GENETIC INVESTIGATION OF PLANT ARCHITECTURE, YIELD, AND DIVERSITY IN WINGED BEAN (*PSOPHOCARPUS TETRAGONOLOBUS* (L.) DC.) FOR BREEDING PROGRAMMES

ALBERTO STEFANO TANZI

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

NOVEMBER 2018

School of Biosciences
University of Nottingham Malaysia,
Jalan Broga 43500, Semenyih, Selangor Darul Ehsan
Malaysia

"*La mia intelligenza? Più che mediocre. I miei unici meriti sono stati impegno e ottimismo*."

("My intelligence? More than mediocre. My only merits have been commitment and optimism")

Rita Levi-Montalcini (1909-2012), Nobel Prize in Physiology and Medicine

Abstract

Winged bean (*Psophocarpus tetragonolobus*) is an underutilised leguminous crop, cultivated mainly in hot and humid countries in Asia and the pacific area. It is traditionally grown as vegetable, for its green pods and tuberous roots, and also as a pulse crop, for its grain, with all edible parts reported to be rich in protein, vitamins, minerals, and fibre. Its plant architecture is composed of long, intertwining stems and lateral branches, which grow indeterminately and impact on final yield as well as farming practices. Yet, this aspect has remained poorly understood. Efforts in winged bean improvement programme also remain constrained by the lack of molecular tools, for example to carry out genetic dissection of the traits of interest, and genetic analysis of germplasm.

Controlled crosses were therefore performed between different genotypes of winged bean in order to investigate morphological traits contributing to its plant architecture, and investigate their effect and interaction with yield components, for pod productivity. After initial assessment of parental and F_1 individuals, a parental combination (M3 × FP15) was chosen in order to establish a segregating population and to carry out phenotypic analyses. These revealed how *stem length* could be predominantly under control of *internode length* (r_s= .80; p< .01), while the average length of lateral branches could impact more on the final number of *pods per plant* (r_s= .44; p< .001) than the *branch number per plant* alone (r_s= .38; p< .001). Together, the results shed light on key morphological traits and their potential interaction with yield components.

Along with crosses and field assessment, molecular markers were developed. First, a set of 18 genic-Simple Sequence Repeat (SSR) markers was validated, and used to initially evaluate the diversity among crossed genotypes, their heterozygosity, and to validate the obtained hybrids. A second set of markers consisted of Single Nucleotide Polymorphisms (SNPs), developed through a genotyping-by-sequencing method on 91 accessions, including parental genotypes, and on the analysed F_2 population.

The SNPs across the genotyped biparental F_2 population have allowed the construction of the first genetic linkage map in winged bean. Nine linkage groups (LG), seemingly corresponding to the 9 chromosomes in winged bean (2n= 2x= 18), were obtained with an average of 199 SNPs per LG. High segregation distortion was revealed by analysing the segregation patterns of the markers, leading first towards the development of a

framework map without distorted marker. A complete map was then constructed by re-introducing distorted markers, resulting in a map with a total length of 1171.6cM, with 395 spaced SNPs (3.1cM/SNP) and an overall 34% of segregation distortion.

QTL analysis was carried out to dissect measured traits, by combining recorded F_2 phenotypic data and the genetic information from the genetic linkage map. Through Interval Mapping (IM), 5 segregating QTLs and 8 putative QTLs were found across 9 traits, all explaining more than 10% of observed phenotypic variance. These included markers linked to genomic regions underlying traits such as *stem length*, *branch number per plant*, *length of branch*, and *dry pod length*.

Using a total of 5891 SNPs discovered across 91 accessions allowed us to perform a preliminary genetic diversity analysis of winged bean germplasm, collected from public and private sources. Genetic distance analyses showed potential agreement between genetic and geographic structure, in contrast with previous studies. The average 7.3% of observed heterozygosity could support the idea of winged bean as a predominantly inbreeding species, although within populations were still found relatively high levels of diversity. Overall fixation index (F_{st}) and pairwise population comparisons revealed moderate significant differentiation between analysed population groups, although South American and African accessions could have been derived from South East Asian and Papua New Guinea stocks.

Overall, the phenotypic analysis provides new insight for the design of winged bean ideotypes with reduced vegetative growth and improved pod productivity, while the QTL analysis gives the first potential markers linked to important traits that could be considered breeding targets. Together, this information provides the initial basis for altering winged bean plant architecture and supporting breeding programmes for the improvement of this crop.

List of Publications

Cheng, A., Chai, H. H., Ho, W. K., Aliyu, S., Abdullah, B., Feldman, A., Kendabie, P., Halim, R. A., **Tanzi, A.**, Mayes, S., Massawe, F. (2017). Molecular Marker Technology for Genetic Improvement of Underutilised Crops. In S. N. A. Abdullah, C. L. Ho, & C. Wagstaff (Eds.), *Crop Improvement - Sustainability Through Leading-Edge Technology* (pp. 47–70). Springer International Publishing. https://doi.org/10.1007/978-3-319-65079-1

Wong, Q., **Tanzi, A. S.**, Ho, W., Malla, S., Blythe, M., Karunaratne, A., Massawe, F., Mayes, S. (2017). Development of Gene-Based SSR Markers in Winged Bean (Psophocarpus tetragonolobus (L.) DC.) for Diversity Assessment. *Genes*, *8*(3), 100. https://doi.org/10.3390/genes8030100

Tanzi A. S., Ho W. K., Massawe F., Mayes S. (2019). Development and interaction between plant architecture and yield-related traits in winged bean (Psophocarpus tetragonolobus (L.) DC.). *Euphytica* 215:36. https://doi.org/10.1007/s10681-019-2359-8

Tanzi A. S., Eagleton G. E., Ho W. K., Quin Nee W., Mayes S., Massawe F. (2019). Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for food and nutritional security: synthesis of past research and future direction. *Planta* 1–21. https://doi.org/10.1007/s00425-019-03141-2

Acknowledgments

To anyone reading this thesis, it's fair to say that I would not have made it without the help of a few people, and this section is dedicated to them.

I would like to start by mentioning my gratitude to Dr Sean Mayes and Dr Festo Massawe for their guidance, encouragement, corrections, trust, and for bearing with my English. Many thanks for giving me this opportunity. Along with them, I would like to thank Dr Wai Kuan Ho (my unofficial supervisor) for her guidance, training, and advice along this project. Many thanks also to Dr Graham Eagleton, which dedication towards winged bean has been an inspiration, for his advice, comments, and for sharing his experience with me.

I would like to thank Crops For the Future, including all the guarantors, for awarding me the DTP scholarship and for the training along the years. Many thanks to the University of Nottingham Malaysia Campus for the support as student. Thanks to Dr Hui Hui Chai for her help in some analysis, and to the Field Research Centre team: GT, Vincent, Marjan, Zharul, Fadhil, Kairi, and all the workers that have made the field work possible… and for letting me drive the tractor.
For their help in the lab, thanks to all the staff in UNMC and CFF.

As lonesome as the PhD path could be, thankfully there were people around making it better every day. Among these, I would like to thank Jennie, Jon, Kasra, Ahmed, and Bhavya, for making me feel a bit more "at home". Thanks also to Norain, Melissa, Nengi, along with other colleagues in UNMC and CFF.
Many thanks to my friends back home and elsewhere: Fuggy, Pava, Mando, Carlo, Olby, Simo, Andy, Vinnie, Aley, Sissi, and Alice. Grazie per non avermi fatto sentire lontano da casa.

As always, I want to thank to my mother, whose support during my studies has allowed me to get this far. The parcels full of food sent over to Malaysia have been much appreciated as well. Thanks to both my parents, for raising me without being afraid to explore (and for introducing me to Chinese food since I was a kid)

Finally, I want to thank Treshi: for all the cookies, for all the times she took a bus to spend time with me, for her support and care, for all the memories, and for being in my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

(Non-)ND	(Non-)Normally distributed	NJ	Neighbour-joining (tree)
%VE	Percentage of variance explained	NoB	Branch number per plant
°C	Celsius	NoBTOT	Total branch number per plant
AFLP	Amplified fragment length polymorphism	NoN	Number of nodes
(G)bp	(Giga) Base pair	NPK	Nitrogen phosphorous potassium
CDR	Complete randomized design	NTP	Nucleoside triphosphate
cM	Centi-Morgan	PCA	Principal component analysis
CRBD	Complete randomized block design	PCoA	Principal coordinate analysis
DArT	Diversity array technology	PCR	Polymerase chain reaction
DArTseq	Diversity array technology sequencing	PIC	Polymorphism information content
DNA	Deoxyribonucleic acid	PMT	Pod maturing time
DtF	Days to first open flower	PoL	Dry pod length
EST	Expressed sequence tags	PPP	Mature pods per plant
g	Gramme	QTL	Quantitative trait locus
GbS	Genotyping-by-Sequencing	RAPD	Random Amplification of Polymorphic DNA
GD	Genetic distance	RE	Restriction enzyme
HSW	Hundred-seeds weight	RFLP	Restriction fragment length polymorphism
IM	Interval mapping	RNA	Ribonucleic acid
ISSR	Inter-simple sequence repeat	SD	Standard deviation
KW	Kruskal-Wallis analysis	SD	Segregation distortion (chapter 4)
L	Litre	SLB	Total sum of branch length
LeL	Leaf length	SNP	Single nucleotide polymorphism
LeW	Leaf width	SPAD	Chlorophyll concentration
LG	Linkage group	SPP	Seeds per pods
LoB	Length of branch	SSR	Simple sequence repeat
LOD	Logarithm of Odds	StL	Stem length
LPP	Number of leaves per plant	UPGMA	Unweighted Pair Group Method with Arithmetic Mean
MAS	Marker-assisted selection	WGR	Weekly growth
Mdn	Median	µg	Microgram, equal to 10^{-6} gramme
mL	Millilitre, equal to 10^{-3} litre	µL	Microlitre, equal to 10^{-6} litre
MQM	Multiple-qtl model	µM	Micromole, equal to 10^{-6} mole

Chapters Synopsis

Every chapter has been written following a research paper format. As such, each is composed of an abstract, introduction, material and methods, results, discussion, conclusions, and reference list. Chapter 1 Section A and chapter 3 have been submitted for publication and are currently under review.

Chapter 1 Section A provides a literature review on winged bean covering past research, in particular about growth, phenology, photo-thermal regulation of development, breeding efforts, molecular tools, and future targets (published in March 2019, doi: 10.1007/s00425-019-03141-2). Section B introduces molecular markers and genotyping-by-sequencing technologies that have been used in this study. This section has been published as book chapter (Springer International Publishing. https://doi.org/10.1007/978-3-319-65079-1). Aims and objectives of this project are presented in the last section (Section C).

Chapter 2 reports the development and validation of a set of genic-SSR marker in winged bean, across selected genotypes. These markers were then used for an initial assessment of the genetic diversity of the parental material, and for the validation of the F_1 hybrids. This chapter was published in March 2017 (*Genes*, *8*(3), 100. https://doi.org/10.3390/genes8030100).

Chapter 3 describes the generation, field assessment, and analysis of a winged bean F_2 population. It reports the phenotypic analysis of the recorded traits related to plant architecture, development, and yield components. The results have highlighted the correlation between some morphological traits, like *branch number per plant*, and yield components, like *pods per plant*. They also providing an initial insight on what traits should be targeted in order to alter plant architecture, and to improve final yield. Additional information is reported in the appendix at the end of the chapter, which include SSR validation of the F_1 hybrids, and further data on morphological and physiological traits. This chapter has been published in January 2019 (doi: 10.1007/s10681-019-2359-8)

Chapter 4 presents the construction of a SNP-based genetic linkage map, using 221 winged bean F_2 population (both phenotyped and non-phenotyped individuals). SNP were generated through Diversity Array Technology sequencing (DArTseq). Two maps were eventually constructed, in order to manage the detected segregation distortion.

Chapter 5 presents the results obtained from QTL analysis for a total of 13 recorded phenotypic traits. The analysis was carried out using the genetic linkage maps presented in chapter 4. Five segregating QTLs, and 8 putative QTL were found distributed across 9 traits. The discussion attempts to match the phenotypic data reported in chapter 3, with the discovered QTLs.

Chapter 6 reports the genetic diversity analysis carried out upon 91 accessions of winged bean, retrieved from genebanks and from private collaborators: across them, SNPs were generated through DArTseq. Frequency and distance-based analysed showed potential agreement between genetic and geographical structure, in contrast with past investigations. A STRUCTURE analysis suggested a maximum number of clusters that could represent most of the structure underlying the analysed germplasm. The results would support more effective selection of individuals, to efficiently include their genetic diversity in improvement programmes.

Chapter 7 integrates the whole project in a general discussion and conclusions. Suggestions for future work are also provided both in terms short-term and long-term targets for the improvement of winged bean, for its wider utilisation in agriculture.

Chapter 1 - General Introduction

Section A) Literature review on Winged bean, and breeding targets

The following section is the original manuscript submitted for publication in September 2018. This, has been accepted for publication in March 2019:

Tanzi AS, Eagleton GE, Ho WK, et al (2019) Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for food and nutritional security: synthesis of past research and future direction. Planta 1–21.
doi: 10.1007/s00425-019-03141-2

Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for food and nutritional security – synthesis of past research and future direction

Alberto Stefano Tanzi[1,2], Graham Ewen Eagleton[3], Wai Kuan Ho[1,2], Quin Nee Wong[1], Sean Mayes[2,4], Festo Massawe[1,2*]

[1]Biotechnology Research Centre, School of Biosciences, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
[2]Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
[3]23 Mulhall Street, Wagstaffe, NSW 2257, Australia
[4] School of Biosciences, Faculty of Science, University of Nottingham Sutton Bonington Campus, Sutton Bonington, Leicestershire LE12 5RD, UK

Address for correspondence: (Tel.: +60 3 8924 8218; E-mail: festo.massawe@nottingham.edu.my)

Abstract

Reliance on a handful of "major" crops has led to decreased diversity in crop species, agricultural systems and in our diets. To reverse this trend, we need to encourage the greater use of minor, "orphan", underutilised species. These could contribute to an increase in crop diversity within agricultural systems, to improve human diets, and to support more sustainable and resilient food production systems. Among these underutilized species, winged bean (*Psophocarpus tetragonolobus*) has long been proposed as a crop for expanded use particularly in the humid tropics. It is an herbaceous perennial legume of equatorial environments and has been identified as a rich source of protein, with most parts of the plant being edible when appropriately prepared. However, to date, limited progress in structured improvement programmes has restricted the expansion of winged bean beyond its traditional confines. In this paper, we discuss the reasons for this and recommend approaches for better use of its genetic resources and related *Psophocarpus* species in developing improved varieties. We review studies on the growth, phenology, nodulation and nitrogen fixation activity, breeding programmes, and molecular analyses. We then discuss prospects for the crop based on the greater understanding that these studies have provided and considering modern plant-breeding technologies and approaches. We propose a more targeted and structured research needed to fulfil the potential of winged bean to contribute to food and nutritional security.

Keywords: winged bean; *Psophocarpus tetragonolobus*; underutilised legume; crop improvement; food security.

1 Introduction

Despite the rich variety of crops once cultivated around the world, since the beginning of the 20th century there has been a consistent decline in the number of species included in our diets. Currently, more than half of our food comes from only three major cereal crops, cultivated at large scale in monocultures. This has resulted in a lack of sufficient agro-diversity (FAO 2010, 2015), making food production systems more vulnerable whenever more stressful conditions occur (drought above all, but pests and diseases too), such that mono-cropping systems could fail on a large scale, undermining production and availability of food (FAO 2010, 2011; Rattalino Edreira et al. 2011; Siebert et al. 2017; Zampieri et al. 2017).. At the same time, lack of genetic diversity hampers the development of improved varieties of major crops, by reducing the availability of beneficial gene-combinations to incorporate into plant breeding programmes. There is now also a growing recognition of the value of the link between agrobiodiversity (defined as the number of different crop species in a system) and sustainability, socio-economic resiliency and improved human health (Chappell and LaValle 2011; Pellegrini and Tasciotti 2014; Khoury et al. 2014; Adhikari et al. 2017; Dwivedi et al. 2017). For these reasons, neglected and underutilised crop species have received increased attention in recent years, thanks to their potential to preserve and increase agrobiodiversity, to improve nutrition, and to tolerate or resist abiotic and biotic stresses (FAO 2010, 2015; Mayes et al. 2012; Padulosi et al. 2013; Chivenge et al. 2015; Massawe et al. 2016; Adhikari et al. 2017; see also the African Orphan Crops Consortium initiative www.africanorphancrops.org). Not that this focus on forgotten crops for fresh purposes is a new phenomenon. In 1974, the US National Academy of Sciences (NAS) convened a select committee to carry out "an extensive survey of underexploited tropical plants" as possible crops for the future. Among 36 species they considered in detail (NAS 1975a), "the exceptional merits" of one particular plant, the winged bean (*Psophocarpus tetragonolobus* L. DC.), was singled out for special attention and for promotion to the agricultural research community (NAS 1975b). This is an underutilised leguminous species, cultivated mainly in hot and humid countries for its green pods, tuberous roots, mature seeds, and leaves. All these have been reported to be rich sources of protein, carbohydrate, vitamins, minerals and fibre (Gillespie and Blagrove 1978; Anonymous 1980; NAS 1981; Kortt 1983; Kadam et al. 1984; Henry et al. 1985; Kantha and Erdman 1986; Mnembuka and Eggum 1995), while particularly effective symbiotic associations with a broad spectrum of rhizobia strains

makes the winged bean a good nitrogen-fixer for low-input and self-resilient agricultural systems (Burkill 1906; Masefield 1961; Anonymous 1980; Ikram and Broughton 1980; Iruthayathas and Vlassak 1982; Klu and Kumaga 1999) .

Despite the international attention dedicated to this crop during the 1970's and 80's, research has been inconsistent ever since, and thus the crop remains cultivated mainly on a subsistence scale. While substantial information came from the initial efforts especially on plant development under different growing conditions and nutritional value, germplasm characterization and trait dissection were difficult in the absence of molecular markers, at that point in time. Hence, a fresh approach could include phenotypic characterization of broader germplasm, the implementation of new technologies for genetic diversity analysis and trait dissection (e.g. Genotyping y Sequencing, GBS), and the translation of knowledge from other crops. This would help to evaluate the variability within species both at a genomic level, while understanding the breeding system and inheritance of key agricultural traits. For some traits, these approaches could bring advantages where past efforts failed, for example, in altering plant growth habit/architecture, in raising and stabilising yield of desired plant parts, and in guaranteeing the nutritional value and safety of end products. Revamped efforts have indeed started to take this path, with recent work on transcriptomic profiles and molecular marker development (Chapman 2015; Vatanparast et al. 2016; Singh et al. 2017; Wong et al. 2017).

The present review provides a detailed analysis of past research on the conditions that influence winged bean as a plant and as a crop. We provide a summary of experiments on the photothermal regulation of key developmental events, of its nodulation and nitrogen fixation activity, of past breeding efforts, and of the key areas for future research. We also discuss the recent molecular approaches, especially those that have investigated genetic diversity and population structure so far. Drawing from past knowledge, and based on current technologies, we propose a roadmap for research and development of winged bean for future agricultural systems.

2 Winged bean in brief

Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) is a dicotylenous plant, taxonomically classified under the Fabaceae family, Papilionoideae subfamily. It has a diploid genome (2*n* = 2x= 18), consisting of a karyotype with three pairs of short and six pairs of long chromosomes (Harder 1992), recently determined at around 1.22Gbp/1C in size (Vatanparast et al. 2016), while previous estimation by flow cytometry suggested an haploid genome size of 782Mbp (Bennett and Smith 1976). It is considered to have a cleistogamous floral system, which would usually imply autogamy, with self-pollination having been observed to take place before the large flowers open in the morning hours (Karikari 1972; Erskine and Bala 1976; Erskine 1980). Such observations have been supported by experiments with bagged flowers (Karikari 1972; Anonymous 1980), suggesting that insects are not required for pollination.. Nonetheless, analysis of phenotypic markers (e.g. stem colour) have revealed a 7.6% of out-crossing during the wet season in Papua New Guinea (PNG), facilitated by carpenter bees (*Xylocopa aruana*) (Erskine 1980). In support of this study, pollen obtained from fully opened flowers was found to be viable for approximately 24 hours (Koshy et al. 2013), while the stigma remained receptive for 33 hours (Senayake and Sumanasinghe 1978), providing evidence for possible cross-pollination events.

Winged bean grows in hot, humid, equatorial countries of Southern Asia, Melanesia and the Pacific area. It can be grown from sea level up to frost-free altitudes of about 2000 meters above sea level, with its main presence found in India, Sri Lanka, Bangladesh, Myanmar, Thailand, Laos, Vietnam, Cambodia, Malaysia, Indonesia, Philippines and PNG (NAS 1975b; Khan 1976; Drinkall 1978; Eagleton 1999). Its centre of origin, though, is still a matter of dispute. Two contrasting hypotheses have been suggested. The first hypothesis suggests the origin of *P. tetragonolobus* to be the African continent, with either *in situ* domestication and then migration (by human activity) of the domesticated species east, or trans-domestication of an African progenitor species from which winged bean has been derived later in the Pacific Asia. In support of an Africa-centric hypothesis, the chromosome number and karyotype pattern of winged bean are consistent with five *Psophocarpus* species from Africa (Khan 1976; Pickersgill 1980; Harder 1992): namely *P. scandens*, *P. grandiflorus*, *P. palustris*, *P. lecomtei*, and *P. lancifolius* (Harder and Smartt 1992). Attempts have been made

to identify the possible wild African progenitor, identified as *P. grandiflorus*, for its morphological resemblance and shared susceptibility to the fungus *Synchytrium psophocarpi* (Smartt 1980; Harder and Smartt 1992), or as *P. scandens*, after the more recent phenetic (Maxted 1990) and cladistic (Fatihah et al. 2012) analyses. The second, alternative, hypothesis postulates that *P. tetragonolobus* is a species distinct from current African members of the genus (Verdcourt and Halliday 1978) and arose through a mechanism of allopatric speciation preceding any purposeful or unwitting processes of domestication. The main limitation of this hypothesis, which shifts the origin of winged bean towards the east, is the lack of wild forms of this species in Asia or Melanesia. Burkill (1906) believed that the linguistic and historical evidence pointed to the origins of winged bean on the western fringes of the Indian Ocean, while Vavilov (1951) placed it in his hypothesised Indian centre of crop domestication. Hymowitz and Boyd (1977), on the other hand, proposed a Melanesian centre of origin based on the discovery of a significant centre of genetic diversity for winged bean in the highlands of PNG. However, Khan (1976) who was the first to assess and report on this centre of diversity, suggested that the winged bean was in fact a relatively recent prehistoric introduction from the west into such area. Turning to modern molecular tools, Yang and colleagues (2018) carried out the first phylogenetic investigation through analysis of three regions in the chloroplast genome, and a nuclear internal transcribed spacer (ITS). The results showed that none of the four *P. tetragonolobus* accessions - originally from Nigeria, Liberia, and Malaysia - fell into any of the other three groups comprising of *Psophocarpus* genus members. The authors suggested the possibility that Verdcourt and Halliday were right in that winged bean could have a distinct history from the rest of the African species, and for which the progenitor has been lost or remains undiscovered. Noteworthy, they also reported success in hybridizing winged bean with *P. scandens*, which could support this latter as the closest relative after all, and whose progeny could form a bridge for future trait introgression from the African material.

Winged bean has a wide range of uses and most parts of the plant are consumed, depending on where the crop is cultivated. The young pods are a popular edible part across all cultivation areas, eaten raw or as a cooked vegetable (NAS 1975b; Sastrapradja and Aminah Lubis 1975), and in Indonesia mature seeds are roasted like peanut and boiled (NAS 1981). In PNG, immature seeds are consumed due to their pea-like taste (Claydon 1975; Okezie and Martin 1980). In Myanmar, Thailand, and PNG, boiled, steamed, baked, fried or roasted

tubers are also consumed (NAS 1975b; Eagleton 1999). Leaves and flowers are also consumed (NAS 1975b; Okezie and Martin 1980; Anonymous 1981), and younger parts are particularly appreciated when infected with the mentioned *S. psophocarpi* in Java, Indonesia (Hymowitz and Boyd 1977). Green or purple young pods have been reported, with presence or absence of specks (Erskine & Khan, 1977; NAS, 1975). The seed coat colour has been reported to range from cream (NAS 1975b; Anonymous 1982; Mohanty et al. 2013), black (Klu and Kumaga 1999), brown (NAS 1975b; Klu and Kumaga 1999; Mohanty et al. 2013), to purple and mottled colours (NAS 1975b).

3 Synthesis of past research

Past efforts were, in part, directed by a dedicated research centre based in Sri Lanka, funded by the International Council for Development of Underutilized Plants, ICDUP (Anonymous 1980). Among research-led initiatives, an international trial was set up to evaluate a set of accessions from across different countries, not dissimilar from the recent worldwide evaluation of Quinoa (*Chenopodium quinoa*) (Bazile et al. 2016). Great emphasis was dedicated to five main areas: 1) growth and phenology assessment in different growing conditions (Herath and Ormrod 1979; Anonymous 1980, 1981, 1982; Rüegg 1981; Misra and Misra 1985; Okubo et al. 1992); 2) development of high seed-yielding varieties (Nangju and Baudoin 1979; Erskine 1981a; Arumugan and Perera 1989); 3) nutritional profiling of edible parts (Anonymous 1980; Kortt 1980; Gross 1983; Kadam et al. 1984; Prakash et al. 1987); 4) nodulation ability (Hildebrand et al. 1981c; Iruthayathas and Herath 1981; Iruthayathas and Vlassak 1982, 1987; Iruthayathas et al. 1985) and 5) biotic and abiotic-stress tolerance studies (Karikari 1978; Thompson and Haryono 1979; Rüegg 1981; Price et al. 1982; Weil and Khalil 1986; Prakash et al. 1987). Following on from the initial concerted research effort, winged bean research activities have expanded to include further characterisation of anti-nutritional factors in mature seeds (Umemoto et al. 1992; Habu et al. 1992, 1997; Mukhopadhyay 2000), factors affecting growth (Schiavinato and Válio 1996a, b), and recently development of genetic resources such as molecular markers for genetic diversity assessment (Mohanty et al. 2013; Chen et al. 2015; Wong et al. 2017; Yang et al. 2018) or clonal fidelity (Koshy et al. 2013b), and transcriptome studies (Chapman 2015; Vatanparast et al. 2016; Wong et al. 2017).

3.1 Vegetative growth

In winged bean, a good vegetative growth of established seedlings requires warm, well lit, humid conditions. Several studies have shown that a day temperature of around 27°C and a night temperature above 20°C is optimal (Wong and Schwabe 1979; Herath and Ormrod 1979; Anonymous 1980, 1981; Wong 1983; Schiavinato and Válio 1996a). Reduced plant survival rate and plant height was observed under high temperature of $36/31^0$C as compared to $27/22^0$C, from the accessions collected from upland and lowland winged beans accessions. The growth suppression was found to continue even when the temperature was reduced to $33/28^0$C (Anonymous 1980). Lenz and Broughton (1981) working with the PNG accession UPS99 demonstrated that soil temperature was more critical than temperature above ground, suggesting that the optimum soil temperature for both vegetative growth and root nodule development was around 25°C, whereas photosynthetic rates in the plant shoots could be sustained at 35°C, and N-fixation (measured by acetylene reduction assay) in the nodules at 30°C. However, significant growth variation in response to different conditions has been reported. Herath and Ormrod (1979), working with a range of PNG and Sri Lankan accessions showed differences in leaf area, stomatal density, and total chlorophyll content among cultivars, and in response to temperature and daylength. The PNG accessions showed higher stomatal density than Sri Lankan accessions, while both groups showed higher values at a 25/20°C temperature regime than at 30/25°C. Sri Lankan material was able to achieve greater shoot biomass accumulation than PNG individuals under any treatment combination, perhaps due to different photosynthetic rates, but, overall, the highest dry weights were achieved either at 25/20°C with 14 hours photoperiod, or at 30/15°C with 11 hours photoperiod.

In another experiment, a lowland Malaysian accession (M14/4) grown at 8 hours daylength was found to achieve the highest total dry matter accumulation at a temperature of 26/26°C, with growth progressively decreasing at 32/18°C, 20/14°C, and 18/18°C. The day temperatures of 18°C was found to depress growth more than the night temperature of 14°C. At an optimum temperature of 26/18°C, M14/4 reported total dry matter per plant, relative growth rate, net assimilation rate, leaf number, main-stem length and main stem node number to be higher at 16-hour rather than 8-hour daylength regime (Wong and Schwabe 1979; Wong 1983), a finding generally supported by other studies (Sinnadurai and Nyalemegbe 1979; Schiavinato and Válio 1996a). However, at higher temperature regimes (32/22°C) there was no significant increase in biomass accumulation under 16-hour, compared to 8-hour daylength. This was, to an extent, in agreement with the

previous findings from Herath and Ormrod (1979). However, at a constant 13-hour photoperiod, a higher temperature regime of 30/25°C led to higher dry mass accumulation and leaf area than under 25/20°C, without significant differences between these two regimes with regard to plant height or number of leaves (Schiavinato and Válio 1996a). Taken together, these finding suggest that winged bean has an integrated response to temperature and daylength, with higher biomass accumulation under day-temperature between 25-30°C and short-day photoperiod. However, substantial variation was observed between different accessions, when more than one was analysed, for traits like stomatal density, which could lead to different responses under sub-optimal growth conditions (e.g. longer daylength and higher temperature).

In the field, winged bean is less tolerant to soil moisture deficits than most tropical legumes. In assessing 36 common leguminous crop species found in the tropics, Rachie (1977) classified winged bean as one of six crops grown mainly in the humid zone requiring an annual rainfall of 1500 mm and above. It is also virtually the only one in the humid zone that can be found growing at altitudes over 1000 meters a.s.l.. Nevertheless, in Ghana winged bean has been grown without irrigation, although the yield reduced below its expected maximum when the annual rainfall dropped below 1120 mm (Anonymous 1978, 1980). Similarly, Nangju and Baudoin (1979) found that winged bean (TPt2) seed yield was almost double that of cowpea (*Vigna unguiculata*), but lower than elite cultivars of soybean (*Glycine max*), pigeon pea (*Cajanus cajan*), and jack bean (*Canavalia enformis*) when grown at Nigerian site (Ibadan 7°N) with annual rainfall of 1200 mm. On the other hand, winged bean was the top performer in another site with a moist environment of 2400 mm annual rainfall (Onna 5°N). This would imply that winged bean has greater water requirement for optimal growth compared to other legumes. A further drought screening characterisation with a larger set of winged bean germplasm from different ecotypes could reveal variation in response to water deficit stress.

3.2 Effects of daylength and temperature on reproductive development

Winged bean is a short-day species, exhibiting photothermal sensitivity during developmental progression towards flowering, pod production, and initiation of tuberous roots. Flower bud development seems inhibited by long daylength (more than 13-hour daylength), although this depends on day/night temperature as well as genotype (see also 3.1 above). A study conducted in Ghana (Accra, $5^{0}30'$ N) using the accession TPt1 from

IITA (International Institute of Tropical Agriculture) showed days to 50% flowering to vary from 40 to 130 days between December (daylength of 11 hours 50 minutes) and March (daylength 12 hours 25 minutes) planting (Sinnadurai and Nyalemegbe 1979). The authors attributed this shift in flowering timing principally to the change in daily photoperiod. Further studies carried out by Eagleton (1983) included TPt1 along with another 24 accessions from Malaysia, PNG, Sri Lanka, Thailand, and Indonesia. Trials were conducted with trellising/staking structures across a range of planting dates between October and December in Malaysia (Serdang 3°8'N) and in two locations in Western Australia (between 31° and 34°S). Overall, accessions began to flower during March in the Australian environment, irrespective of planting date or temperature conditions during vegetative growth stage, and significantly later than in the Malaysian trial (57 days in average *vs* a minimum of 116 days in Australia). TPt1, followed by UPS99 (PNG) were the least affected accessions by change of daylength between the two environments, while M13/1 (Malaysia) showed greater dry mass accumulation during the same period. Eagleton and colleagues concluded that for subtropical conditions it could be feasible to exploit the variation in accessions collected from different ecotypes. Seed yield in Australia could be improved by sowing in spring, and in combination with early flowering, fast growing genotypes.

In addition to field studies, experiments conducted under controlled environments have provided more evidence on how temperature and daylength (alone or in combination) control the phenology of winged bean, and also revealed the genotypic variation that exists among different winged bean accessions (see a summary of the experiments in Table 1). For example, Herath & Ormrod (1979) reported that none of the six Sri Lankan and fifteen PNG accessions flowered under 30/25 °C temperature regime at either 11- or 14-hour photoperiod. Flowering was observed only in the 11-hour photoperiod in a 25/20°C temperature, with PNG accessions producing significantly more racemes (14) and flowers (31) within the first 60 days of growth, compared to Sri Lankan material (average of 5 racemes and 15 flowers), where 3 accessions failed to flower completely. By contrast, Rüegg (1981) found that in short daylength (12 hours) the actual time to flowering in accessions from PNG, Nigeria, Ghana, and Costa Rica was shortened by warmer temperatures (27/23°C) compared to cooler conditions (22/18°C), leading to the highest grain yield for all accessions in the experiment. The UPS99 accessions (PNG) was among the earliest to flower under the optimal experimental conditions, and the least affected under lower temperature compared to the other accessions. This overall trend reported by Rüegg, in

disagreement with Herath and Ormrod's findings, was also reported by Eagleton and colleagues (Eagleton 1983) in a controlled environment experiment with another four winged bean accessions of Malaysian and mainly PNG origins.

Table 1: Summary of accessions and treatment imposed in some of the main phenological studies in controlled conditions on flowering in winged bean.

Reference	Accession (Origin)	Daylength and Temperature (day/night) regimes
Eagleton et al. (Anonymous 1980)	UPS32, UPS47, UPS121 (PNG) (Malaysia) M13/1	10-, 12-, 14-hour; 26/16°C, 27/22°C, 33/28°C
Herath & Ormrod, 1979	SL3, SL7, SL8, SL11, SL17, SL18 (Sri Lanka) UPS31, UPS45, UPS46, UPS59, UPS61, UPS66, UPS78, UPS80, UPS102 (PNG)	11-, 14-hour; 25/20°C, 30/25°C
Wong & Schwabe, 1979	M14/4 (Malaysia)	8-, 16-hour 32/18°C, 26/26°C, 26/14°C
Rüegg (1981)	Kade1-26 UPS32, UPS122 (PNG) TPt8 (retrieved from IITA)	12- hour 22/18°C, 27/23°C
Uemoto et al., 1982	nine, including UPS31, UPS62, UPS99 (PNG) 1 (Sri Lanka); 1 (Thailand)	8-,11-,13-hour; 20°C, 25°C, 30°C (constant day/night)

Uemoto and colleagues (1982) investigated raceme budding in nine PNG and two Asian (from Sri Lanka and Thailand) accessions mainly under 8-, 11-, or 13-hour daylength regimes, with temperatures of 20°, 25°, or 30°C. They concluded 12 hours being the critical daylength threshold for most of the genotypes, although UPS99 was able to produce buds at up to 15 hours of daylength when kept at 20°C. Moreover, in a following experiment, when UPS99 seedlings were raised in inductive conditions of 20°C mean temperature and 8-hour daylength until second trifoliate leaf stage, they continued to produce plentiful buds even when the daylength was subsequently extended to 16 hours (8-hour natural light and 8-hour incandescent light). The relative stability of the phenology of UPS99 in response to photothermal variability (Rüegg 1981; Uemoto et al. 1982; Eagleton 1983) was demonstrated also in the first international replicated field trials, in which UPS99 was the earliest among ten PNG accessions and three Asian accessions to initiate flowering, across twelve tropical and subtropical locations ranging from 1°N (Singapore) to 27°N (Nepal). Even so, the overall mean for the number of days to flower across all sites was as much as 62 days after sowing (DAS) for UPS99, 67 DAS for all PNG accessions taken as a group, and 87 DAS for the mean of the three Asian accessions (Anonymous 1981; Eagleton et al. 1985).

Wong and Schwabe (1979) and Wong (1983) inferred from Malaysian accession (M14/4), under a thorough controlled environment examination that the 'critical' daylength for flower induction was between 11h15min and 12h15min, a conclusion that was not dissimilar from the studies mentioned so far. It was also reported that a reduction in light intensity, during part of the photoperiod, lowered this critical daylength. At short, potentially inductive daylengths, the optimal day temperature was found to be 26°C, whereas 32°C or 18°C inhibited flowering, again in broad agreement with most of the other studies. In apparent contrast to the study by Uemoto and colleagues (1982) with UPS99, Wong (Anonymous 1981) reported that further bud development through to anthesis in the Malaysian accession did not take place if plants were returned to long days without the sufficient number of short days.

Rüegg (reported in Anonymous 1980) identified a very useful variant of the IITA accession TPt8, during controlled environment studies. The author followed the development of four diverse accessions, TPt8 (IITA), Kade1-26, UPS32 and UPS121 (PNG), at two temperature regimes finely calibrated to simulate field conditions at upland and lowland sites in PNG; namely, Wau at 1200 m a.s.l. (22/18°C) and Lae at sea level (27/23°C). Flowering was initiated even at a daylength of 16 hours when the temperature was held at 22/18°C but not at 27/23°C. In general agreement with the mentioned studies, Schiavinato and Válio (1996a) found in an Indonesian accession that at a daylength of 13 hours, a temperature regime of $25/20^0$C was required for flower induction (flowering occurred within 52 days after germination for all individuals); flowering was instead delayed in regimes of $20/15^0$C or $30/25^0$C (96 and 63 days after germination, respectively).

In several studies (see Table 2) individuals exhibiting beneficial reduced photothermal sensitivity have been found in some of the accessions. This demonstrates that photothermal-insensitive genotypes could be selected from large screening of germplasm, potentially leading to identification of different individuals with different genetic base for this trait, as shown in crops like soybean (Watanabe et al. 2012; Xu et al. 2013a) and common bean (*Phaseolus vulgaris*) (Kwak et al. 2008).

Table 2: Accessions and variants with particular phenological traits across studies.

Reference	Location	Description

Kim (Anonymous 1981)	Seoul, South Korea (37°34' N)	variant of UPS63 (PNG) that flowered earlier than UPS99
Okubo et al., 1992; Endo et al., 1993	Kyushu University, Japan; Iwate prefecture, Japan (39°42' N)	Detection of several completely photo-insensitive varieties for summer pod production; two of these, KUS8 and KUS12, planted in May yielded 19 and 16ton/ha of pods respectively
Rüegg (Anonymous 1980); Eagleton & Sandover, 1984	Controlled environment; Kununurra, Australia (15°46'S)	Early flowering variant of TPt8 renamed TPt8-ETH1 was found less photothermal sensitive. Tested in Australia, it flowered 64 DAS

3.3 Photothermal regulation of tuberization

The TPt8-ETH1 variant found by Rüegg (Anonymous 1981) not only was the first to flower at the peak of the annual daylength and temperature cycle in the field experiment in Australia (Eagleton & Sandover, 1984) (see Table 2), producing its first mature seed pods 132 days after planting, but it also gave the highest yield of root tubers, 939 gm/m^2 when all plots were harvested in mid-August, eight months after planting. Rüegg (1981) demonstrated that Tpt8 and UPS122 produced higher tuber yield at 22/18°C than at 27/23°C, the reverse of what was found for pod and seed yields. This concurred with the findings of extensive multisite field trials in PNG which showed that upland locations produced higher tuber yields on average than lowland sites; while, again, the reverse was the case for pod yields, as suggested by Kesavan, and Stephenson and colleagues (Anonymous 1981). These findings have also been supported by Schiavinato and Válio (1996a) who found that a temperature of 30/25°C, individuals failed to initiate any tubers while at 25/20°C they produced significantly more tubers with a higher overall root dry weight yield than a temperature of 20/15°C. There is ample evidence to suggest that tuber development in winged bean requires short days (Okubo et al., 1992; Wong & Schwabe, 1979; K.C. Wong, 1983). For example, Wong and Schwabe (1979) reported that M14/4 plants grown under 16-hour daylength produced no tubers. Wong (1983) also found that reducing the intensity of daylight below 1100 µmol $m^{-2}s^{-1}$ inhibited tuberisation. As was the case for flowering, plants had to be at least five weeks old before tubers could be initiated by inductive daylengths (see also Okubo et al., 1992). However, unlike for flowering, tuber development proceeded even under non-inductive conditions. Okubo and colleagues (1992) reported that some winged bean accessions (see Table 2) that appeared to have reduced photoperiod sensitivity for flowering, were also relatively photoperiod insensitive for tuber initiation, a finding that echoes a report by Eagleton and Sandover (1984).

Overall, the findings have shown that both temperature and daylength control the onset of flowering and tuberization in winged bean. The variation shown by accessions from different areas might be related to the ecotype and where such accessions have been cultivated and undergone the initial adaptation process, such that there is the potential to find improved material that could lead to the development of early maturing and higher yielding varieties of winged bean, able to flower and produce tubers beyond the current geographical distribution.

3.4 Nodulation, N_2 fixation activity, and applications

Winged bean plant has great capacity to nodulate, fix nitrogen and survive on a range of tropical soils: from poor acidic clay and loam soils in Puerto Rico (Anonymous 1980), to sandy, swamp peats and heavy clay soils in Myanmar and PNG (Burkill 1906; Khan et al. 1977). Across repeated experiments in different locations in Malaysia, Masefield (1957) reported fresh nodule yields of up to 21 grams per plant, which exceeded cowpea, common bean, groundnut (*A. hypogaea*) and *G. soya*. In Puerto Rico, on an uninoculated oxisoil with no history of prior legume cultivations, Harding and colleagues (1978) observed greater numbers and weight of nodules from eight winged bean accessions (from PNG and IITA collection) than from four other tropical legumes [soybean (*G. max* cv. "Ransom"), cowpea, hyacinth bean (*Lablab purpureus*), and pigeon pea (*C. cajan*). Using sandy loam soil collected from a plot where winged bean was previously grown, Iruthayathas & Herath (1981) showed that nodulation could begin as early as the second week after planting. Nodule distribution in accession from Indonesia (LBNC3, LBNC1), Nigeria (TPt2), PNG (UPS122), and Sri Lanka (SLS1, SLS6), was uniform across the rooting system, although with a higher number in the 30 to 60mm root-zone, and a sharp decrease below 90mm, probably due to decreasing aerobic conditions. However, in a follow-up experiment with 4 of the previous accessions, nodule distribution was influenced by the strain of *Rhizobium*, with one in particular (SRI-1, the only one isolated from a winged bean field in Sri Lanka) mostly forming nodules from 90mm and below (Iruthayathas and Vlassak 1982).

Winged bean is a relatively non-selective plant capable of fixing nitrogen in association with a broad-spectrum cowpea rhizobial group as shown in a controlled environment experiment in Papua New Guinea (Elmes 1976), and in field trials in Maryland, USA, using the TPt1 accession from IITA (Hildebrand et al. 1981a). Studies

using rhizobia extracted from different species and genera (e.g. *Arachis*, *Desmodium*, *Dolichos*, *Phaseolus*, *Pueraria*, *Vigna*) have further demonstrated the ability of winged bean to form nodules with a broad spectrum of strains (Ikram and Broughton 1980; Iruthayathas and Vlassak 1987). Such studies reported that some rhizobia were as effective in fixing nitrogen and enhancing winged bean plant growth as were rhizobial strains isolated from field-grown winged bean itself. It should be mentioned that Harding and colleagues (1978) in Puerto Rico reported significant differences in nodulation between winged bean accessions while Kumarasinghe and Kumarasinghe (Anonymous 1981) in Sri Lanka found genotypic differences not only in nodulation but in symbiotic effectiveness and nitrogen fixation. This was followed up by Iruthayathas and colleagues (1985) who classified UPS31, UPS45, UPS121, UPS122 (PNG), SLS6 (Sri Lanka), and TPt8 (IITA) as relatively low nodulating (as fresh nodule weight) N_2 fixing (as C_2H_2 reduction assay) accessions, while LBNC1, SLS44, TPt1 and THAI25-01 (respectively from Indonesia, Sri Lanka, IITA, Thailand) were classified as relatively high ones. The reported variability in nodulation and nitrogen fixation activity could be a significant advantage for future breeding programmes, aiming to study and to develop improved winged bean varieties for mixed agricultural systems. A better characterization of the relatively "low" nodulation and N-fixing accessions might reveal individuals that are able to use more efficiently the fixed nitrogen which could be combined with high fixation accessions through breeding. On a molecular level, the identification of specific targets for improving nodulation could start from the reported Kunitz protease inhibitor exclusively localised in senescent nodule cells in winged bean (Manen et al. 1991). Delays in the expression of this protein, potentially involved in the pathway towards nodule senescence, could lead to prolonged life of nodules in winged bean, to support nitrogen demand even during seed late developmental stages. Such an opportunity is investigated also in soybean (van Wyk et al. 2014; Yuan et al. 2016) and *Medicago truncatula* (Sheokand et al. 2005).

Another breeding target to improve winged bean nodulation could be through seed coat colour. In an assessment of γ-radiation M_4 mutants from UPS122 and cv Kade 6/16, it was reported that that lighter seed coat colour was associated with increased nodules per plant, likely due to the alterations in flavonoid biosynthetic pathways that could influence *nod* genes expression during symbiotic initiation between plant and

rhizobia bacteria (Klu and Kumaga 1999). Similar observations have been made in common bean (Hungria et al., 1991; Hungria & Phillips, 1993) and other species (Mandal et al. 2010).

Nitrogen fixation in winged bean is comparable to other ureide-transporting legumes such as cowpea (*Vigna unguiculata*), Desmodium (*D. intortum*), Siratro (*Macroptilium atropurpureum*) and soybean (Pate et al. 1980; Yoneyama et al. 1986). As demonstrated by Hildebrand and colleagues (1981), a high nitrate treatment (15mM solution) suppressed all nodulation in winged bean and also in a commercial soybean cultivar (cv Williams). The biomass of all above-ground parts was greater in the high nitrate treatment but, not surprisingly, the increase was greater in the soybean than in the winged bean accessions. At the low nitrate level (0.75mM solution), nitrogen fixation per unit nodule mass, as measured by acetylene reduction, was significantly higher in one PNG winged bean accession (introduced to U.S. as "IL13") than in the soybean or the other two winged bean accessions. Leaf nitrate reductase level was significantly higher in all winged bean than in the soybean, and higher for all in the high nitrate treatment than in the low one. All winged bean accessions were significantly higher than the soybean cultivar for leaf and stem protein content in both treatments, but in the high nitrate treatment total nitrogen accumulation (including non-protein nitrogen) per plant was higher in the commercial soybean cultivar. This might be due to the selection of soybean individuals in high-input conditions, where higher inorganic fertilizer uptake rate would be directly or indirectly selected.

Given the nitrogen fixation capability of winged bean, a few attempts have been made to test its potential in agricultural systems. In an experiment conducted in Kentucky (37°N) intercropping at high planting density with maize led to lower grain yield for the latter, while increasing total biomass and N yield per hectare (Hikam et al. 1991, 1992). More encouraging results were shown in crop rotation experiments in pots, by comparing winged bean to common bean, cowpea (*Vigna unguiculata*), and mung bean (*V. radiata*). Winged bean was found to provide an equivalent of 4 to 8g/m^2 of N fertilizer, and to improve the most N-uptake, N-recovery efficiency, and N-agronomic efficiency compared to the other legumes. Overall, the yield of rice after winged bean was similar to the rice yield after fallowing with additional 8g/m^2 of N fertilizer (Rahman et al. 2014). Such results were further supported by studies on dry matter and nitrogen accumulation and partitioning in Sri Lanka (accession SLS40). With a planting density of 37000 plants/ha, the authors reported an estimated

202kg/ha of N available as green manure, after mature seeds were harvested (Weil and Belmont 1991). Perhaps inspired by these results, Anugroho and colleagues (2010) compared winged bean (*P. tetragonolobus* (L.) DC. cv Urizun) to velvet bean (*Mucuna pruriens*), as a common cover crop (Buckles 1995; Mupangwa et al. 2017). The authors found the first to have a significantly higher N concentration and uptake, and lower C/N ratio, meaning that winged bean could potentially be an optimal substitute to velvet bean as green manure (Anugroho et al. 2010).

3.5 Past Breeding

The first successful controlled crosses were conducted by Erskine & Khan (1977), using the crossing technique previously developed by Erskine & Bala (1976). The authors studied the inheritance of five qualitative traits based on two crosses from three PNG pure lines (UPS31, 4F1054, 4F1061). Within those crosses, purple colour was dominant over green for stem colour, calyx colour, pod wing, and pod specks. A single-gene mechanism was proposed for the inheritance of each of these traits, as well as for dominance of rectangular over flat pod shape. Moreover, between the genes controlling stem colour and calyx colour, and between pod wing colour and pod specks, they detected chromosomal linkage. It was not long before Erskine and colleagues carried out the first controlled crossing programme to study the inheritance of quantitative traits in winged bean, using a partial diallel cross design between five well-studied PNG accessions (UPS81, UPS66, UPS89, UPS121 and UPS122), one Indonesian accession from Java (UPS132), and a single plant selection from TPt1 (relabelled as UPS132) (Anonymous 1980; Erskine 1981a). F_1 plants along with their parents were evaluated at Port Moresby (PNG) in the wet season, and F_2 arrays with parents in the following dry season. Initial analysis focused on seed yield, which indicated significant additive and dominance gene effects, with positive heterosis over the mid-parent mean across the 21 F_1 families (averaging 25.6%) (Erskine 1981a). Pod-number per plant, rather than seed number per pod or mean seed weight, was the primary determinant of this heterosis. Narrow-sense heritability, determined from mid-parent/F_1 and from F_1/F_2 correlations, was significantly high (0.69 and 0.56 respectively) for seed yield. This suggested that significant gains for this trait would be possible, in that environment, through judicious choice of parent accessions followed by effective selection. Narrow-sense heritability calculated from inter-generational correlations for several characters contributing to the determination of seed yield (namely, leaf-area index; pod number per plant; pod length and seed number/pod; 100-seed weight) were also significant and high. Interestingly, in the diallel analysis of yield of green vegetable pods harvested every five days over nine weeks in a separate Port Moresby experiment with the same set of parents and their F_2 populations, Erskine and Kesavan (1982) found both heterosis above parental means and narrow-sense heritability to be non-significant, along with non-significant differences in number of pods per plant. The authors suggested the predominance of environmental and non-additive gene components in the determination of final green pod yield. On the other hand, there were large differences between populations in measured pod characters (length, width, weight), a result similar to that reported by Jalani and colleagues

(1983) in F_1s from a similar set of crosses. Improvements in vegetable pod yield could be achieved through indirect selection for these pod characteristics. However, the studies did not elucidate differences in some of the plant architecture traits (e.g. branch number, or raceme number per plant) that could contribute to differences in pod yield between populations and should be considered in the selection process as well.

For another important qualitative character, days to first flower (DAP), Erskine analysed the F_1 generation planted out in the long daylengths of December (13 hours photoperiod). The seven parental accessions ranged from 44 to 80 DAS, and narrow-sense heritability based on the correlation with their F_1 arrays was significant and high (0.68). However, in the evaluation of the F_2 generation, planted out in the short daylengths of the following July (12 hours photoperiod), flowering date among parents had a much smaller range from 46 to 51 DAS, and the narrow-sense heritability was zero. High environmental and genotype x environment effects for number of days to first flower was also encountered in a partial diallel involving a similar set of eight parental lines in the sub-tropical latitudes of Perth, Western Australia. There, parental means ranged widely from 109 to 194 DAP when planted on 30 October (in the increasing day lengths of mid-spring), but the narrow-sense heritability for days to flower estimated at 0.17, did not significantly differ from zero (Eagleton 1983). The data once again showed the variation in response to changing growing conditions among germplasm, and the effect of daylength on modulating days to first flower. Unfortunately, ambient factors air and soil temperature (see section 3.2) was not recorded, limiting the interpretation of this results.

Another very important determinant of agronomic performance and manageability of winged bean is plant growth habit. In a germplasm evaluation screen in Serdang (Malaysia), carried out on 147 accessions from several countries in the traditional winged bean growing centres, evaluated for forty-seven characters, the most important character apart from number of days to first flower that distinguished Highland Papua New Guinea accessions as a group from lowland South and Southeast Asian accessions was the number of primary branches (NPB) in the first ten nodes (Eagleton et al. 1985). The diallel analysis referred to above involving one-way crosses between eight representative parent accessions estimated narrow-sense heritability for NPB based on parent/F_1 array correlation in the Perth environment to be 0.65. Furthermore, the correlation between accession means determined in Malaysia and Perth for this character was 0.90, and for F_1 means it was 0.93, supporting

the hypothesis of minimal influence of environmental differences on the expression of this trait. The analysis indicated that apart from a significant additive genetic component for this character there are dominance effects, predominantly unidirectional, favouring high branch number over low branch number. It is possible that this character is determined by a few gene loci only.

Still in PNG, one preliminary study focused specifically at the genetic control of tuberisation in winged bean. Stephenson and colleagues crossed a tuberous PNG accession (UPS122) with another PNG accession (UPS31) which was found to be a poor producer of tubers (Anonymous, 1980; Stephenson et al., 1979). They compared the F_1, F_2, and first backcross generations (BC1 and BC2) in four replicated blocks at the lowland location of Waigani (Port Moresby, PNG), a sub-optimal location for tuber production. The plots were trellised and reproductively pruned to encourage tuber bulking up. Analysis indicated that additive genetic variance for root and tuber yield was greater than either dominance or environmental variance (i.e. among the four replicate blocks). Narrow-sense heritability of root and tuber yield, number of tubers per plant, and haulm yield were 0.45, 0.64, and 0.64 respectively. Haulm yield of the F_1 had exceeded that of the better parent UPS122, and across all plots highly correlated with root and tuber yield. To date, there has been no similar study published in the international literature concerning the genetic control of tuber yield in central Myanmar, another traditional location of winged bean tuber production.

The genetic mechanisms controlling nodulation and nitrogen fixation has received some attention from researchers as well. In crosses between TPt1 (IITA) and the relatively poor nodulator UPS122 (PNG), Iruthayathas and colleagues (1985) analysed the extent of host genetic determination of nodulation and nitrogen fixation, by partitioning variance for these two characters among parental, F_1, F_2 populations. Nodulation showed low broad-sense heritability (0.25), but there was no non-allelic interaction, and additive effects was the largest component of genetic variance. Nonetheless, the high environmental effect on variance would imply that genetic advance towards enhanced nodulation could be made in late generations of the progeny of this cross. Differently, the environmental effect was lower for nitrogen fixation (broad-sense heritability 0.48), and again most of it appeared of additive origin, suggesting substantial genetic advance could be attained in generations as early as F_2. Genetic studies of host resistance to certain diseases that can

limit winged bean have also been reported in the literature. Aminah Lubis and Sastrapradja (Anonymous 1981) produced evidence from hybridization studies indicating that resistance to the false rust disease *Synchytrium psophocarpi* in some accessions from West Java is conferred by two gene loci, at which alleles for susceptibility are dominant to alleles for resistance. Thompson and Haryono (1979) confirmed the presence of resistance to *S. psophocarpi* in the West Javanese accessions and also reported on the occurrence of resistance in Javanese germplasm to the disease yellow mosaic virus. Any controlled hybridisation for germplasm improvement in winged bean should take into account the outcome of controlled crosses carried out by Jalani and colleagues (1983). By analysing 6 crosses in a full diallel design, they found evidence of maternal and paternal effect within each cross on pod length, seed number per pod, and 100-seed weight.

3.6 The Big Friendly Genomics era

Initial analysis of genetic diversity and population structures relied on phenotypic traits known to be controlled by single genes (Erskine and Khan 1977). Analysing three such phenotypic markers, Erskine and Khan (1980) analysed the structure of the winged bean population upon samples collected across multiple locations in highland PNG. Eventually, they were able to support the idea of winged bean being a self-fertilising species, with poor gene-flow even between proximate locations. Subsequent studies, on much broader germplasm from wider geographical range, confirmed the basic picture of a cleistogamous species comprising landraces of predominantly inbred lines with a low level of cross-pollination but suggested a more complex population structure for the species as a whole. Papua New Guinea highland accessions on the one hand, Indonesian and Burmese accessions on the other, appear to be more highly domesticated off-shoots of a vegetable species that through most of its distribution is a useful, but low-maintenance, sprawling vegetable crop, subject to little purpose-driven selection pressure (Sastrapradja and Aminah Lubis 1975; Thompson and Haryono 1980; Anonymous 1981; Chandel et al. 1984). Such studies begun to evaluate winged bean diversity and population structure of germplasm collection based on phenotypic observations: knowledge on the genetic diversity is indeed important for selecting material to include in breeding programmes that is as diverse as possible. However, it is only with the advent of molecular technologies that researchers could start looking at polymorphisms in DNA sequences: independently from the subjective recording of a phenotype, and free from environmental effects. We had to wait till 2013 to have the first study of this kind in winged bean, when

Mohanty and colleagues (2013) analysed DNA of 24 accessions from at least four sources including India, Nigeria, and Thailand, through 13 Random Amplification of Polymorphic DNA (RAPD) and 7 Inter Simple Sequence Repeat (ISSR) molecular markers. They found ISSR amplification to be the more effective among the two approaches, but by combining the two sets of data they obtained a total of 167 clearly identifiable markers, of which 138 (83%) proved to be polymorphic across the range of the screened accessions. Cluster analysis applied to the cumulative data from the two sets of primers, grouped the 24 accessions into seven clusters that the authors judged not to be "in congruence with their geographic affiliations". Later, Chen and colleagues (2015) worked on a different set of 45 winged bean accessions from eight countries (PNG, Indonesia, Thailand, Nigeria, Colombia, Sri Lanka, Costa Rica, and China), using a set of five ISSR markers. These latter generated 67 identifiable marker fragments of which only 44 (65.7%) revealed polymorphism. Despite the lower degree of polymorphism compared to Mohanty and colleagues' findings, their four groups derived from cluster analysis led to the same conclusions: there was no consensus between genetic distance and geographical origin of the accessions.

A different approach to developing useful genomic markers for genetic studies is one which focusses on RNA rather than DNA nucleotide sequences. Genic SSR, for example, have been useful in genetic diversity analysis, as well as in linkage analysis for applying marker assisted selection (MAS) in breeding (Varshney et al. 2005). The first attempt in this direction resulted in the sequencing, assembly, and annotation of transcriptomes from four "orphan" (i.e. underutilised) crops, among which figured the winged bean accession TPt1, also identified as Ibadan Local-1 (Chapman 2015). The second transcriptomic analysis sequenced, annotated, and compared scored Simple Sequence Repeat (SSR) markers from two Sri Lankan accessions, one of which was collected in the field and the other maintained in the USDA seed bank since 1984 (Vatanparast et al. 2016). When comparing the transcripts obtained from these two Sri Lankan accessions, the authors found a high degree of similarity between the two [less than 200 Single Nucleotide Polymorphisms (SNPs)]. However, when they compared the combined SSR library of the Sri Lankan accessions with that of Chapman's TPt1, they identified 5,190 SNPs, showing a more distant genetic relationship. In 2017, two additional studies led to publication of transcriptomes. Singh and colleagues developed data from two accessions held in India differing for tannin contents. The expression analysis led to the isolation of candidate genes involved in phenylpropanoid pathway,

and the differential expression level were validated through real time-PCR (Singh et al. 2017). Later on, another study developed genic SSR markers from transcriptome data generated from a Malaysian accession. From this, the authors went on to validate 18 SSR markers upon nine genotypes originated from five countries (Malaysia, Bangladesh, Sri Lanka, Indonesia, and PNG) (Wong et al. 2017). Five of these markers were later used by Yang and colleagues to screen 53 accessions from IITA, U.S. Department of Agriculture (USDA), and National Agriculture and Food Research Organization (NARO, Japan) genebanks using ten SSRs in total, for genetic distance and population structure analysis. Principal Coordinates Analysis (PCoA) results indicated no clustering by country of origin, even for those with several accessions included, and the population structure suggested four subpopulations, each comprising of accessions from multiple countries, spanning in some case between South America, Asia, and Africa. Overall, these findings were in agreement with previous genetic analysis through ISSR and RAPD markers, in which molecular markers-based grouping did not agree with geographical origins (Mohanty et al. 2013; Chen et al. 2015). The authors, given also the suggested presence of admixture from clusters obtained through STRUCTURE analysis, suggested the possibility of crossing between genetic material coming from different origins, probably thanks to transport/sharing of seeds. Regarding these findings, two things are worth mentioning: the first one is the international trial organised by the ICDUP (see section 3), which did promote seed sharing, especially from South East Asian countries and PNG. Following on from this, the presence of accessions collected in South American countries (e.g. Costa Rica and Colombia), must be counted as suspicious, given that winged bean had never been an endogenous crop. Accessions deposited in international collections (e.g. USDA) and collected from Colombia, carry indeed the "UPS" label, the trademark of PNG material used in the past studies. The second source of bias in grouping by geographical origin could be the incorrect relabelling of the accessions used in studies. There are, indeed, clues spread across papers and reports about the true origin of accessions re-labelled by different institutes. An immediate, and rather important, example is the TPt1 accession deposited at the IITA genebank (Nigeria). Also known as Ibadan-Local 1 and usually referred to as from Nigeria, this may well have originated in PNG (see Erskine & Kesavan (1982) citing K. O. Rachie), and was also relabelled UPS132 by the same Erskine (1981).

From this standpoint, it might be necessary to understand genetic diversity analysis in winged bean, to allow the selection of material to phenotype. It would make sense to genotype deposited accessions and retrieved germplasm, without relying particularly on the declared geographical origin. In this context, high-throughput genotyping by sequencing (GbS) technologies could aid in obtaining a finer picture of the germplasm currently held, to later identify regions where to mine additional diversity or to select efficiently truely diverse material for phenotyping and breeding.

4 Winged bean: a perspective

Winged bean has great potential to contribute to food and nutritional security globally. Its ability to produce green pods, edible leaves, and tuberous roots makes it an ideal horticultural crop, and its mature seeds are a good source of protein and other nutrients making winged bean a potential pulse crop for large scale production in the humid and sub-humid tropics (NAS 1975b; Gross 1983; Kadam et al. 1984; Misra and Misra 1985; Lepcha et al. 2017). The nutritional profile of these edible parts could play a major role in diet improvement of people especially in areas where major vegetable crops require high-inputs, limiting consumers' access to vegetables. Furthermore, production of mature seeds could help reduce the demand and dependence on major crops such as soy bean for production of value-added products like tofu, and plant-based milk. In low-input systems, there is potential to grow winged bean on a relatively wide range of soils, because of its propensity to nodulate with different *Rhizobia* and its efficient nitrogen fixation activity. Additionally, winged bean's ability to store high concentrations of N in its biomass makes it possible to incorporate the crop into different cropping systems such as rotation intercropping and as a cover crop.

4.1 Photothermal sensitivity

For winged bean to fulfil its potential, several constraints, which limit its productivity and hence its wider adoption as a crop, must be tackled. The first is its photothermal sensitivity, given the effects that this has on its development and final yield, which also limits the potential expansion of winged bean cultivation areas. The combined effect of temperature and daylength on flowering is not unique to winged bean. In a well-known short-day legume, soybean, there are at least 10 loci controlling flower initiation, and for most of them, recessive alleles confer the observed degrees of photoperiod insensitivity (Xu et al. 2013b; Zhai et al. 2014; Weller and Ortega 2015; Zhang et al. 2016). Genetic pathways underlying mechanisms such as flowering are likely to be conserved across taxa (Hecht 2005; Kwak et al. 2008; Zhang et al. 2016), making it possible to harness information from a model species, such as soybean, and transfer it to a less-researched species, such as winged bean. An approach could be the development of linkage maps from mapping populations, followed by discovery of syntenic regions of the genomes of well characterised species, and identification of candidate genes for traits of interest. Such an approach has been implemented for identify flowering QTL in faba bean (*Vicia faba*) (Catt et al. 2017). The advantage of generating mapping populations would also be the concurrent

study and exploitation of the natural variation for the trait of interest. Luckily, such variation has been reported in winged bean, in particular between sets of germplasms coming from different regions, such as Papua New Guinea and South East Asia (Thompson and Haryono 1980; Erskine 1981b; Eagleton 1983).

4.2 Growth habit and plant architecture

Winged bean has intertwining vines that grow indeterminately, with a variable number of lateral branches (NAS 1975b; Khan and Erskine 1978; Eagleton 1983; Schiavinato and Válio 1996b). Until now, all published reports indicate that winged bean has a climbing growth habit, and naturally occurring determinate bush-type growth forms have not been observed in the species. A few investigations aimed at identify influences on plant structure have been published, by testing different physical supports and eventually concluding that trellis was needed in order to promote lateral branching and boost pod production (NAS 1975a; Anonymous 1978, 1981; Nangju and Baudoin 1979; Schiavinato and Válio 1996b). The number of lateral branches, as reported earlier (Eagleton et al. 1985), showed a significant degree of variation among germplasm from different areas, and could be a major target for obtaining desired plant architecture ideotypes in winged bean. However, there is still not a clear investigation focused on the growth, development, and inter-relationship among winged bean plant architecture traits (e.g. lateral branching, stem length, node number and internode length, leaves per plant), neither has there been research on how each of these influence yield-related traits (e.g. pods per plant).

In order to obtain dwarf or determinate winged bean cultivars, several researchers have attempted to use mutagenesis (Khan and Brock 1975; Jalani 1976; Jugran et al. 1986; Klu 1996; Quan et al. 2011; Chethan and Chikkadevaiah 2014). Unfortunately, from the published literature there is no clear evidence on whether such attempts have succeeded in producing either of the desired phenotypes.

Perhaps, TILLING (targeting induced local lesions in genomes) or ecoTILLING are possible approaches to obtain determinate growth phenotypes, although this method requires sequence information of target genes. The genomic basis of a determinate shoot apex meristem has been discovered in other leguminous crops; for example, soybean (Tian et al. 2010), chickpea (Hegde 2011), common bean (Kwak et al. 2012; Repinski et al. 2012), and cowpea (Dhanasekar and Reddy 2015), where the determinate shoot phenotypes are reported to be related to mutations in genes orthologous with the *A. thaliana TERMINAL FLOWER1* (*TFL1*) genes. Such

information could be used to identify candidate genes involved in the same mechanism in winged bean and being the target for TILLING methods. Another approach could be again the use of the conserved syntenic mapping approach for identification of QTL underlying morphological traits, as recently demonstrated in another underutilized species, bambara groundnut (*Vigna subterranea*) (Ho et al. 2017).

Alternatively, a more classical approach could focus on accessions grown predominantly for tuberous production, rather than for pods and seeds (NAS 1975b; Khan et al. 1977; Khan and Erskine 1978; Eagleton 1999), which could have undergone an initial selective pressure. Phenotypic screening for naturally occurring determinate growth habit could concentrate efforts on the winged bean tuber production fields of Central Myanmar, where crops are routinely grown without trellising, on a scale up to a hectare (Eagleton, personal communication). Exploration of these tuber growing areas for variation in growth habit and other useful variants is long overdue.

4.3 Mature seeds consumption

Mature winged bean seeds contain on average around 15-18% fat and 30-37% crude protein, and a favourable amino acid and micronutrient composition (Černý et al. 1971; Claydon 1975; Gillespie and Blagrove 1978; Garcia and Palmer 1980; Hildebrand et al. 1981b; Gross 1983; Kadam et al. 1984). Remarkably, tuberous root contains 13-20% protein (Anonymous 1980; Kortt and Caldwell 1984) and up to 25% in some case (Claydon 1975). Comprehensive assessments of the nutritional properties of winged bean seeds and the other edible parts have being adequately reviewed elsewhere (see for example Claydon 1975, 1983; Kailasapathy and Sandrasegaram 1983; Kadam et al. 1984; Kantha and Erdman 1984; Kailasapathy and Macneil 1985; Prakash et al. 1991; Lepcha et al. 2017). However, for mature seeds, two points should be mentioned. First, a hard coat that adheres tightly to the cotyledons could hamper cookability and digestion of winged bean mature seeds. Henry and colleagues studied cooking time across tropical legumes, where winged bean required almost double the time (220 minutes) for cooking compared to kidney beans (*P. vulgaris*) (time to texture judged acceptable by Sri Lankan housewives) (Henry et al. 1985). It must be pointed out, though, that the authors analysed only one cultivar ("Chimbu" from PNG) against legumes that have historically been used as pulses (e.g. pigeon pea, black lentils, kidney bean, soybean). Furthermore, soybean, a pulse widely consumed for its

seeds and a well-researched crop with improved varieties, took only 10 minutes less to be cooked in the same experiment (210 minutes) (Henry et al. 1985). Clearly, cooking traits and especially the hard-to-cook phenomenon in winged bean is an important aspect that effects consumer acceptability and is worth researching. Indeed, the development of easy-to-cook varieties of winged bean without compromising on its nutritional content should be one of the breeding objectives in any future winged bean breeding programme.

The second point concerns anti-nutritional factors in the winged bean seed coat, where decreased concentration remains a major target for improving net protein utilisation. A significant step forward in this direction is the identification of genes involved in tannin content (Singh et al. 2017), as well as identification of other anti-nutritional factors (Umemoto et al. 1992; Habu et al. 1996, 1997; Vatanparast et al. 2016). Cooking methods have also being investigated, with autoclaving or boiling being the most promising for seed processing and safe consumption (Černý and Addy 1973; Kadam et al. 1987). Screening for thinner seed coat might bring benefits, given the role it has in water uptake during soaking. Variation in water uptake during soaking, recorded across different accessions, could be harnessed for this purpose in breeding programmes on winged bean seed coat improvement (Anonymous 1980; Ekpenyong and Borchers 1980; Henry et al. 1985; Deshpande and Cheryan 1986; Sambudi and Buckle 1991).

4.4 Future directions

Despite the latest phylogenetic analysis by Yang and colleagues (2018), the centre of origin of winged bean remains a matter that requires further investigations. Research efforts in this direction, based on larger collections of the cultivated *P. tetragonolobus* and related *Psophocarpus* species, could yield a clearer picture of the taxonomy, origin and evolution of winged bean, and allow for a strategic exploitation of the resources for future improvement. Such analysis ought to be carried out on germplasm (or subsets of) currently held in different countries and institutes. The declared geographical origin for each accession should be treated carefully, looking at codes such as “UPS” (usually from PNG) or “SLS” (usually from Sri Lanka), and taking it as indication, rather than certainty.

Renovated efforts could mirror, to an extent, what the dedicated international centre did back during the 70s and 80s, by creating a panel of common accessions for phenotypic screening, and for development of populations with shared molecular information. Such steps are critical also in the development of mapping populations, linkage mapping, and Genome Wide Association Studies (GWAS) for trait discovery, as shown in soybean (Li et al. 2018), chickpea (*Cicer arientinum*)(Bajaj et al. 2016; Saeed and Darvishzadeh 2017), mung bean (Noble et al. 2018), and common bean (Tock et al. 2017). This would allow us also to combine phenotypic and genetic data across different environments and assess the degree of adaptability and plasticity of winged bean. Photoperiod sensitivity is likely to demand the most out of these combined efforts, especially for the required screening of natural or mutated individuals and the identification of the molecular basis of insensitive phenotypes. For all kind of screenings, an important component would be the development of common descriptors in order to phenotype the analysed germplasm, perhaps starting by updating those available at https://www.bioversityinternational.org/e-library/publications/detail/winged-bean-descriptors-revised/.

Winged bean, as for any other crop, is grown in multiple contexts and its growth and development have implications on farming practices and costs. It is therefore natural to call for the design of ideotypes that fit into such contexts, through studies that clarify the effect of the plant architecture on yield, and the best trade-offs between them (see Figure 1 for potential breeding targets). To stress this concept, winged bean can be cultivated on a variety of supporting structures (e.g. strings, trellises, single stalk), as well as on the ground without any support. Any desired alteration to the winged bean plant architecture and its potential effects on developmental processes would depend on the desired product or final farming purposes (for example for pods, seeds, tubers production) and the distinction between supported and unsupported cultivation methods. It should also be noted that individuals with longer growing cycles and extended flowering period are desirable in certain conditions (e.g. as a cover crop where biomass is the objective) and so are early maturing types, e.g. for seed production. Moreover, an extended flowering period might be preferred by small-scale farmers and for domestic use, as this ensures a continuous source of food and products to sell in local markets over an extended period. Post-harvest research could have a role here, as well as in preserving green pods and tubers for larger

distribution areas, but unfortunately no research has been conducted so far on this. Figure 1 presents possible breeding targets for winged bean.

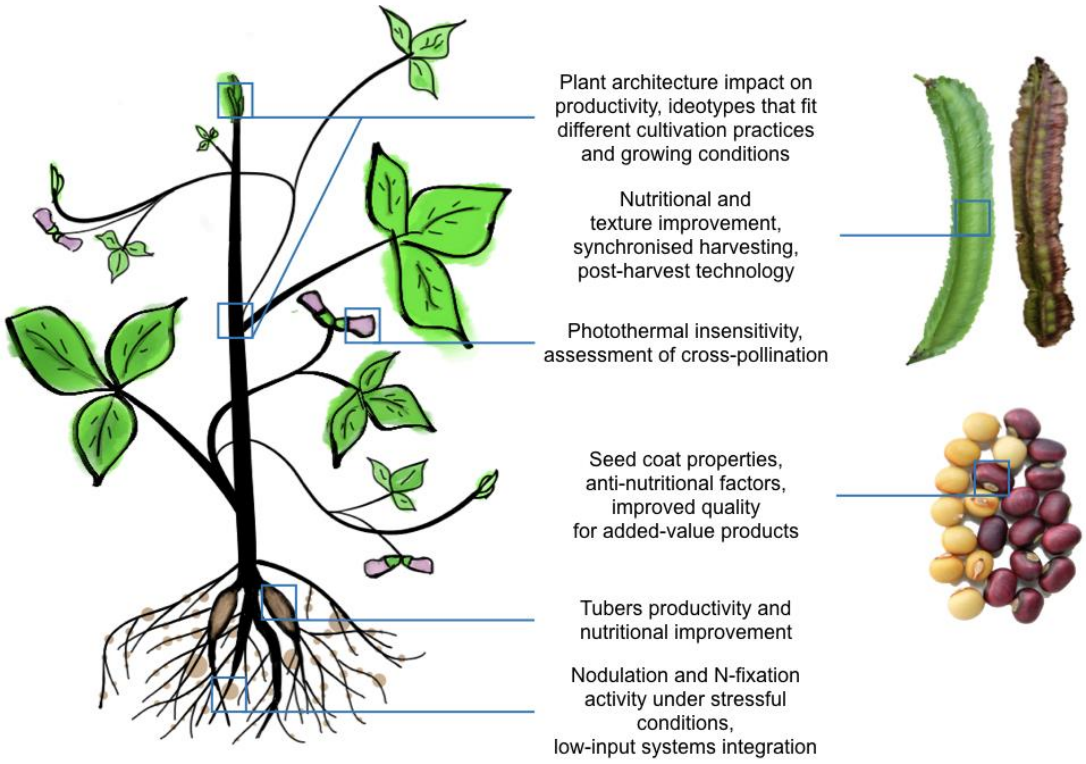


Figure 1: Possible breeding targets for winged bean improvement programmes.

Perhaps the least explored area in the past research is the ability of winged bean to tolerate abiotic stresses, like drought, salinity, or waterlogging. After his observations Karikari suggested a list of characters for potential identification of drought tolerant individuals in winged bean, which included, among others, number and weight of nodules per plant and stem thickness (Karikari 1978). This might be useful to have a set of traits for initial screening aimed to identify more drought-tolerant material. Another study investigated the effect of salt stress on nitrogenase activity, which showed activity to be significantly more hampered in winged bean than in soybean, although winged bean had higher plant N, and less foliage injury (Weil and Khalil 1986). The authors concluded that winged bean, here represented by just one accession commonly grown in Sri Lanka (SLS47), was similar to the soybean cv. "Lee", considered to be one of the most salt tolerant soybean cultivar and used, among others, for the discovery of the *Ncl* gene involved in this trait (Weil and Khalil 1986; Do et al. 2016). Such findings would encourage further research to explore salt tolerance attributes of winged bean. In addition to drought-tolerance and photothermal sensitivity, the ability of winged bean to withstand abiotic

stresses will not only determine the extent of adoption beyond its current areas of cultivation but will also determine its position in low-input agricultural systems.

Lessons learned from past research experiences are a key component of the toolkit required to develop winged bean into 'a one-stalk supermarket'. Although the plurality of winged bean edible parts offers the opportunities for breeders to develop this crop in different ways for commercial purposes, research should not lose sight of the traditional modes of consumption. Winged bean is more likely to be a horticultural crop first: immature pods, tubers, green seeds, and leaves should be the main target for increased productivity and nutritional quality. Mature seeds could gain more value if improved and introduced into pipelines to produce value-added products. Food technologist could adapt what is already present for crops like soybean, to generate alternative products with different nutritional profiles and taste, like tofu, tempeh, and milk (Omachi et al. 1983; Kantha et al. 1983; Prakash et al. 1987; Jinsong Yang and Haisheng Tan 2011). Creation of cultivars that could bring high quality pods and seed productivity, should both benefit from decreasing fibre in the pod walls: these, in fact, could make pods harder to eat, and contribute to seed dispersal through pod shattering (Koinange et al. 1996).

Unfortunately, no relevant information is available about national production of winged bean. Being mainly cultivated on small scale, in home gardens, and sold at local markets, makes it difficult to have an actual idea of the extent of cultivated area, production, and market price, an issue in common with other minor crop species (Khoury et al. 2014). This lack of information makes it harder to include crops like winged bean into a more comprehensive assessment of the diversity of national food production systems, and to advocate for policies and research funding schemes to improve such sources of food and income. We hope that, as for other crops, data will be collected for winged bean, where its presence in local markets is a significant source of food and income.

Finally, for the development of underutilised crops such as winged bean, past research is critical and yet difficult to access first hand. For this reason, Crops For the Future (CFF) is establishing wingedbeannetwork.org, to build upon the work started with wingedbeanflyer.org by Graham Eagleton, as a

platform to share digitalised copies of past papers, and to engage with farmers, researchers, and private companies. The aim is to share research results, and to promote collaborations among the different players involved in the research and cultivation of winged bean.

5 Conclusions

The United Nations, through the declaration of the Sustainable Development Goals, have called for an end to hunger, to achieve food security, to improve nutrition, and to promote sustainable agriculture (UN, 2015). There is growing data supporting the advantages of including more crops into our agricultural systems, to achieve food security. A case is made for vegetables and pulses to improve, above all, human nutrition. Winged bean, can be grown as both, given its edible green pods, immature seeds, tuberous roots, leaves, and mature seeds. This underutilised crop has potentially beneficial traits related to its nodulation and nitrogen fixation activity that could support low-input farming and improve soil conditions. This review has explored past findings related to the controlling mechanisms underlying traits like growth, flowering, and yield of the different edible parts of winged bean. There are still several production and plant architectural constraints that remain unresolved since then, including the indeterminate growth habit, photothermal sensitivity, inconsistent productivity for edible parts, anti-nutritional content of seeds, and the beneficial effects of winged bean nodulation. Some guidelines, in order to have a fresh approach and resolve these constraints, have been suggested.

Improvement programmes should keep in mind the conditions where winged bean is grown, with regards to cultivation methods and ways of consumption. Research efforts can prioritise the development of high yielding individuals, the design of ideotypes for pods and tuberous root production, the improvement of seed and pod nutritional values, and their quality for the market. It is strongly recommended to have a panel of winged bean accessions shared among research institutes, for generating information that can be available and shared across them. Regarding this, it is crucial first to keep track of the material shared or retrieved from genebanks, with a labelling system consistent across research groups; second, to utilise a common set of descriptors and procedures to phenotype the germplasm. Starting from this, molecular tools (GBS, RNA-Seq, GWAS, and genome sequencing) can bring forward the efforts to tackle critical traits. In such regard, the generation of

winged bean mapping populations, linkage map, and sequenced genome would boost the translation of information already developed in other crops and the dissection of traits of interest in winged bean.

Acknowledgements

The authors would like to thank Crops For the Future (CFF) and the University of Nottingham Malaysia Campus (UNMC) for financial support. The authors wish to dedicate this paper to the memory of Quin Nee Wong.

References

Adhikari L, Hussain A, Rasul G (2017) Tapping the potential of neglected and underutilized food crops for sustainable nutrition security in the mountains of Pakistan and Nepal. Sustain 9:. doi: 10.3390/su9020291

Anonymous (1980) The Winged Bean Flyer (3 No.1). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Anonymous (1981) The Winged Bean Flyer (3 No.2). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Anonymous (1982) The Winged Bean Flyer (4 No.1). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Anonymous (1978) The winged bean - Papers Presented at the 1st International Symposium on Developing the Potentials of the Winged Bean (Manila, Philippines). Philippine Council for Agriculture and Resources Research, Los Banos, Laguna (Philippines)

Anugroho F, Kitou M, Kinjo K, Kobashigawa N (2010) Growth and Nutrient Accumulation of Winged Bean and Velvet Bean as Cover Crops in a Subtropical Region. Plant Prod Sci 13:360–366. doi: 10.1626/pps.13.360

Arumugan S, Perera ALT (1989) A study of the genetic architecture of some yield parameters of winged bean using model-fitting techniques. Trop Agric Res 1:

Bajaj D, Srivastava R, Nath M, et al (2016) EcoTILLING-Based Association Mapping Efficiently Delineates Functionally Relevant Natural Allelic Variants of Candidate Genes Governing Agronomic Traits in Chickpea. Front Plant Sci 7:450. doi: 10.3389/fpls.2016.00450

Bazile D, Pulvento C, Verniau A, et al (2016) Worldwide Evaluations of Quinoa: Preliminary Results from Post International Year of Quinoa FAO Projects in Nine Countries. Front Plant Sci 7:850. doi: 10.3389/fpls.2016.00850

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc London Ser B Biol Sci 274:227–274

Buckles D (1995) Velvetbean: A "new" plant with a history. Econ Bot 49:13–25. doi: 10.1007/BF02862271

Burkill IH (1906) Goa beans in India. Agric Ledger 4:101–114

Catt SC, Braich S, Kaur S, Paull JG (2017) QTL detection for flowering time in faba bean and the responses to ambient temperature and photoperiod. Euphytica 213:125. doi: 10.1007/s10681-017-1910-8

Černý K, Addy HA (1973) The winged bean (*Psophocarpus palustris* Desv.) in the treatment of kwashiorkor. Br J Nutr 29:105. doi: 10.1079/BJN19730082

Černý K, Kordylas M, Pospíšil F, et al (1971) Nutritive value of the winged bean (*Psophocarpus palustris* Desv.). Br J Nutr 26:293. doi: 10.1079/BJN19710035

Chandel KPS, Pant KC, Arora RK (1984) Winged Bean in India. NBPGR Sci. Monogr. No. 8. National Bureau of Plant Genetics Resources, New Delhi, India

Chapman MA (2015) Transcriptome Sequencing and Marker Development for Four Underutilized Legumes. Appl Plant Sci 3:1400111. doi: 10.3732/apps.1400111

Chappell MJ, LaValle LA (2011) Food security and biodiversity: Can we have both? An agroecological analysis. Agric Human Values 28:3–26. doi: 10.1007/s10460-009-9251-4

Chen D, Yi X, Yang H, et al (2015) Genetic diversity evaluation of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) using inter-simple sequence repeat (ISSR). Genet Resour Crop Evol 62:823–828. doi: 10.1007/s10722-015-0261-3

Chethan KG, Chikkadevaiah (2014) The effect of gamma rays on yield and yield attributes in winged bean (Psophocarpus tetragonolobus (L)DC) genotypes. Bioinfolet 11:953–956

Chivenge P, Mabhaudhi T, Modi AT, Mafongoya P (2015) The potential role of neglected and underutilised crop species as future crops under water scarce conditions in Sub-Saharan Africa. Int J Environ Res Public Health 12:5685–5711. doi: 10.3390/ijerph120605685

Claydon A (1975) A review of the nutritional value of the winged bean *Psophocarpus tetragonolobus* (L.) DC. with special reference to Papua New Guinea. Sci New Guinea 3:103–114

Claydon A (1983) Potential of winged bean pods and their products in Papua New Guinea. Qual Plant Plant Foods Hum Nutr 32:167–177. doi: 10.1007/BF01091337
Deshpande SS, Cheryan M (1986) Microstructure and Water Uptake of Phaseolus and Winged Beans. J Food Sci 51:1218–1223. doi: 10.1111/j.1365-2621.1986.tb13089.x
Dhanasekar P, Reddy KS (2015) A novel mutation in *TFL1* homolog affecting determinacy in cowpea (Vigna unguiculata). Mol Genet Genomics 290:55–65. doi: 10.1007/s00438-014-0899-0
Do TD, Chen H, Hien VTT, et al (2016) *Ncl* Synchronously Regulates Na+, K+, and Cl-in Soybean and Greatly Increases the Grain Yield in Saline Field Conditions. Sci Rep 6:1–10. doi: 10.1038/srep19147
Drinkall MJ (1978) False Rust Disease of the Winged Bean. Trop Pest Manag 24:160–166. doi: 10.1080/09670877809411605
Dwivedi SL, Lammerts van Bueren ET, Ceccarelli S, et al (2017) Diversifying Food Systems in the Pursuit of Sustainable Food Production and Healthy Diets. Trends Plant Sci 22:842–856. doi: 10.1016/j.tplants.2017.06.011
Eagleton G (1999) Winged Bean in Myanmar, Revisted. Econ Bot 53:342–352
Eagleton GE (1983) Evaluation of genetic resources in the winged bean (*Psophocarpus tetragonolobus* (L.) DC.) and their utilisation in the development of cultivars for higher latitudes. University of Western Australia
Eagleton GE, Khan TN, Erskine W (1985) Winged bean (*Psophocarpus tetragonolobus* (L.) DC. In: Summerfield RJ, J. RE (eds) Grain Legume Crops. Collins, London, pp 624–657
Eagleton GE, Sandover S (1984) Adaptation of winged bean to Kununurra clays. Dry Season 1984. Western Australia Department of Agriculture (Kununurra Regional Office)
Ekpenyong TE, Borchers RL (1980) Effect of cooking on the chemical composition of winged beans (*Psophocarpus tetragonolobus*). J Food Sci 45:1559–1560. doi: 10.1111/j.1365-2621.1980.tb07562.x
Elmes RPT (1976) Cross-inoculation relationships of *Psophocarpus tetragonolubus* and its *rhizobium* with other legumes and rhizobia. Papua New Guinea Agric J 27:53–57
Erskine W (1981a) Inheritance of some economic characters in winged bean. Paper presented at the 2nd international seminar of winged bean. Colombo, Sri Lanka
Erskine W (1981b) Heritability and combining ability of vegetative and phenological characters of winged beans (*Psophocarpus tetragonolobus* (L.) D.C.). J Agric Sci 96:503. doi: 10.1017/S0021859600034456
Erskine W (1980) Measurements of the cross-pollination of winged bean in Papua New Guinea. SABRAO J. 12:11–14
Erskine W, Bala AA (1976) Crossing technique in winged bean. Trop. Grain Legum. Bullettine 6:32–35
Erskine W, Kesavan V (1982) Genetic Variability in the Green Pod Production of Winged Bean. J Hortic Sci 57:209–213. doi: 10.1080/00221589.1982.11515042
Erskine W, Khan TN (1977) Inheritance of pigmentation and pod shape in winged bean. Euphytica 26:829–831. doi: 10.1007/BF00021714
FAO (2010) The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome: Food and Agriculture Organization of the United Nations
FAO (2015) Coping with climate change – the roles of genetic resources for food and agriculture. Rome: Food and Agriculture Organization of the United Nations
FAO (2011) The State of Food Insecurity: How Does International Price Volatility Affect Domestic Economies and Food Security? Rome: Food and Agriculture Organization of the United Nations
Fatihah HNN, Maxted N, Rico Arce L (2012) Cladistic analysis of *Psophocarpus* Neck. ex DC. (Leguminosae, Papilionoideae) based on morphological characters. South African J Bot 83:78–88. doi: 10.1016/j.sajb.2012.07.010
Garcia V V., Palmer JK (1980) Proximate analysis of five varieties of winged beans, *Psophocarpus tetragonolobus* (L.) DC. Int J Food Sci Technol 15:469–476. doi: 10.1111/j.1365-2621.1980.tb00965.x
Gillespie J, Blagrove R (1978) Isolation and Composition of the Seed Globulins of Winged Bean, *Psophocarpus tetragonolobus* (L.) DC. Aust J Plant Physiol 5:357. doi: 10.1071/PP9780357
Gross R (1983) Composition and protein quality of winged bean (*Psophocarpus tetragonolobus*). Qual Plant Plant Foods Hum Nutr 32:117–124. doi: 10.1007/BF01091332
Habu Y, Fukushima H, Sakata Y, et al (1996) A gene encoding a major Kunitz proteinase inhibitor of storage organs of winged bean is also expressed in the phloem of stems. Plant Mol Biol. doi: 10.1007/BF00041408
Habu Y, Peyachoknagul S, Sakata Y, et al (1997) Evolution of a multigene family that encodes the Kunitz chymotrypsin inhibitor in winged bean: A possible intermediate in the generation of a new gene with a distinct pattern of expression. Mol Gen Genet. doi: 10.1007/s004380050392
Habu Y, Peyachoknagul S, Umemoto K, et al (1992) Structure and Regulated Expression of Kunitz Chymotrypsin Inhibitor Genes in Winged Bean [*Psophocarpus tetragonolobus* (L.) DC.]. J Biochem 111:249–258. doi: 10.1093/oxfordjournals.jbchem.a123745
Harder DK (1992) Chromosome Counts in *Psophocarpus*. Kew Bull 47:529. doi: 10.2307/4110581
Harder DK, Smartt J (1992) Further evidence on the origin of the cultivated winged bean, *Psophocarpus tetragonolobus* (L.) DC. (Fabaceae): Chromosome numbers and the presence of a host-specific fungus. Econ Bot 46:187–191. doi: 10.1007/BF02930637

Harding J, Lugo Lopez. M. A., Perez Escolar R (1978) Promiscuous root nodulation of winged beans on an oxisol in Puerto Rico (*Psophocarpus tetragonolobus*). Trop. Agric.
Hecht V (2005) Conservation of Arabidopsis Flowering Genes in Model Legumes. Plant Physiol 137:1420–1434. doi: 10.1104/pp.104.057018
Hegde VS (2011) Morphology and genetics of a new found determinate genotype in chickpea. Euphytica 182:35–42. doi: 10.1007/s10681-011-0447-5
Henry CJK, Donachie PA, Rivers JPW (1985) The winged bean. Will the wonder crop be another flop? Ecol Food Nutr 16:331–338. doi: 10.1080/03670244.1985.9990872
Herath HMW, Ormrod DP (1979) Effects of Temperature and Photoperiod on Winged Beans [*Psophocarpus tetragonolobus* (L.) D.C.]. Ann Bot 43:729–736. doi: 10.1093/oxfordjournals.aob.a085686
Hikam S, MacKown CT, Poneleit CG, Hildebrand DF (1991) Growth and N Accumulation in Maize and Winged Bean as Affected by N Level and Intercropping. Ann Bot 68:17–22
Hikam S, Poneleit CG, MacKown CT, Hildebrand DF (1992) Intercropping of Maize and Winged Bean. Crop Sci 32:195. doi: 10.2135/cropsci1992.0011183X003200010040x
Hildebrand DF, Blevins DG, Hymowitz T (1981a) Cross-inoculation specificity of *Psophocarpus tetragonolobus* (L.) DC. Plant Soil 60:139–142. doi: 10.1007/BF02377119
Hildebrand DF, Chaven G, Hymowitz T, et al (1981b) Variation in protein and oil content of winged bean seeds as measured by near infrared light reflectance. Agron J 73:623–625
Hildebrand DF, Harper JE, Hymowitz T (1981c) Effects of nitrate level on nitrogen metabolism in winged bean and soya bean. Ann Bot 48:307–313. doi: 10.1093/oxfordjournals.aob.a086128
Ho WK, Chai HH, Kendabie P, et al (2017) Integrating genetic maps in bambara groundnut [*Vigna subterranea* (L) Verdc.] and their syntenic relationships among closely related legumes. BMC Genomics 18:1–9. doi: 10.1186/s12864-016-3393-8
Hungria M, Joseph CM, Phillips DA (1991) Rhizobium nod Gene Inducers Exuded Naturally from Roots of Common Bean (*Phaseolus vulgaris* L.). Plant Physiol 97:759–764. doi: 10.1104/pp.97.2.759
Hungria M, Phillips DA (1993) Effects of a Seed Color Mutation on Rhizobial nod -Gene-Inducing Flavonoids and Nodulation in Common Bean. Mol Plant-Microbe Interact 6:418–422. doi: 10.1094/MPMI-6-418
Hymowitz T, Boyd J (1977) Origin, ethnobotany and agricultural potential of the winged bean-*Psophocarpus tetragonolobus*. Econ Bot 31:180–188. doi: 10.1007/BF02866589
Ikram A, Broughton WJ (1980) Rhizobia in tropical legumes-VII. Effectiveness of different isolates on *Psophocarpus tetragonolobus* (L.) DC. Soil Biol Biochem 12:77–82. doi: 10.1016/0038-0717(80)90105-4
Iruthayathas EE, Herath HMW (1981) Nodule formation and distribution during the establishment stage of six selections of winged bean. Sci Hortic (Amsterdam) 15:1–8. doi: 10.1016/0304-4238(81)90055-8
Iruthayathas EE, Vlassak K (1982) Symbiotic specificity in nodulation and nitrogen fixation between winged bean and *Rhizobium*. Sci Hortic (Amsterdam) 16:313–322. doi: 10.1016/0304-4238(82)90030-9
Iruthayathas EE, Vlassak K (1987) Promiscuity of some legume species with winged bean *Rhizobium* strains. Sci Hortic (Amsterdam) 31:1–9. doi: 10.1016/0304-4238(87)90100-2
Iruthayathas EE, Vlassak K, Laeremans R (1985) Inheritance of nodulation and N2 fixation in winged beans. J Hered 76:237–242. doi: 10.1093/oxfordjournals.jhered.a110084
Jalani BS (1976) A preliminary study of the effects of gamma- irradiation on the growth of winged bean *Psophocarpus tetragonolobus* (L.) DC. Malaysian Agric Res 5:139–144
Jalani BS, Benong M, Ismail N (1983) Evaluation of F1 hybrids on winged bean (*Psophocarpus tetragonolobus*). 4–8
Jinsong Yang, Haisheng Tan (2011) Study on winged bean milk. In: 2011 International Conference on New Technology of Agricultural. IEEE, pp 814–817
Jugran HMM, Nath P, Banerji BKK, Datta SKK (1986) Gamma ray induced dwarf mutant of winged bean. J Nucl Agric Biol 15:175–178
Kadam SS, Salunkhe DK, Luh BS (1984) Winged bean in human nutrition. C R C Crit Rev Food Sci Nutr 21:1–40. doi: 10.1080/10408398409527395
Kadam SS, Smithard RR, Eyre MD, Armstrong DG (1987) Effects of heat treatments of antinutritional factors and quality of proteins in winged bean. J Sci Food Agric 39:267–275. doi: 10.1002/jsfa.2740390310
Kailasapathy K, Macneil JH (1985) Baking Studies with Winged Bean (*Psophocarpus tetragonolobus* L.DC) Flour-Wheat Flour Blends. J Food Sci 50:1672–1675. doi: 10.1111/j.1365-2621.1985.tb10562.x
Kailasapathy K, Sandrasegaram C (1983) Preliminary studies on winged bean leaf protein concentrates. In: Sri Lanka Assoc Adv Sci. pp 129–146
Kantha SS, Erdman JW (1984) The winged bean as an oil and protein source: A review. J Am Oil Chem Soc 61:515–525. doi: 10.1007/BF02677021
Kantha SS, Erdman JW (1986) Winged Bean as a Source of Protein: Recent Advances. pp 206–219
Kantha SS, Hettiarachchy NS, Erdman JW (1983) Laboratory Scale Production of Winged Bean Curd. J Food Sci 48:441–444. doi: 10.1111/j.1365-2621.1983.tb10761.x
Karikari SK (1972) Pollination requirements of winged beans (*Psophocarpus* spp. Neck) in Ghana. Ghana J Agric Sci

5:235–239
Karikari SK (1978) Characters in the selection of Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) tolerant to drought. Kade
Khan TN (1976) Papua New Guinea: A centre of genetic diversity in winged bean (*Psophocarpus tetragonologus* (L.) D c.). Euphytica 25:693–705. doi: 10.1007/BF00041608
Khan TN, Bohn JC, Stephenson RA (1977) Winged beans cultivation in Papua New Guinea. World Crop. Livest.
Khan TN, Brock RD (1975) Mutation breeding in winged bean. In "Regional Seminar on the Use of Induced Mutation in Improvement in Grain Legume Production in South East Asia." 8–12
Khan TN, Erskine W (1978) The Adaptation of Winged Bean (*Psophocarpus tetragonolobus* (L.) DC.) in Papua New Guinea. Aust J Agric Res 281–289
Khoury CK, Bjorkman AD, Dempewolf H, et al (2014) Increasing homogeneity in global food supplies and the implications for food security. Proc Natl Acad Sci 111:4001–4006. doi: 10.1073/pnas.1313490111
Klu GYP (1996) Efforts to accelerate domestication of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) by means of induced mutations and tissue culture. Wageningen Agricultural University
Klu GYP, Kumaga FK (1999) Testing of induced mutants of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for nodulation and phenotypic performance. Ghana J Sci 39:55–62
Koinange EMK, Singh SP, Gepts P (1996) Genetic Control of the Domestication Syndrome in Common Bean. Crop Sci 36:1037. doi: 10.2135/cropsci1996.0011183X003600040037x
Kortt AA (1983) Comparative studies on the storage proteins and anti-nutritional factors from seeds of *Psophocarpus tetragonolobus* (L.) DC. from five South-East Asian countries. Qual Plant Plant Foods Hum Nutr 33:29–40. doi: 10.1007/BF01093735
Kortt AA (1980) Isolation and properties of a chymotrypsin inhibitor from winged bean seed (*Psophocarpus tetragonolobus* (L) Dc.). Biochim Biophys Acta - Protein Struct 624:237–248. doi: 10.1016/0005-2795(80)90243-3
Kortt AA, Caldwell JB (1984) Characteristics of the proteins of the tubers of winged bean (*Psophocarpus tetragonolobus* (L.) DC). J Sci Food Agric 35:304–313. doi: 10.1002/jsfa.2740350310
Koshy EP, Alex BK, John P (2013a) Pollen Viability Studies in *Psophocarpus tetragonolobus* (L.) DC. Int J Appl Nat Sci 2:65–68
Koshy EP, Alex BK, John P (2013b) Clonal fidelity studies on regenerants of *Psophocarpus tetragonolobus* (L.) DC. using RAPD markers. The Bioscan 8:763–766
Kwak M, Toro O, Debouck DG, Gepts P (2012) Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). Ann Bot 110:1573–1580. doi: 10.1093/aob/mcs207
Kwak M, Velasco D, Gepts P (2008) Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). J Hered 99:283–291. doi: 10.1093/jhered/esn005
Lenz F, Broughton WJ (1981) Growth, photosynthesis and transpiration in *Psophocarpus tetragonolobus* (L.) DC. cultivar UPS 99. Photosynth Res 2:259–268. doi: 10.1007/BF00056263
Lepcha P, Egan AN, Doyle JJ, Sathyanarayana N (2017) A Review on Current Status and Future Prospects of Winged Bean (*Psophocarpus tetragonolobus*) in Tropical Agriculture. Plant Foods Hum Nutr 72:225–235. doi: 10.1007/s11130-017-0627-0
Li D, Zhao X, Han Y, et al (2018) Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. Genomics. doi: 10.1016/J.YGENO.2018.01.004
Mandal SM, Chakraborty D, Dey S (2010) Phenolic acids act as signaling molecules in plant-microbe symbioses. Plant Signal Behav 5:359–368. doi: 10.4161/psb.5.4.10871
Manen JF, Simon P, Slooten JC Van, et al (1991) A Nodulin Specifically Expressed in Senescent Nodules of Winged Bean Is a Protease Inhibitor. Plant Cell 3:259–270. doi: 10.1105/tpc.3.3.259
Masefield GB (1957) The nodulation of annual leguminous crops in Malaya. Emp J Exp Agric 25:
Masefield GB (1961) Root nodulation and agricultural potential of the leguminous genus *Psophocarpus*. Trop Agric (Trinidad Tobago) 38:225–229
Massawe F, Mayes S, Cheng A (2016) Crop Diversity: An Unexploited Treasure Trove for Food Security. Trends Plant Sci 21:365–368. doi: 10.1016/j.tplants.2016.02.006
Maxted N (1990) A phenetic investigation of *Psophocarpus* Neck. ex DC. (Leguminosae-Phaseoleae). Bot. J. Linn. Soc. 102:103–122
Mayes S, Massawe FJ, Alderson PG, et al (2012) The potential for underutilized crops to improve security of food production. J Exp Bot 63:1075–1079. doi: 10.1093/jxb/err396
Misra PN, Misra G (1985) Studies on the adaptability of winged bean (*Psophocarpus tetragonolobus* (L.) DC) to marginal soil conditions under semi-arid sub tropics. Qual Plant Plant Foods Hum Nutr 35:145–151. doi: 10.1007/BF01092130
Mnembuka B V., Eggum BO (1995) Comparative nutritive value of winged bean (*Psophocarpus tetragonolobus* (L) DC) and other legumes grown in Tanzania. Plant Foods Hum Nutr 47:333–339. doi: 10.1007/BF01088271
Mohanty CS, Verma S, Singh V, et al (2013) Characterization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.)

based on molecular , chemical and physiological parameters. Am J Mol Biol 3:187–197. doi: 10.4236/ajmb.2013.34025
Mukhopadhyay D (2000) The molecular evolutionary history of a winged bean α-chymotrypsin inhibitor and modeling of its mutations through structural analyses. J Mol Evol. doi: 10.1007/s002399910024
Mupangwa W, Thierfelder C, Ngwira A (2017) Fertilization strategies in conservation agriculture systems with maize-legume cover crop rotations in Southern Africa. Exp Agric 53:288–307. doi: 10.1017/S0014479716000387
Nangju D, Baudoin JP (1979) Performance of Winged Bean (*Psophocarpus Tetragonolobus* (L.) DC.) in Nigeria. J Hortic Sci 54:129–136. doi: 10.1080/00221589.1979.11514860
NAS (1975a) Underexploited Tropical Plants with Promising Economic Value. Washington, D.C.
NAS (1975b) The Winged Bean: A High-Protein Crop for the Humid Tropics. National Academy of Sciences, Washington, D.C.
NAS (1981) The Winged Bean: A High-Protein Crop for the Humid Tropics, 2nd edn. National Academy of Sciences, Washington, D.C.
Noble TJ, Tao Y, Mace ES, et al (2018) Characterization of Linkage Disequilibrium and Population Structure in a Mungbean Diversity Panel. Front Plant Sci 8:2102. doi: 10.3389/fpls.2017.02102
Okezie BO, Martin FW (1980) Chemical composition of dry seeds and fresh leaves of winged bean varieties gronw in the U.S. and Puerto Rico. J Food Sci 45:1045–1051. doi: 10.1111/j.1365-2621.1980.tb07509.x
Okubo H, Masunaga T, Yamashita H, Uemoto S (1992) Effects of photoperiod and temperature on tuberous root formation in winged bean (*Psophocarpus tetragonolobus*). Sci Hortic (Amsterdam) 49:1–8. doi: 10.1016/0304-4238(92)90137-2
Omachi M, Ishak E, Homma S, Fujimaki M (1983) Manufacturing of tofu [bean curd] from winged bean [*Psophocarpus tetragonolobus*] and the effect of presoaking and cooking. J Japanese Soc Food Sci Technol 30:
Padulosi S, Thompson J, Rudebjer P (2013) Fighting poverty, hunger and malnutrition with neglected and underutilized species: needs, challenges and the way forward. Biodiversity International, Rome
Pate JS, Atkins CA, White ST, et al (1980) Nitrogen Nutrition and Xylem Transport of Nitrogen in Ureide-producing Grain Legumes. Plant Physiol 65:961–5
Pellegrini L, Tasciotti L (2014) Crop diversification, dietary diversity and agricultural income: Empirical evidence from eight developing countries. Can J Dev Stud 35:211–227. doi: 10.1080/02255189.2014.898580
Pickersgill B (1980) Cytology of two species of winged bean, *Psophocarpus tetragonolobus* (L.) DC. and *P. scandens* (Endl.) Verdc. (Leguminosae). Bot J Linn Soc 80:279–291. doi: 10.1111/j.1095-8339.1980.tb01665.x
Prakash D, Misra PN, Misra PS (1987) Amino acid profile of winged bean (*Psophocarpus tetragonolobus* (L.) DC): a rich source of vegetable protein. Plant Foods Hum Nutr 37:261–264. doi: 10.1007/BF01091791
Prakash D, Misra PN, Misra PS (1991) Variation in protein and trypsin inhibitor activity with maturity and amino acid composition of winged bean (*Psophocarpus tetragonolobus*) seeds. J Sci Food Agric 57:623–626. doi: 10.1002/jsfa.2740570414
Price T V., Drinkall MJ, Munro PE (1982) Studies of the incidence, severity and distribution of false rust (Synchytrium psophocarpi), leaf spot (Pseudocercospora psophocarpi) and powdery mildew (Oidium sp.) on winged bean (Psophocarpus tetragonolobus) in Papua New Gui. Ann Appl Biol 101:473–482. doi: 10.1111/j.1744-7348.1982.tb00848.x
Quan MH, Jiang XH, Chen DM, et al (2011) Identification and Analysis of A New Dwarf Mutant of Winged Bean. Abstract retrieved May 27th 2013
Rachie KO (1977) The nutritional role of grain legumes in the lowland humid tropics. In: Ayanabe A, Dart PJ (eds) Biological Nitrogen Fixation in Farming Systems of the Tropics. Wiley
Rahman MM, Islam AM, Azirun SM, Boyce AN (2014) Tropical Legume Crop Rotation and Nitrogen Fertilizer Effects on Agronomic and Nitrogen Efficiency of Rice. Sci World J 2014:1–11. doi: 10.1155/2014/490841
Rattalino Edreira JI, Budakli Carpici E, Sammarro D, Otegui ME (2011) Heat stress effects around flowering on kernel set of temperate and tropical maize hybrids. F Crop Res 123:62–73. doi: 10.1016/J.FCR.2011.04.015
Repinski SL, Kwak M, Gepts P (2012) The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. Theor Appl Genet 124:1539–1547. doi: 10.1007/s00122-012-1808-8
Rüegg J (1981) Effects of Temperature and Water Stress on the Growth and Yield of Winged Bean (*Psophocarpus Tetragonolobus* (L.) DC.). J Hortic Sci 56:331–338. doi: 10.1080/00221589.1981.11515009
Saeed A, Darvishzadeh R (2017) Association analysis of biotic and abiotic stresses resistance in chickpea (Cicer spp.) using AFLP markers. Biotechnol Biotechnol Equip 1–11. doi: 10.1080/13102818.2017.1333455
Sambudi H, Buckle KA (1991) Characteristics of winged bean (*Psophocarpus tetragonolobus*) seeds during soaking and boiling. J Sci Food Agric 57:585–595. doi: 10.1002/jsfa.2740570411
Sastrapradja S, Aminah Lubis SH (1975) *Psophocarpus tetragonolobus* as minor garden vegetable in Java. Bogor, Indonesia
Schiavinato MA, Válio IFM (1996a) Influence of photoperiod and temperature on the development of Winged Bean plants. RBrasFisiolVeg 8:105–110
Schiavinato MA, Válio IFM (1996b) Influence of staking on development of Winged Bean plants. RBrasFisiolVeg 8:99–

103
Senayake YDA, Sumanasinghe VAD (1978) Stigma receptivity in winged bean (*Psophocarpus tetragonolobus* (L.) DC.). SABRAO J Breed Genet 10:116–119
Sheokand S, Dahiya P, Vincent JL, Brewin NJ (2005) Modified expression of cysteine protease affects seed germination, vegetative growth and nodule development in transgenic lines of *Medicago truncatula*. Plant Sci 169:966–975. doi: 10.1016/j.plantsci.2005.07.003
Siebert S, Webber H, Zhao G, Ewert F (2017) Heat stress is overestimated in climate impact studies for irrigated agriculture. Environ Res Lett 12:054023. doi: 10.1088/1748-9326/aa702f
Singh V, Goel R, Pande V, et al (2017) De novo sequencing and comparative analysis of leaf transcriptomes of diverse condensed tannin-containing lines of underutilized *Psophocarpus tetragonolobus* (L.) DC. Sci Rep 7:1–13. doi: 10.1038/srep44733
Sinnadurai S, Nyalemegbe KD (1979) Effect of photoperiod on the nodulation flowering and seed yield of winged bean (*Psophocarpus tetragonolobus* (L.) DC.). Ghana J Agric Sci 12:27–30
Smartt J (1980) Some observations on the origin and evolution of the winged bean (*Psophocarpus tetragonolobus*). Euphytica 29:121–123. doi: 10.1007/BF00037256
Stephenson RA, Kesavan V, Claydon A, et al (1979) Studies on tuber production in winged bean (*Psophocarpus tetragonolobus* (L.) DC.). Proceedings of the 5th International Symposium on Tropical Root and Tuber Crops. Laguna, Philippines
Thompson AE, Haryono SK (1979) Sources of resistance to two important diseases of winged bean, *Psophocarpus tetragonolobus* (L.) DC. HortScience 14:532–533
Thompson AE, Haryono SK (1980) Winged bean: unexploited tropical food crop. HortScience 15:233–237
Tian Z, Wang X, Lee R, et al (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci 107:8563–8568. doi: 10.1073/pnas.1000088107
Tock AJ, Fourie D, Walley PG, et al (2017) Genome-Wide Linkage and Association Mapping of Halo Blight Resistance in Common Bean to Race 6 of the Globally Important Bacterial Pathogen. Front Plant Sci 8:1170. doi: 10.3389/fpls.2017.01170
Uemoto S, Fujieda K, Nonaka M, Nakamoto Y (1982) Effects of photoperiod and temperature on the raceme budding of winged beans (*Psophocarpus tetragonolobus*). Bull Inst Trop Agric Kyushu Univ 5:59–70
Umemoto K, Habu Y, Ohno T (1992) Analysis of differential accumulation of winged bean Kunitz chymotrypsin inhibitor mRNA species by a sequence-specific termination method. Plant Mol Biol. doi: 10.1007/BF00028904
van Wyk SG, Du Plessis M, Cullis CA, et al (2014) Cysteine protease and cystatin expression and activity during soybean nodule development and senescence. BMC Plant Biol 14:294. doi: 10.1186/s12870-014-0294-3
Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: Features and applications. Trends Biotechnol 23:48–55. doi: 10.1016/j.tibtech.2004.11.005
Vatanparast M, Shetty P, Chopra R, et al (2016) Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Sci Rep 6:1–14. doi: 10.1038/srep29070
Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. In: Chronica Botanica 13. Ronald Press, New York
Verdcourt B, Halliday P (1978) A Revision of *Psophocarpus* (*Leguminosae-Papilionoideae-Phaseoleae*). Kew Bull 33:191. doi: 10.2307/4109575
Watanabe S, Harada K, Abe J (2012) Genetic and molecular bases of photoperiod responses of flowering in soybean. Breed Sci 61:531–543. doi: 10.1270/jsbbs.61.531
Weil RR, Belmont GS (1991) Dry matter and nitrogen accumulation and partitioning in field grown winged bean. Exp Agric 27:323–328. doi: 10.1017/S0014479700019049
Weil RR, Khalil NA (1986) Salinity Tolerance of Winged Bean as Compared to that of Soybean. Agron J 78:67. doi: 10.2134/agronj1986.00021962007800010015x
Weller JL, Ortega R (2015) Genetic control of flowering time in legumes. Front Plant Sci 6:207. doi: 10.3389/fpls.2015.00207
Wong KC (1983) Effects of daylength, temperature, light intensity and applied growth substances on the growth, flowering and tuberization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.). Universiti Putra Malaysia
Wong KC, Schwabe WW (1979) Effects of daylength and day/night temperature on the growth, flowering and tuber formation of winged bean (*Psophocarpus tetragonolobus* (L.) DC.)
Wong Q, Tanzi A, Ho W, et al (2017) Development of Gene-Based SSR Markers in Winged Bean (*Psophocarpus tetragonolobus* (L.) DC.) for Diversity Assessment. Genes (Basel) 8:100. doi: 10.3390/genes8030100
Xu M, Xu Z, Liu B, et al (2013a) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol 13:91. doi: 10.1186/1471-2229-13-91
Xu M, Xu Z, Liu B, et al (2013b) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol 13:1. doi: 10.1186/1471-2229-13-91
Yang S, Grall A, Chapman MA (2018) Origin and diversification of winged bean (*Psophocarpus tetragonolobus* (L.) DC.), a multipurpose underutilized legume. Am J Bot 105:888–897. doi: 10.1002/ajb2.1093

Yoneyama T, Fujita K, Yoshida T, et al (1986) Variation in Natural Abundance of 15N among Plant Parts and in 15N/14N Fractionation during N2 Fixation in the Legume-Rhizobia Symbiotic System. Plant Cell Physiol 27:791–799. doi: 10.1093/oxfordjournals.pcp.a077165

Yuan S, Li R, Wang L, et al (2016) Search for Nodulation and Nodule Development-Related Cystatin Genes in the Genome of Soybean (*Glycine max*). Front Plant Sci 7:1595. doi: 10.3389/fpls.2016.01595

Zampieri M, Ceglar A, Dentener F, Toreti A (2017) Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. Environ Res Lett 12:064008. doi: 10.1088/1748-9326/aa723b

Zhai H, Lü S, Liang S, et al (2014) *GmFT4*, a homolog of *FLOWERING LOCUS T*, is positively regulated by *E1* and functions as a flowering repressor in soybean. PLoS One 9:. doi: 10.1371/journal.pone.0089030

Zhang X, Zhai H, Wang Y, et al (2016) Functional conservation and diversification of the soybean maturity gene *E1* and its homologs in legumes. Sci Rep 6:29548. doi: 10.1038/srep29548

Section B) Molecular marker development and applications

Cheng, A., Chai, H. H., Ho, W. K., Aliyu, S., Abdullah, B., Feldman, A., Kendabie, P., Halim, R. A., **Tanzi, A. S.**, Mayes, S., Massawe, F. (2017). Molecular Marker Technology for Genetic Improvement of Underutilised Crops. In S. N. A. Abdullah, C. L. Ho, & C. Wagstaff (Eds.), *Crop Improvement - Sustainability Through Leading-Edge Technology* (pp. 47–70). Springer International Publishing. https://doi.org/10.1007/978-3-319-65079-1

Published book chapter on molecular markers development and applications in underutilised crops. The information reported in sections 3.2.4, 3.2.5, and 3.2.6 in particular, introduce molecular markers systems and technologies that have been implemented in this project.

Molecular Marker Technology for Genetic Improvement of Underutilised Crops

Acga Cheng, Hui Hui Chai, Wai Kuan Ho, Aliyu Siise Abdullah Bamba, Aryo Feldman, Presidor Kendabie, Razlin Azman Halim, **Alberto Tanzi**, Sean Mayes, and Festo Massawe

A. Cheng
Biotechnology Research Centre, School of Biosciences, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

H.H. Chai
Biotechnology Research Centre, School of Biosciences, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

W.K. Ho • A. Tanzi • F. Massawe (*)
Biotechnology Research Centre, School of Biosciences, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia e-mail: festo.massawe@nottingham.edu.my

A.S.A. Bamba
Biotechnology Research Centre, School of Biosciences, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia CSIR-Savannah Agricultural Research Institute, Off Tamale-Tolon Road, Nyankpala, Tolon District, N/R, Ghana

A. Feldman • R.A. Halim
Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

P. Kendabie
Plant and Crop Sciences, School of Biosciences, The University of Nottingham, Sutton Bonington Campus, Leicestershire, LE12 5RD Loughborough, UK

S. Mayes
Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia Plant and Crop Sciences, School of Biosciences, The University of Nottingham, Sutton Bonington Campus, Leicestershire, LE12 5RD Loughborough, UK

3.1 Introduction

With the average global temperatures predicted to increase above 3.0 °C by mid-century, agricultural productivity is likely to significantly decline in many countries, particularly those currently in the hottest, most arid regions of the world (Abraham et al. 2014). There is growing evidence that rising temperatures are stifling the production of many important staples, notably maize (*Zea mays*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*), due to their sensitivity to water shortage and heat stress (Khoury et al. 2014). For example, maize yields in Africa are modelled to decrease by between 22 and 35% by 2030 because of the increased variability of rainfall patterns and changes in local temperatures (Ngwira et al. 2012; Shi and Tao 2014). Climate change, coupled with increasing global demand for food as the world population and its buying power grows, could also trigger a wave of price hikes for basic food commodities (Masters 2010).

While mainstream research has often focused on how to develop improved varieties of the world's mere 20 major food crops, thousands of indigenous and underutilized crops have been overlooked. Some of the underutilised (also called neglected, orphan or minor) crops are likely to be more resilient to harsh conditions than the major cereal crops as a result of the low-input farming systems in marginal environments in which they have been selected (Mayes et al. 2011). Until recently, unlocking the potential value of underutilised crops has not been seen as a mainstream approach to addressing the effects of climate change on agriculture. Although some of these crops, like quinoa (*Chenopodium quinoa*) and bambara groundnut (*Vigna subterranea* (L.) Verdc.), have recently received more research attention (although not necessarily more research funding), existing knowledge on the genetic potential of many other promising underutilised species, such as the high-protein legume winged bean (*Psophocarpus tetragonolobus*), remains limited (Massawe et al. 2016). There is, therefore, an urgent need for more innovative research to boost the productivity of potential future crops and molecular marker (or genetic marker) technology is considered to be a critical component in this endeavour. This far-reaching technology has indeed revolutionised the way plant genetics research is conducted (Semagn et al. 2006; Govindaraj et al. 2015) (Fig. 3.1).

Molecular markers are capable of detecting allelic variation or modification (e.g. methylation) in DNA sequence, and have been applied over a number of years to answer many complex biological questions in plant and animal science. These range from simple genetic fingerprinting for quality control in breeding programmes, to complex questions in population ecology and germplasm structure (Joshi et al. 1999; Schlotterer 2004). The last two decades have seen substantial investments in large-scale genetic and genomic resource development on major crops, including the development of simple sequence repeats (SSRs), expressed sequence tags (ESTs) and single nucleotide polymorphism (SNPs) (Schlotterer 2004; Kesawat and Kumar 2009). This is particularly the case for the major cereals, strengthening the applications of marker-assisted selection (MAS) (Gupta et al. 2010; Miah et al. 2013; Huang et al. 2015).

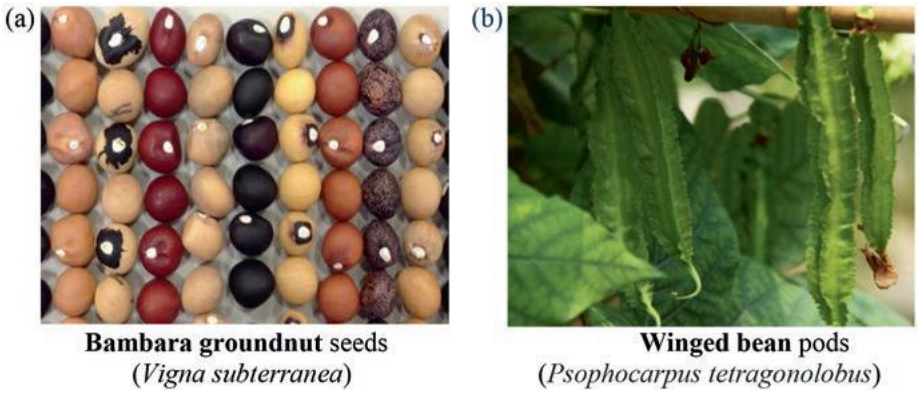


Fig. 3. 1 Examples of underutilised crops *Source:* http://www.cropsforthefuture.org/

Compared to conventional breeding, MAS:

1. Does not require the plant to express the desired phenotypic trait to be able to select for plants carrying the trait.
2. Is not affected by the environment.
3. Is able to recognise heterozygous individuals, for example, hybrids can be identified by co-dominant markers.
4. Makes gene pyramiding possible by introgressing multiple quantitative trait loci (QTL) into an individual.
5. Is more time- and resource-efficient.

Advances in molecular biology, principally the invention of polymerase chain reaction (PCR) in the mid-1980s, have resulted in a wealth of molecular marker systems with potentially widespread utility for detecting genetic variation in plants (Semagn et al. 2006; Kumar et al. 2009). Although most research to date have been dedicated to major crops, the use of molecular marker technology to investigate the genetic basis of important traits for a range of underutilised species is increasing (Moe et al. 2012). The first section of this chapter will

highlight the types of marker systems commonly used in crop improvement, followed by a thorough discussion on the past, current, and potential utilisation of molecular marker technology for the selection of the underutilised crop examples in the subsequent sections.

3.2 Commonly Used Molecular Marker Systems for Crop Improvement: Principles and Progress

Molecular marker systems can reveal genetic variation (or polymorphism) in DNA sequence, often arising from mutations that occur during DNA replication or repair. Plant molecular genetics has progressed dramatically since the establishment of the first molecular markers, allozymes (or isozymes), in the 1960s. Allozymes work on the principle of the electrophoretic mobility of protein variants in enzymes modified by amino acid substitution, when proteins are run on gel matrixes. Specific enzymatic stains can reveal the position of the protein bands on a gel, with amino acid
changes effecting the migration of the band (Schlotterer 2004). In the past, these biochemical markers were widely used in plant population genetic studies, including measurements of population structure and divergence (Troconis-Torres et al. 2002). The popularity of allozymes decreased with the advent of DNA-based markers, which have higher discrimination level at the individual level and for which
there are far larger number of markers available. These include restriction fragment length polymorphisms (RFLPs) (Tanksley et al. 1989), and a suite of PCR-based markers, such as SSRs (Varshney et al. 2005) and SNPs (Ganal et al. 2009). This section presents information on the most commonly used DNA-based marker systems in plants (Table 3.1).

3.2.1 Restriction Fragment Length Polymorphism (RFLP)

The discovery of restriction enzymes (or restriction endonucleases) in the 1960s has led to the first DNA-based marker system known as RFLP (Schlotterer 2004). The marker system exploits variation in homologous DNA sequences, such as base pair changes within the endonuclease recognition sequence. The resulting fragments are separated by size using gel electrophoresis, followed by Southern blotting and hybridisation in solution of the filter to a labelled and denatured probe that identifies the locus of interest (Tanksley et al. 1989). Restriction enzymes are generally classified into four types; Type I, II, III, and IV, on the basis of cleavage position, sequence specificity, subunit composition, and cofactor requirements (Tomar 2010). A list of known restriction enzymes, along with their recognition sequences and cleavage sites, is provided in REBASE, a fully curated database available at http://rebase.neb.com/rebase/rebase.html (Roberts et al. 2003).
RFLPs were one of the most commonly used technologies to construct DNA fingerprints (Weising et al. 1994) until the emergence of PCR-based marker systems. RFLPs were predominantly used to measure genetic diversity and construct genetic linkage maps for most major crop plants, notably maize, wheat and rice (Tanksley et al. 1989; Schlotterer 2004). Nonetheless, RFLPs are still very much in use today, mainly because of their high reproducibility, co-dominant nature and relatively low cost per data point. They have also been one of the major marker systems to confirm the presence and copy number of transgenes (Bisen 2014).

Table 3. 1 Comparison of advantages and disadvantages offered in different DNA-based marker systems

Marker	Advantages	Disadvantages
RFLPs	• High reproducibility	• Requires large amounts of high quality DNA
	• No sequence information required	• Difficult to automate
		• Involve radioactive methods
		• Cloning and characterisations of probe are required
RAPDs	• Inexpensive	• Low reproducibility
	• No sequence information required	• Mainly dominant
	• Produces a large number of bands, which can be further characterised individually	• Difficult to analyse
		• Difficult to automate
		• Cross-study comparisons are difficult
AFLPs	• High reproducibility	• Mainly dominant
	• A large number of loci can be analysed simultaneously	• Difficult to automate
SSRs	• Highly informative	• High mutation rate
	• High heterozygosity	• Complex mutation behaviour
	• High reproducibility	• Low to moderate abundance
	• Low ascertainment bias	• Special preparation required for cross-study comparison
	• Easy to isolate	• Sequence information required
SNPs	• Low mutation rate	• Substantial rate heterogeneity among sites
	• Easy to automate	• Can be expensive to isolate
	• High abundance	• Ascertainment bias
		• Low information content of a single SNP

Adopted from Schlotterer 2004, Kumar et al. 2009

3.2.2 Random Amplified Polymorphic DNA (RAPD)

The discovery of PCR has led to the development of a handful of novel genetic assays characterised by the selective amplification of defined DNA fragments. In 1990, the first PCR-based marker system, RAPD, was described based on the amplification of random DNA fragments using short synthetic primers of arbitrary nucleotide sequence (Williams et al. 1990). These primers, usually 10 bp in length, occur at such high frequency in the genome that two random binding sites orientated to allow PCR are likely to occur a number of times in the genome of most species, leading to multiple band profiles. The amplified fragments are separated by gel electrophoresis, stained with ethidium bromide, and visualised under ultraviolet light. RAPD polymorphisms are indicated by the presence or absence of bands of particular sizes in comparison to other individuals, and hence treated as dominant markers (Williams et al. 1990). Inherited in a Mendelian fashion, their polymorphisms can arise as the result of base substitutions at the primer annealing sites or INDELs in the regions between the sites. The RAPD technology represented a useful tool for plants with no existing markers, such as the endangered seven-son flower (*Heptacodium miconioides* Rehd.) (Liu et al. 2007a). From its inception until now, RAPDs have been largely used for genetic diversity and phylogenetic studies at the genus or species level, with the most recent example being *Pistacia* (Iranjo et al. 2016). RAPDs possess almost all the advantages of a PCR-based marker, offering rapid results, technical simplicity and no need for species-specific development. Beyond these advantages, RAPD analysis uses only small amounts of DNA (10–100 ng per reaction), and requires no prior knowledge of the sequences of the target genome (Arif et al. 2010). One inherent drawback of RAPDs is their low reproducibility, and thus highly standardized experimental procedures are required to achieve reproducible profiles. The reproducibility testing of RAPD in plants by a number of European laboratories revealed that the technique) is stable within labs, but not between labs (Jones

et al. 1997). Another major drawback of RAPDs is their dominant nature of inheritance, which limits their use for genetic mapping and marker-assisted selection (MAS) (Bardakci 2000).

3.2.3 Amplified Fragment Length Polymorphism (AFLP)

AFLP (Vos et al. 1995) is another marker technology that has been widely used to assess the genetic diversity and phylogenetic studies in plants (Mba and Tohme 2005). This dominant marker system can be used in any species, including non-model organisms, as no prior sequence information is required. Unlike RAPDs, AFLPs are highly reproducible because they use a combination of restriction digestion and PCR amplification (Arif et al. 2010). However, the AFLP method is more technically challenging than RAPD, requiring highly purified genomic DNA. The standard AFLP protocol begins with the digestion of genomic DNA with two restriction enzymes, generally one enzyme with an average cutting frequency (e.g. 6-base cutter *Eco*RI), and the other with a higher cutting frequency (e.g. 4-base cutter *Mse*I). The resulting DNA fragments are then ligated to specific adapters, and a subset of these fragments are amplified using primer pairs that consist of common sequences of the adapter and arbitrary nucleotides extending into the unique sequence adjacent to the restriction enzyme recognition site. The amplified fragments, usually within 60–500 bp size range, are separated by the polyacrylamide gel electrophoresis (PAGE) or on a capillary sequencer. The former is rather cumbersome, time-consuming and labour intensive, the latter relatively expensive (Lin et al. 1996; Cheng et al. 2012).
The standard AFLP protocol has been modified by Ranamukhaarachchi et al. (2000) to develop a rapid plant genetic characterisation analysis with improved intensity and specificity, which is convenient to perform and reliable. The modified AFLP protocol involves the use of standard agarose as the medium of electrophoresis, and is carried out with only a single restriction enzyme and one type of adapter. However, as the gel resolution decreases, the chance of mis-calling bands as identical when they are in fact different in size increases. For high-throughput genotyping, specific AFLPs) can sometimes be converted into co-dominant and sequence-specific markers, such as sequence tagged sites (STSs) and cleaved amplified polymorphic sites (CAPSs) (Meksem et al. 2001).

3.2.4 Simple Sequence Repeat (SSR)

Plant genomes typically contain a large number of SSRs; DNA sequence motifs of up to 6 bp units repeated in tandem and widely scattered throughout the genome. Most SSR repeat motifs are composed of di-, tri-, or tetra-nucleotide sequences, with $(GA)_n$ and $(AT)_n$ being the most frequent repeat types observed (Powell et al. 1996). Highly reproducible, locus-specific, simple detection and abundance are all characteristics of SSRs, which have made them one of the most preferred PCR-based markers for assessing genetic diversity, population structure and evolution analysis in plants (Fu et al. 2013). Due to their co-dominant nature, high polymorphism and random distribution within the genome, SSRs are extensively used for genotyping and genetic map construction. For instance, a genetic linkage map consisting of 15 linkage groups with 406 SSRs covering 1143.5 cM of the genome was developed using a pseudo-testcross population derived from an intraspecific cross of two tea plants (*Camellia sinensis* varieties). The QTL analysis indicated that nine QTLs associated with catechins content were clustered into four linkage groups (LG03, LG11, LG12 and LG15), providing information for functional studies to improve tea plants (Ma et al. 2014).
Nonetheless, despite their many advantages, the de novo development of SSRs can be costly and time-consuming (Yu and Li 2007). Unlike sequence-arbitrary methods, such as AFLP, the sequence of each SSR locus needs to be known to allow the development of locus-specific, co-dominant, primers in the unique flanking regions of each SSR repeat. Classically, SSR analysis involves the construction of genomic DNA libraries. The preparation of these libraries is a tedious process, and the specific DNA sequencing required is expensive (Edwards et al. 1996). Nevertheless, the development of advanced molecular tools over the past two decades has enabled establishment of cost-effective, high-throughput methods for SSR analysis. These include the mining of sequences from DNA databases (e.g. those deposited at NCBI), and the sequencing method where the whole genome or parts of the genome (e.g. exome or transcriptome) are sequenced (Saeed et al. 2016). This has facilitated the development of Expressed Sequence Tag (EST) or gene-based SSRs (Zhao et al. 2012), which have higher recovery rates than genomic SSRs (Ellis and Burke 2007; Zhao et al. 2012).
Numerous studies have found that EST-SSRs are generally more transferable across taxonomic boundaries than the genomic SSRs (Ellis and Burke 2007). Despite their relatively low polymorphism (Yu and Li 2007),

EST-SSRs have higher transferability rates, are more practical, more informative, and having a lower frequency of null alleles. These markers have been developed and utilised in many crop species, including several primary oil crops such as soybean (*Glycine max*) and rapeseed (*Brassica napus*) (Ellis and Burke 2007; Lai et al. 2013).

The SSRs have often been the marker chosen in studies of crops lacking of resources and species-specific information. One recent example is a study conducted on common vetch (*Vicia sativa*). The study reported the transferability of SSRs from *Vicia sativa* subsp. sativa across 22 *Vicia* species, which ranged between 32 and 82% (Raveendar et al. 2015). There have been increasing numbers of SSR being identified through the utilisation of the next-generation sequencing (NGS) technology particularly from transcriptomic data as the genic SSR are in the conserved coding regions, which, in turn, are highly transferable to related taxa and useful in association or comparative mapping. For example, a total of 3011 genic SSR have been identified as potential markers from 26 M reads of transcriptome data derived from root, stem and leaf tissues of rice bean (*Vigna umbellata* L.), with mono-nucleotide repeat motifs being the most abundant (47.3%) (Chen et al. 2016). Nevertheless, SSR motifs within coding sequence are constrained, as changes in repeat number will effect protein translation and can lead to a frameshift mutation. For this reason, trinucleotides repeat until are most common in coding sequence and most likely to be polymorphic, while dinucleotide motifs can be found in the upstream and downstream transcribed, but not translated regions. The biggest problem with such dinucleotide motifs is that the flanking sequence region is often highly constrained for primer design (Lin and Kussell 2012; Gonthier et al. 2015).

3.2.5 Single Nucleotide Polymorphism (SNP)

From the late 1990s to the early 2000s, the SSRs had indeed pervaded modern plant breeding and genetics research. The dominance of SSRs has been, however, replaced by SNPs during the last decade. Despite having lower polymorphism rates than the SSRs because of their biallelic nature (Kumpatla et al. 2012), SNPs have emerged as the current markers of choice because they are more stable, efficient, and amenable to high- and ultra-high-throughput automation. SNPs represent variation in a DNA sequence that affects only a single nucleotide; adenine (A), cytosine (C), guanine (G) or thymine (T) (Kumpatla et al. 2012). They are the most abundant class of polymorphisms in both plant and animal genomes, making them particularly suited for high-resolution mapping. Inherited as co-dominant markers, SNPs are able to discriminate efficiently between homozygous and heterozygous alleles (Arif et al. 2010). Nonetheless, the SNP system is not without limitations. The discovery of

SNPs can be a difficult and complex process, depending on the level of genome complexity and the availability of reference genome sequences.

Although numerous protocols for SNP discovery have been described, the typical SNP genotyping protocol consists of three major components: target amplification, allelic discrimination and product detection (Chen and Sullivan 2003). The last half-decade has seen tremendous progress in SNP genotyping technologies, such as the TaqMan assay (allele specific hybridisation), SNP arrays (e.g. GoldenGate assay, BeadArray, microarray technology, KASP array), genotype-by-sequencing (GbS) and resequencing for variant calling. Each of these technologies renders a unique combination of throughput, accuracy, scale, and cost (Kumar et al. 2012; Huang et al. 2015). With the reduction in the cost of NGS technology, both de novo and reference-based SNP discovery and genotyping are now relatively feasible for various plant species.

Reference-based SNP identification using resequencing is more applicable to those model and major crops that have a complete or draft genome. For studies of underutilised crops, the GbS approach would be the choice of interest given ex ante sequence information is not required. NGS-derived SNPs have been recently reported in leguminous species such as alfalfa (*Medicago sativa*) and common bean (*Phaseolus vulgaris*) (Huang et al. 2015). More recently, restriction site associated DNA (RAD) sequencing has been adopted to produce 1894 high quality SNP markers and together with 68 SSRs, they were mapped onto seven chromosomes in barley (Ren et al. 2016). The genetic map allowed the identification of a new recessive dwarfing gene *btwd*1 at 0.7 cM and 0.9 cM on chromosome 7H, giving a new insight into the use of SNPs markers for positional cloning of the gene and MAS in barley (Ren et al. 2016). While automation is often the most crucial factor in selecting a genotyping system for a large-scale SNP profiling), other factors such as cost, throughput, and data quality obtained are also of considerable importance (Huang et al. 2015).

3.2.6 Diversity Arrays Technology (DArT)

Generally, in the GbS approach, the genome complexity of a species is reduced using restriction enzymes prior to sequencing. For example, with the appropriate choice of methylation-sensitive restriction enzymes in a combination with a frequent cutting enzyme, repetitive fractions of the genome can be avoided targeting informative low copy sequences (Sánchez-Sevilla et al. 2015). There have been more than 30,000 wheat germplasm accessions at the International Maize and Wheat Improvement Centre (CIMMYT) genotyped using the GbS technique, specifically the Diversity Arrays Technology (DArT)-seq pipeline, under its Seeds of Discovery (SeeD) initiative (http://seedsofdiscovery.org/about/genotyping-platform/). Developed in the early 2000s, DArT offers a high-throughput genotyping platform for detecting sequence polymorphisms in a single experiment without species-specific sequence information.
Over the past few years, DArT marker system has been widely used for genetic diversity analysis, evaluation of population structure of a collection of germplasm, construction of genetic linkage maps, and alignment of genetic maps of different segregating populations (Phung et al. 2014; Sánchez-Sevilla et al. 2015). There are two types of platforms for DArT markers development, namely microarray-based DArT and DArTseq (Cruz and Kilian 2013). Practically, the DArT is a complexity reduction method that involves the use of methylation-sensitive restriction enzymes such as *Pst*I/*Taq*I (Semagn et al. 2006) or *Pst*I/*Bst*NI (Cruz and Kilian 2013). Genomic 'representations' are selected through amplification of primers complementary to the adaptors after genomic DNA is isolated and fragmented using enzymes, followed by the ligation of restricted fragments with adaptors. In effect, the technique amplifies the remaining *Pst*I-*Pst*I fragments. A 'Diversity Panel' is then generated when selected genomic 'representations' are purified and spotted onto a microarray in the DArT Arrays method (Jaccoud et al. 2001). When individual DNA samples are hybridised onto this 'Diversity Panel', polymorphic DArT markers are identified based on hybridisation signal intensities and assembled into a 'genotyping array' for routine genotyping consisting of presence and absence scores (Jaccoud et al. 2001; Semagn et al. 2006).
The DArTseq technique is suitable for detailed exploitation of genetic diversity and population structure, gene discovery for molecular breeding, high-resolution genetic mapping and MAS as no sequence information is required (Cruz and Kilian 2013). DArTseq genotyping was reported to produce 2835 polymorphic markers in *Eucalyptus* as compared to 1088 microarray-based DArT markers (Sansaloni et al. 2011). Due to the capability of producing a large quantity of polymorphic markers at a similar cost, DArTseq genotyping is considered to be more cost effective than microarray-based DArT.

3.3 Utilisation of Molecular Markers in the Studies of Underutilised Crops

It is apparent that there has been continuous progress and evolution of molecular marker systems used to meet the needs of modern plant breeding programmes. The selection of preferred molecular markers depends immensely on the crop species of interest, the intended research outcomes and the financial resources available (Kesawat and Kumar 2009). In the present era, DNA-based marker systems, notably SSRs and SNPs, have become increasingly important to underutilized crops by way of improving access and utilisation of germplasm resources, genetic
analysis of breeding populations, parental selection and predicting progeny performance, marker-assisted selection, marker-enriched backcross breeding, comparative
mapping, and gene isolation, function and manipulation. These systems
have found their niche applications in various fields, which are discussed in detail
in Sects. 3.2 and 3.3.

3.3.1 Underutilised Crops: What Are They and Where Do They Come From?

Underutilised crops are, by and large, indigenous crop species which are still used by communities at a local or regional level, but are largely overlooked by researchers, producers and consumers at a global level (Mayes et al. 2011). These underutilized crops are often bestowed with greater stability to cope with harsher and more

dynamic environmental conditions (Thies 2000). Some of these crops, especially for underutilised cereals such as quinoa and teff (*Eragrostis tef*), are well adapted to the socio-economic conditions in their native habitats, and are preferred by both local farmers and consumers (Thies 2000; Cheng et al. 2015). Table 3.2 shows some examples of underutilised crop species and their known centre(s) of diversity.

Table 3. 2 List of recently available genomes or draft genomes on some underutilised crops

Crop example	Crop type	Centre(s) of diversity	Genome size (Mbp)	Year	References
Foxtail millet (*Setaria italica*)	Cereal	Asia, Mediterranean/ Southwest Asia	~490	2012	Bennetzen et al. (2012)
Muskmelon (*Cucumis melo*)	Vegetable	South Mexico, Central America	~375	2012	Garcia-Mas et al. (2012)
Peach (Prunus persica)	Fruit	Asia (Chinese Centre)	~265	2012	Verde et al. (2013)
Pigeon pea (*Cajanus cajan*)	Legume	Asia (Indian Centre)	~835	2012	Varshney et al. (2012)
Tausch's goatgrass (*Aegilops tauschii*)	Cereal	Middle East	~4400	2013	Jia et al. (2013)
Quinoa (Chenopodium quinoa)	Cereal	Latin America	~1450	2017	Jarvis et al. (2017)

The term 'underutilised', or other common synonymous terms such as 'orphan', 'neglected' and 'minor', does not necessarily reflect the crops' geographical distribution, nor their social or economic implications. For instance, many crop species are consumed as staple by millions of people within their centres of origin, but their poor marketability and limited accessibility make them largely underutilised in both social and economic terms (Padulosi and Hoeschle-Zeledon 2004). One of the key reasons behind their underutilisation is that they are neglected by mainstream research, mainly because they are not considered global crops and are mostly produced and consumed subsistently (Naylor et al. 2004; Massawe et al. 2016).
Nevertheless, in recent years, a small number of underutilised crops, Bambara groundnut, winged bean, amaranth, and proso millet (*Panicum millaceum*), to name a few, has attracted worldwide research attention. This is mainly due to their ability to thrive in extreme environments with the increasing concern on global warming and other negative future climate changes, as well as their potential to contribute to nutritional security (Ray et al. 2013). The genome of some of these crops have also been fully sequenced (Table 3.2). Molecular genetics and genomics research on underutilised crops have gained momentum in the present decade, spurred largely by the advent of powerful yet affordable molecular marker technology. A handful of marker systems are being used extensively to investigate the genetic basis of agronomically important traits for a range of underutilised species, as discussed below.

3.3.2 Summary of Recently Developed DNA-Based Markers in Underutilised Crops

Recent years have been marked by significant progress in the development of DNA-based marker systems in underutilised crop species. Specifically, in the past 5 years, numerous markers representing various marker types, such as RAPDs, AFLPs, SSRs, and SNPs, have been used for genetic analysis in many potential underutilised crops (summarised in Table 3.3). Protein-coding sequences, functionally characterized genes and EST sequences have been made available to build up new generation markers like genic SSRs and SNPs. Given that SSRs and SNPs have rapidly emerged as the genetic markers of choice in view of their abundance, high transferability across related taxa, and amenability to automated analysis (Kumar et al.; as summarised in Table 3.1), this section focuses chiefly on the development of these marker systems.
SSR marker system has been extensively used in genetic mapping and molecular breeding in underutilised crops, in particular underutilised cereals/pseudo-cereals and legumes (Zeid et al. 2012; Bohra et al. 2014).

Major progress has been seen in the development of EST-SSRs in underutilised species in the past decade, leading by a couple of highly potential underutilised species, including quinoa and finger millet (*Eleucine coracana*) (Dawson et al. 2007). EST-SSRs have also been developed for genetic analysis, for example, in adzuki bean (*Vigna angularis*) (Chen et al. 2015). SNP marker system has become more popular than SSRs in light of their higher abundance and suitability for automatic allele calling (Arias et al. 2012). The advent of NGS has allowed the identification of large collections of SNPs for some underutilised species, including Tausch's goatgrass (*Aegilops tauschii*) (Mochida and Shinozaki 2013). At present, SNP discovery is being carried out in many other underutilised species such as teff (Cheng et al. 2015), and roselle (*Hibiscus sabdariffa*) (Melo et al. 2016).

3.3.3 Development of Choice of Marker Systems for the Development of Some Underutilised Crops

3.3.3.1 Underutilised Cereal and Pseudo-cereal

Cereals are a major source of food in the world, and of the approximately 50,000 known edible plant species, more than 10,000 are cereals (Ji et al. 2013). Being staples that feed almost the entire human population, cereals are grown in far greater quantities than any other type of crop. Maize, wheat, and rice alone account for nearly two-thirds of the world's plant-derived food energy (Ji et al. 2013). Although the annual production of major cereals has seen a steady increase since the dawn of the Green Revolution, rapid changes in climate pose a huge threat to the sustainability of the current productions. This is, perhaps, the chief reason why there has been a marked shift in research attention from a handful of major cereals to a wider range of cereals in the past decade (Pingali 2012). Two Andean cereal grains; quinoa and grain amaranth, have recently taken the spotlight as 'superfoods', due largely to their glutenfree property and nutrient composition (Massawe et al. 2016). They are often referred to as pseudo-cereals, as they do not belong to the grass family (Gramineae) but have similar uses as true cereals. Quinoa has had a slightly higher popularity over the past half-decade, receiving numerous titles and honours including 'Queen of Superfoods' and 'International Year of Quinoa' (http://www.fao.org/quinoa-2013/en/).

Quinoa has been an underutilised priority species since the early 2000s (Williams and Haq 2002). During the early years, its molecular studies were performed using AFLP, RAPD and SSR marker systems (Maughan et al. 2004). The first genetic linkage map of quinoa, covered an estimated 60% of its genome, was developed in 2004 based on AFLPs, RAPDs and SSRs (Maughan et al. 2004). The subsequent studies were based primarily on SSRs, whereby more than 400 SSRs were developed and characterised (Jarvis et al. 2008). Christensen et al. (2007) used the developed SSRs to assess the genetic diversity among 121 accessions of quinoa within the United States Department of Agriculture (USDA) collection, reporting that the accessions can be clustered into two major groups: one comprising accessions from the lowlands of Chile (Coastal ecotype), and the other comprising accessions from the Andean Highlands (Altiplano ecotype). Tártara et al. (2012) conducted a more recent SSR-based genetic diversity analysis on quinoa using 35 germplasm accessions from Northwest Argentina.

The first EST libraries (424 ESTs) for quinoa were described in 2005 (Coleset al. 2005) using developing seed and floral tissue. The study yielded 38 SNPs and 13 INDELs based on 20 EST sequences derived from five quinoa accessions. Maughan et al. (2012) identified 14,178 putative SNPs for quinoa using pyrosequencing from five mapping populations with 511 were developed into KASPar-based SNP array to genotype 113 accessions of C. quinoa and eight from Chenopodium taxa. Subsequently, an integrated SNP-based linkage map was generated using two recombinant inbred line (RIL) populations. The map consisted of 29 linkage groups with an average distance of 3.1 cM between SNPs. The same research group also sequenced amaranth from four parental lines using the same technique, yielding 27,658 putative SNPs (Maughan et al. 2009). The developed SNPs were then used to generate the first complete linkage map of the *Amaranthus* genus (Maughan et al. 2011). In this map, a total of 411 SNPs were assigned to 16 linkage groups, spanning 1288 cM with an average distance of 3.1 cM between SNPs (Maughan et al. 2011). In 2014, Illumina Hi-Seq RNA sequencing (RNA-seq) technology has been used to perform the first large transcriptome analysis of drought-induced stress in quinoa (Raney et al. 2014).

Table 3. 3 DNA-based markers developed and genetic analysis of selected underutilised crops

Crop example	Marker type(s)	Research description	References
Bambara groundnut (*Vigna subterranean*)	SSR	SSR-based analysis of genetic diversity of Ghanaian bambara groundnut landraces	Abdullah Bamba and Massawe (2013)
	SSR, DArT	SSR-based analysis of genetic diversity and population structure in bambara groundnut landraces	Molosiwa et al. (2015)
	SSR, DArT	Construction of linkage map and QTL analysis of phenotypic traits in bambara groundnut	Ahmad et al. (2016)
Cucurbitis (*Cucurbitaceae*)	AFLP, SSR, SNP	Construction of genetic map of and QTL analysis of fruit quality in melon	Harel-Beja et al. (2010)
	SNP	Construction of ultrahigh-density linkage map for cultivated cucumber using SNP genotyping array	Rubinstein et al. (2015)
Finger millet (*Elusine coracana*)	RAPD SSR	RAPD-based analysis of genetic diversity in finger millet	Kumari and Pande (2010)
	Genic-SSR	QTL analysis of blast disease resistance in finger millet	Babu et al. (2014)
Grain amaranth (*Amaranthus* spp.)	SNP	Development and analysis of SNPs in grain amaranth; characterisation of the first linkage map in the genus	Maughan et al. (2011)
	SSR	SSR-based analysis of genetic diversity and population structure in *Amaranthus* species	Suresh et al. (2014)
Pigeon pea (*Cajanus cajan*)	RAPD	Identification of RAPDs linked to plant type gene in pigeon pea	Dhanasekar et al. (2010)
Proso millet (Panicum miliaceum)	SSR	SSR-based analysis of genetic diversity and phylogeography of proso millet across Eurasia	Hunt et al. (2011)
Teff (Eragrostis tef)	SSR	Linkage-map construction and QTL analysis of yield and lodging resistance in an enhanced SSR-based map	Zeid et al. (2011)
	SSR	SSR-based analysis of genetic diversity and relationships within and among E. tef, E. pilosa and E. curvula	Zeid et al. (2012)
Winged bean (Psophocarpus tetragonolobus)	SSR and SNP	Development of SSRs and SNPs in two Sri Lankan winged bean accessions	Vatanparast et al. (2016)
	Genic-SSR	Development of gene-based SSRs in winged bean for diversity assessment	Wong et al. (2017)

3.3.3.2 Underutilised Legumes

Legumes, members of the Leguminosae (or Fabaceae) family, are an important component for low-input agriculture. Their desirability rests on their ability to fix atmospheric nitrogen, which consequently increases soil fertility for the production of the neighbouring or subsequent non-legume crop cycle such as rice (Rahman et al. 2014). While legumes are one of the most nutritious, versatile and relatively more affordable foods available all over the world, the utilisation of most leguminous species is relatively low with groundnut and soybean being exceptions due to their high oil and protein contents. Nevertheless, the importance of several leguminous species has been recently highlighted, as in the case for bambara groundnut (Massawe et al. 2016). The groundwork for the genetic improvement in bambara groundnut began in the late 1990s, focusing primarily on increasing the understanding of its genetic diversity within and between landraces (Massawe et al. 2005). High levels of polymorphism among African bambara groundnut landraces were observed using different DNA-based marker systems, such as AFLP (Massawe et al. 2002) and RAPD (Massawe et al. 2003). These findings are further supported by a study conducted by Abdullah Bamba and Massawe (2013), which demonstrated the high level of SSR polymorphism among a collection of Ghanaian bambara groundnut landraces. The first attempt in the development of SSR markers for bambara groundnut, which yielded ten SSRs, were reported by Basu et al. (2007). More recently, a total of 75 SSRs have been developed for this legume through various sequencing approaches for the analysis of its genetic diversity and population structure, as well as for the selection of pure lines from landraces (Molosiwa et al. 2015). SSR marker system has also been utilised to assess the breeding system and varietal purity of bambara groundnut (Ho et al. 2016). The first linkage map of bambara groundnut was developed from a F3 population derived from a cross between two single genotypes derived from Tiga Nicuru (a landrace from Mali) and DipC (a landrace from Botswana) using 238 SSR and DArT array markers. In this map, a total of 36 QTLs were detected at eight linkage groups, associating with 19 phenotypic traits (Ahmad et al. 2016). These detected QTLs could be useful for future breeding programmes in bambara groundnut.

3.3.3.3 Other Types of Underutilised Crops

An abundance of vegetables, fruits, roots, and tubers has been important source of food and medicine for millennia, and yet, many of them are still widely underutilized and underexploited (Ebert 2014). Nonetheless, in the case of vegetables, numerous conservation and development initiatives have already been launched and championed by the World Vegetable Centre (AVRDC) (Ojiewo et al. 2010). Today, AVRDC maintains the largest vegetable germplasm collection of approximately 61,000 accessions, including about 12,000 accessions of indigenous species.

The Cucurbitaceae, also known as cucurbit or vine crop family, provides a wide variety of vegetables such as squash, melon, and several Cucurbita species with edible fruits. One example of potential underutilised cucurbits in which genetics have been fairly studied is the egusi (Citrullus colocynthis), also called bitter apple, bitter cucumber or desert gourd. Closely related to cultivated watermelon (Citrullus lanatus), egusi is an important medicinal plant that strives in arid environments (Dane et al. 2006). Several marker systems have been used to estimate the genetic diversity among egusi and cultivated watermelon accessions, including RAPD (Levi et al. 2001), AFLP and SSR (Nimmakayala et al. 2010; Hwang et al. 2011). Two major clusters were observed using 965 AFLP and EST-SSR markers, separating 27 watermelon cultigens from four wild-type species. Si et al. (2009) found dynamic gene expression changes in response to drought stress in egusi root tissues using cDNA-AFLP. Time point leaf transcriptomic changes at seedling stage was further elaborated by Wang et al. (2014) utilising NGS technology. This drought tolerant species has been used as a model to elucidate genes responsible for stress tolerance in other cucurbit crops through genetic manipulation (Wang et al. 2014).

Yam (Dioscorea spp.), among a handful of tropical root and tuber crops, is an important staple food for millions of people in Africa. The white (D. rotundata) and yellow Guinea (D. cayenensis) yams are the two most important yam species in Africa, while the water yam (D. alata) is the most widely distributed species in the world (Mignouna and Abang Asiedu 2007). The initial effort in yam genetics began with the development of

its polymorphic DNA-based markers in the late 1990s. The markers of choice in early studies of yam genetic variation were AFLP (Mignouna et al. 1998) and RAPD (Dansi et al. 2000). In 2003, a set of 36 SSRs was developed and the efficiency of these SSRs was assessed together with the previously developed AFLPs and RAPDs in analysing the genetic diversity among 45 accessions of D. alata (Mignouna et al. 2003). Each of the three marker systems showed polymorphism among the accessions, with AFLP revealing the highest level of polymorphism (Mignouna et al. 2003).

The first two comprehensive linkage maps of yams, one for D. alata (Mignouna et al. 2002a) and the other for D. rotundata (Mignouna et al. 2002b), were developed concurrently based on AFLPs. Since then, considerable progress has been made in the development of molecular tools for mapping and QTL analysis in yams (Petro et al. 2011). Narina et al. (2011) developed the first extensive public Dioscorea EST collection, yielding a total of 44,757 EST sequences. More recently, genomic resources for D. alata have been generated by Saski et al. (2015), including 288,505 high quality SNPs and more than 30,000 genomic SSRs. The availability of these resources will cater tools for wider compatibility across different Dioscorea species and for MAS in their breeding improvement programmes (Saski et al. 2015).

3.4 Potential Use of Molecular Marker Technology in Genetic Improvement of Underutilised Crops

The evolution of molecular marker technology constantly offers advanced tools for wider compatibility across different crop species, genetic analysis of complex traits and for MAS in crop breeding programmes (De La Fuente et al. 2013; Saski et al. 2015). In the case of major crops, the new generation marker systems such as SSR and SNP have been routinely used in the MAS breeding programmes worldwide to develop improved or new varieties (Gupta et al. 2010). The success of MAS depends largely on several important factors, including the genetic distance of the target gene(s) and the marker(s), and the number of target genes to be selected (Francia et al. 2005). The efficiency of MAS can be enhanced if the markers used are more closely linked to major genes or target loci (Liu et al. 2007b). It should be noted that, in any event, identification and validation of markers tightly linked to gene(s) controlling the target trait must be done prior to MAS. It is evident that large-scale MAS has been successful in many major cereals, notably wheat, maize, and rice (Liu et al. 2007b; Gupta et al. 2010; Miah et al. 2013). Unfortunately, this is not the case for most of the underutilised crops (Kumar et al. 2011). Although the pace of development of markers and genomic resources has been hastened in some potential underutilised crops as discussed in Sect. 3.3, the overall progress in the use of MAS as part of their breeding programmes has been limited. To this end, researchers are urged to start employing the developed resources for improving underutilised crops, and making MAS an integral part of the current and future breeding programmes (Kumar et al. 2012). As for the less developed underutilised crops that have enormous potential, the winged bean being a notable example (Wong et al. 2016), the development of adequate genetics and genomic resources is a crucial first step towards their genetic improvement (Abdullah Bamba et al. 2015a, b).

On a positive note, the potential benefits and application of genomics resources developed in model and/or major plant species has been explored in some underutilized species. One of the approaches for cross-species study is through cross-hybridisation of heterologous nucleic acids from target species, i.e. underutilized crops, onto the Affymetrix oligonucleotide based microarray of a reference species, i.e. major crop (http://affy.arabidopsis.info/xspecies/). Several studies have reported the use of cross-species microarray, for instance, the development of more than 1000 single feature polymorphisms (SFPs) in cowpea through detection and validation using soybean genome array (Das et al. 2008). In general, the application of knowledge from model and/or major crops to underutilised crops may be effective in the following areas (Liu et al. 2007b):

Analyses of crop genetic diversity

Identification of useful alleles or perfect molecular markers

Development of new or improved varieties with specific allele integration using MAS schemes

Cloning and transfer of desirable alleles or traits among taxa

We are certain that the ever-decreasing cost of marker technologies, and availability of fully automated high-throughput approaches in this post-genomics era will contribute immensely to the research for underutilised crop improvement.

3.5 Conclusions and Future Prospects

Biodiversity loss, population growth and climate change are all interconnected and interdependent challenges facing global agriculture today. There is, in fact, no single recipe to ensure global food security in light of these challenges. From an agricultural perspective, diversification of global food sources and systems with underutilised crops is one of the most sensible means to deal with hunger and malnutrition in the climate change era. The slow progress made in the utilisation of molecular technology in a number of underutilised crops with high potential has been attributed to the limited efforts and finite resources dedicated to developing their genomic resources. Greater emphasis should be placed on generating a sophisticated molecular markers and genomic resources for each of these crops, which will lay a critical foundation for their future molecular breeding efforts.

References

Abdullah Bamba SA, Massawe F (2013) Microsatellites based marker molecular analysis of Ghanaian bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces alongside morphological characterization. Genet Resour Crop Evol 60:777–787

Abdullah Bamba SA, Massawe F, Mayes S (2015a) Beyond landraces: developing improved germplasm resources for underutilized species—a case for Bambara groundnut. Biotechnol Genet Eng Rev 30:37–41. doi:10.1080/02648725.2014.992625

Abraham B, Araya H, Berhe T, Edwards S, Gujja B, Khadka RB, Sen D, Koma Y, Sharif A, Styger E, Uphoff N, Verma A (2014) The system of crop intensification: reports from the field on improving agricultural production, food security, and resilience to climate change for multiple crops. Agric Food Sci 3:1–12

Ahmad SA, Redjeki ES, Ho WK, Aliyu S, Mayes K, Massawe F, Kilian A, Mayes S (2016) Construction of a genetic linkage map and QTL analysis in bambara groundnut. Gen 59:459– 472. doi:10.1139/gen-2015-0153

Arias RS, Borrone JW, Tondo CL, Kuhn DN, Irish BM, Schnell RJ (2012) Genomics of tropical fruit tree crops. In: Schnell RJ, Priyadarshan PM (eds) Genomics of tree crops. Springer Science, New York, pp 209–239

Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH et al (2010) A brief review of molecular techniques to assess plant diversity. Int J Mol Sci 11(5):2079–2096

Babu BK, Dinesh P, Agrawal PK, Sood S, Chandrashekara C, Bhatt JC, Kumar A (2014) Comparative genomics and association mapping approaches for blast resistant genes in finger millet using SSRs. PLoS One 9(6):e99182

Bardakci F (2000) Random amplified polymorphic DNA (RAPD) markers. Turk J Biol 25:185–196

Basu S, Roberts JA, Azam-Ali SN, Mayes S (2007) Development of microsatellite markers for bambara groundnut (*Vigna subterranea* L. Verdc.)—an underutilized African legume crop species. Mol Ecol Notes 7:1326–1328

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep P, Feng L, Vaughn JN et al (2012) Reference genome sequence of the model plant Setaria. Nat Biotechnol 30(6):555–561

Bisen PS (2014) Laboratory protocols in applied life sciences. CRC Press, Boca Raton

Bohra A, Pandey MK, Jha UC, Singh B, Singh IP, Datta D, Chaturvedi SK, Nadarajan N, Varshney RK (2014) Genomics-assisted breeding in four major pulse crops of developing countries:

present status and prospects. Theor Appl Genet 127(6):1263–1291

Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenomics J 3:77–96

Chen H, Chen X, Tian J, Yang Y, Liu Z, Hao X et al (2015) Development of gene-based SSR markers in rice bean (*Vigna umbellata* L.) based on transcriptome data. PLoS One 11(3):e0151040

Chen HL, Chen X, Tian J, Yang Y, Liu Z, Hao X et al (2016) Development of gene-based SSR markers in rice bean (*Vigna umbellata* L.) based on transcriptome data. PLoS One 11(3):e0151040

Cheng A, Ismanizan I, Mohamad O, Habibuddin H (2012) Simple and rapid molecular techniques for identification of amylose levels in rice varieties. Int J Mol Sci 13:6156–6166

Cheng A, Mayes S, Dalle G, Demissew S, Massawe F (2015) Diversifying crops for food and nutrition security—a case

of teff. Biol Rev 92:188–198
Christensen SA, Pratt DB, Pratt C, Nelson PT, Stevens MR, Jellen EN et al (2007) Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (*Chenopodium quinoa* Willd.) using microsatellite markers. Plant Gen Res 5:82–95 Coles ND, Coleman CE, Christensen SA, Jellen EN, Stevens MR, Bonifacio A, Rojas-Beltran JA,
Fairbanks DJ, Maughan PJ (2005) Development and use of an expressed sequenced tag library in quinoa (*Chenopodium quinoa* Willd.) for the discovery of single nucleotide polymorphisms.
Plant Sci 168(2):439–447 Cruz VMV, Kilian A, Dierig DA (2013) Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop Lesquerella and related species. PLoS One 8:e64062
Dane F, Liu J, Zhang C (2006) Phylogeography of the bitter apple, *Citrullus colocynthis*. Genet Resour Crop Evol 54:327–336
Dansi A, Mignouna HD, Zoudjihekpon J, Sangare A, Asiedu R, Ahoussou N (2000) Identification of some Benin Republic's Guinea yam cultivars using randomly amplified polymorphic DNA. Genet Resourc Crop Evol 47:619–662
Das AK, Anjaneluyu ASR, Gadekar YP, Singh RP, Pragati H (2008) Effect of full-fat soy paste and textured soy granules on quality and shelf-life of goat meat nuggets in frozen storage. Meat Sci 80:607–614
Dawson IK, Guarino L, Jaenicke H (2007) Underutilised plant species: impacts of promotion on biodiversity. Position Paper No. 2. ICUC, Colombo, Sri Lanka
De La Fuente GN, Frei UK, Lubberstedt T (2013) Accelerating plant breeding. Trends Plant Sci 18:667–672
Dhanasekar P, Dhumal KN, Reddy KS (2010) Identification of RAPD markers linked to plant type gene in pigeonpea. Indian J Biotechnol 9:58–63
Ebert AW (2014) Potential of underutilized traditional vegetables and legume crops to contribute to food and nutritional security, income and more sustainable production systems. Sustainability 6:319–335
Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758–760
Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132
Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Valè G (2005) Marker assisted selection in crop plants. Plant Cell Tissue Org 82:317–342
Fu D, Ma B, Mason A, An Z (2013) MicroRNA-based molecular markers: a novel PCR-based genotyping technique in brassica species. Plant Breed 132(4):75–81
Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217
Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G et al (2012) The genome of melon (*Cucumis melo* L.) Proc Natl Acad Sci U S A 109(29):11872–11877
Gonthier P, Sillo F, Lagostina E, Roccotelli A, Cacciola OS, Stenlid J, Garbelotto M (2015) Selection processes in simple sequence repeats suggest a correlation with their genomic location: insights from a fungal model system. BMC Genomics 16:1107
Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int 2015:431–487
Gupta PK, Kumar J, Mir RR, Kumar A (2010) Marker assisted selection as a component of conventional plant breeding. Plant Breed Rev 33:145–217
Harel-Beja R, Tzuri G, Portnoy V, Lotan-Pompan M, Lev S et al (2010) A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. Theor Appl Genet 121(3):511–533
Ho WK, Muchugi A, Muthemba S, Kariba R, Mavenkeni BO, Hendre P, Song B, Deynze V, Massawe F, Mayes S (2016) Use of microsatellite markers for the assessment of bambara groundnut breeding system and varietal purity before genome sequencing. Genome 59(6):427–431
Huang CW, Lin YT, Ding ST, Lo LL, Wang PH, Lin EC, Liu FW, Lu YW (2015) Efficient SNP discovery by combining microarray and lab-on-a-chip data for animal breeding and selection. Microarrays 4:570–595
http://rebase.neb.com/rebase/rebase.html
http://seedsofdiscovery.org/about/genotyping-platform/
http://www.cropsforthefuture.org/
Hunt HV, Campana MG, Lawes MC, Park Y-J, Bower MA, Howe CJ et al (2011) Genetic diversity and phylogeography of broomcorn millet (*Panicum miliaceum* L.) across Eurasia. Mol Ecol 20:4756–4771
Hwang J, Ang JK, Son B, Kim K, Park Y (2011) Genetic diversity in watermelon cultivars and related species based on AFLPs and EST-SSRs. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 39(2):285
Iranjo P, Nabati Ahmadi D, Sorkheh K, Memeari HR, Ercisli S (2016) Genetic diversity and phylogenetic relationships between and within wild *Pistacia* species populations and implications for its conservation. J Forestry Res 27(3):685–697
Jaccoud DK, Peng D, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information

independent genotyping. Nucleic Acids Res 29(4):e25
Jarvis DE, Kopp OR, Jellen EN, Mallory MA, Pattee J, Bonifacio A, Coleman CE, Stevens MR, Fairbanks DJ, Maughan PJ (2008) Simple sequence repeat marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.) J Genet 87:39–51
Jarvis DE, Ho YS, Lightfoot DJ, Schmockel SM, Li B et al (2017) The genome of *Chenopodium quinoa*. Nature 542(7641):307–312
Ji Q, Xu X, Wang K (2013) Genetic transformation of major cereal crops. Int J Dev Biol 57:495–508
Jia J, Zhou S, Kong X et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496(7443):91–95
Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F et al (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390
Joshi SP, Ranjekar PK, Gupta VS (1999) Molecular markers in plant genome analysis. Crop Sci 16:102–105
Kesawat MS, Kumar BD (2009) Molecular markers: it's application in crop improvement. J Crop Sci Biotechnol 12:169–181
Khoury CK, Bjorkman AD, Dempewolf H, Ramirez-villegas J, Guarino L, Jarvis A, Rieseberg LH, Struik PC (2014) Increasing homogeneity in global food supplies and the implications for food security. Proc Natl Acad Sci U S A 111:4001–4006
Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009) Potential of molecular markers in plant biotechnology. Plant Omics J 2(4):141–162
Kumar J, Choudhary AK, Solanki RK, Pratap A (2011) Towards marker-assisted selection in pulses: a review. Plant Breed 130:297–313
Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics 2012:831460. doi:10.1155/2012/831460
Kumari K, Pande A (2010) Study of genetic diversity in finger millet (*Eleusine coracana* L. Gaertn) using RAPD markers. Afr J Biotechnol 9(29):4542–4549
Kumpatla SP, Buyyarapu R, Abdurakhmonov IY, Mammadov JA (2012) Genomics-assisted plant breeding in the 21st century: technological advances and progress. Plant Breed:131–184
Lai PS, Ho WS, Pang SL (2013) Development, characterization and cross-species transferability of expressed sequence tag-simple sequence repeat (EST-SSR) markers derived from kelampayan tree transcriptome. Biotechnology 12:225–235
Levi A, Thomas CE, Keinath AP, Wehner TC (2001) Genetic diversity among watermelon (*Citrullus lanatus* and *Citrullus colocynthis*) accessions. Gen Res Crop Evol 48:559
Lin WH, Kussell E (2012) Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. Nucleic Acids Res 40(6):2399–2413
Lin J, Kuo J, Ma J (1996) A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. Nucleic Acids Res 24(18):3649–3650
Liu P, Yang YS, Hao CY, Guo WD (2007a) Ecological risk assessment using RAPD and distribution pattern of a rare and endangered species. Chemosphere 68:1497–1505
Liu S, Banik M, Yu K, Park SJ, Poysa V, Guan Y (2007b) Marker-assisted election (MAS) in major cereal and legume crop breeding: current progress and future directions. Int J Plant Breed 1:74–88
Ma JQ, Yao MZ, Ma CL, Wang XC, Jin JQ, Wang XM, Chen L (2014) Construction of a SSR-based genetic map and identification of QTLs for Catechins content in tea plant *(Camellia sinensis*). PLoS One 9(3):e93131
Massawe FJ, Dickinson M, Roberts JA, Azam-Ali SN (2002) Genetic diversity in bambara groundnut landraces (*Vigna subterranea* (L.) Verdc) revealed by AFLP markers. Gen 45:1175–1180
Massawe FJ, Roberts JA, Azam-Ali SN, Davey MR (2003) Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc) landraces assessed by random amplified polymorphic DNA (RAPD) markers. Genet Resour Crop Evol 50:737–741
Massawe F, Mwale SS, Azam-Ali SN, Roberts JA (2005) Breeding in bambara groundnut (*Vigna subterranea* (L.) Verdc.): strategic considerations. Afr J Biotechnol 4(6):463–471
Massawe F, Mayes S, Cheng A (2016) Crop diversity: an unexploited treasure trove for food security. Trends Plant Sci 21(5):365–368
Masters WA (2010) Africa's turnaround: from crisis to opportunity in African agriculture. In: Food and financial crises: impacts on Sub-Saharan Africa. CAB International, Wallingford
Maughan PJ, Bonifacio A, Jellen EN, Stevens MR, Coleman CE, Ricks M et al (2004) A genetic linkage map of quinoa (*Chenopodium Quinoa*) based on AFLP, RAPD, and SSR markers. Theor Appl Genet 109(6):1188–1195
Maughan PJ, Yourstone SM, Jellen EN, Udall JA (2009) SNP discovery via genomic reduction, barcoding and 454-pyro-sequencing in amaranth. Plant Gen 2:260–270
Maughan PJ, Smith SM, Jellen EN, Fairbanks D (2011) Development, characterization and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.)
Plant Gen 4:1–10 Maughan PJ, Smith SM, Rojas-Beltran JA, Elzinga D, Raney JA, Jellen EN, Bonifacio A, Udall

JA, Fairbanks DJ (2012) Single nucleotide polymorphism identification, characterization, and linkage mapping in quinoa. Plant Genome 5:114–125
Mayes S, Massawe F, Anderson PG, Roberts JA, Azam-Ali SN, Hermann M (2011) The potential of underutilized crops to improve security of food production. J Exp Bot 63:1075–1079
Mba C, Tohme J (2005) Use of AFLP markers in surveys of plant diversity. Methods Enzymol 395:177–201
Meksem K, Ruben E, Hyten D, Triwitayakorn K, Lightfoot DA (2001) Conversion of AFLP bands into high-throughput DNA markers. Mol Genet Genomics 265(2):207–214
Melo ATO, Bartaula R, Hale L (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by- sequencing data. BMC Bioinformatics 17:29
Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KN, Latif MA (2013) A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. Int J Mol Sci 14(11):22499–22528
Mignouna HD, Abang Asiedu R (2007) Advances in yam (*Dioscorea* spp.) genetics and genomics proceedings of the 13th ISTRC symposium, pp 72–81
Mignouna HD, Ellis NTH, Asiedu R, Ng QN (1998) Analysis of genetic diversity in Guinea yams (*Dioscorea* spp) using AFLP fingerprinting. Trop Agric (Trinidad) 75:224–229
Mignouna HD, Mank RA, Ellis THN, van den Bosch N et al (2002a) A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. Theor Appl Genet 105:726–735
Mignouna HD, Mank RA, Ellis THN et al (2002b) A genetic linkage map of Guinea yam (*Dioscorea rotundata* L.) based on AFLP markers. Theor Appl Genet 105:716–725
Mignouna HD, Abang MM, Fagbemi SA (2003) A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata* Poir.) germplasm characterisation. Ann Appl Biol 142:269–276
Mochida K, Shinozaki K (2013) Unlocking Triticeae genomics to sustainably feed the future. Plant Cell Physiol 54(12):1931–1950
Moe KT, Gwag JG, Park YJ (2012) Trends in genomics and molecular marker systems for the development of some underutilized crops. Gen Genom 34(5):451–466
Molosiwa OO, Aliyu S, Stadler F, Mayes K, Massawe F, Kilian A, Mayes S (2015) SSR marker development, genetic diversity and population structure analysis of Bambara groundnut [*Vigna subterranea* (L.) Verdc.] landraces. Gen Res Crop Evol 62(8):1225–1243
Narina SS, Buyyarapu R, Kottapalli KR, Sartie AM, Ali MI, Robert A et al (2011) Generation and analysis o expressed sequence tags (ESTs) for marker development in yam (*Dioscorea alata* L.) BMC Genomics 12:100
Naylor RL, Falcon WP, Goodman RM, Jahn MM, Sengooba T, Tefera H, Nelson RJ (2004) Biotechnology in the developing world: a case for increased investments in orphan crops. Food Policy 29:15–44
Ngwira AR, Thierfelder C, Lambert DM (2012) Conservation agriculture systems for Malawian smallholder farmers: longterm effects on crop productivity, profitability and soil quality. Renew Agric Food Syst 28(4):350–363
Nimmakayala P, Tomason YR, Jeong J, Ponniah SK, Karunathilake A, Levi A, Perumal R, Levi UK (2010) Genetic reticulation and interrelationships among citrullus species as revealed by joint analysis of shared AFLPs and species-specific SSR alleles. Plant Gen Res 8(1):16–25
Ojiewo C, Tenkouano A, Oluoch M, Yang R (2010) The role of AVRDC—the world vegetable centre in vegetable value chains. Afr J Hort Sci 3:1–23
Padulosi S, Hoeschle-Zeledon I (2004) Underutilized plant species: what are they? LEISA 20(1):56
Petro D, Onyeka TJ, Etienne S, Rubens S (2011) An intraspecific genetic map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. Euphytica 179:405–416
Phung NTP, Mai CD, Mournet P, Frouin J, Droc G, Ta NK, Jouannic S, Lê LT, Do VN, Gantet P, Courtois B (2014) Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. BMC Plant Biol 14:371
Pingali PL (2012) Green revolution: impacts, limits, and the path ahead. Proc Natl Acad Sci U S A 109:12302–12308
Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1(7):215–222
Rahman MM, Aminul MI, Sofian MA, Amru NB (2014) Tropical legume crop rotation and nitrogen fertilizer effects on agronomic and nitrogen efficiency of rice. Sci World J. doi:10.1155/2014/490841
Ranamukhaarachchi DG, Kane ME, Guy CL, Li QB (2000) Modified AFLP technique for rapid genetic characterization in plants. Biotechniques 29(4):858–859
Raney JA, Reynolds DJ, Elzinga DB, Page J, Udall JA, Jellen EN, Bonfacio A, Fairbanks D, Maughan PJ (2014) Transcriptome analysis of drought induced stress in *Chenopodium quinoa*. Am J Plant Sci 5:338–357
Raveendar S, Lee GA, Jeon YA, Lee YJ, Lee JR, Cho GT, Cho JH, Park JH, Ma KH, Chung JW (2015) Cross-amplification of *Vicia sativa* subsp. sativa microsatellites across 22 other *Vicia* species. Molecules 20(1):1543–1550
Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050.

PLoS One 8:1–8
Ren X, Wang J, Liu L, Sun G, Li C, Luo H, Sun D (2016) SNP-based high density genetic map and mapping of btwd1 dwarfing gene in barley. Sci Rep 6:31741. doi:10.1038/srep31741
Roberts RJ, Vincze T, Posfai J, Macelis D (2003) REBASE: restriction enzymes and methyltransferases. Nucleic Acids Res 31(1):418–420
Rubinstein M, Katzenellenbogen M, Eshed R, Rozen A, Katzir N, Colle M et al (2015) Ultrahigh-density linkage map for cultivated cucumber (*Cucumis sativus* L.) using a single-nucleotide polymorphism genotyping array. PLoS One 10(4):e0124101
Saeed B, Baranwal VK, Khurana P (2016) Comparative transcriptomics and comprehensive marker resource development in mulberry. BMC Genomics 17:98
Sánchez-Sevilla JF, Horvath A, Botella MA, Gaston A, Folta K, Kilian A, Denoyes B, Amaya I (2015) Diversity arrays technology (DArT) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (Fragaria × ananassa). PLoS One 10(12):e0144960
Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proc 5(Suppl 7):54
Saski CA, Bhattacharjee R, Scheffler BE, Asiedu R (2015) Genomic resources for water yam 55(*Dioscorea alata* L.): analyses of EST-sequences, de novo sequencing and GBS libraries. PLoS One 10(7):e0134031
Schlotterer C (2004) The evolution of molecular markers—just a matter of fashion? Nat Rev Genet 5:63–69
Semagn K, Bjørnstad A, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. Afr J Biotechnol 5(25):2540–2568
Shi W, Tao F (2014) Vulnerability of African maize yield to climate change and variability during 1961–2010. Food Secur 6:471–481
Si Y, Zhang C, Meng S, Dane F (2009) Gene expression changes in response to drought stress in *Citrullus colocynthis*. Plant Cell Rep 28:997–1009
Suresh S, Chung JW, Cho GT, Sung JS, Park JH, Gwag JG, Baek HJ (2014) Analysis of molecular genetic diversity and population structure in *Amaranthus* germplasm using SSR markers. PlantBiosyst 148(4):635–644
Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Nat Biotechnol 7:257–264
Tártara SMC, Manifesto MM, Bramardi SJ, Bertero HD (2012) Genetic structure in cultivated quinoa (*Chenopodium quinoa* Willd.), a reflection of landscape structure in Northwest Argentina. Conserv Genet 13:1027-1038
Thies E (2000) Promising and underutilized species, crops and breeds. GTZ, Eschborn, Germany
Tomar RS (2010) Molecular markers and plant biotechnology. New India Publishing, New Delhi
Troconis-Torres IG, Rojas-López M, Hernández-Rodríguez C, Villa-Tanaca L, Maldonado-Mendoza
IE et al (2002) Biochemical and molecular analysis of some commercial samples of chilli peppers from Mexico. J Biomed Biotechnol 2012:873090
Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55
Varshney RK, Chen W, Li Y et al (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30(1):83–89
Vatanparast M, Shetty P, Chopra R, Doyle JJ, Sathyanarayana N, Egan AN (2016) Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Sci Rep 6:29070. doi:10.1038/srep29070
Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F et al (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45(5):487–494
Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414
Wang Z, Hu H, Goertzen LR, McElroy JS, Dane F (2014) Analysis of the *Citrullus colocynthis* transcriptome during water deficit stress. PLoS One 9(8):e104657
Weising K, Nybom H, Pfenninger M, Wolff K, Meyer W (1994) DNA fingerprinting in plants and fungi. CRC Press, Boca Raton
Williams JT, Haq N (2002) Global research on underutilized crops. An assessment of current activities and proposals for enhanced cooperation. ICUC, Southampton
Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535
Wong QN, Massawe F, Mayes K, Blythe M (2016) Molecular genetic tools to support genetic improvement of winged bean (*Psophocarpus tetragonolobus*) for food and nutrition security. Acta Hortic 1110:1–6
Wong QN, Tanzi AS, Ho WK, Malla S, Blythe M et al (2017) Development of gene based SSR markers in winged bean (*Psophocarpus tetragonolobus* (L.) DC.) for diver-sity assessment. Genes 8(3):100

Yu H, Li Q (2007) Development of EST-SSRs in the Mediterranean blue mussel, Mytilus galloprovincialis. Mol Ecol Notes 7(6):1308–1310
Zeid M, Belay G, Mulkey S, Poland J, Sorrells ME (2011) QTL mapping for yield and lodging resistance in an enhanced SSR-based map for tef. Theor Appl Genet 122:77–93
Zeid M, Assefa K, Haddis A, Chanyalew S, Sorrells ME (2012) Genetic diversity in tef (*Eragrostis tef*) germplasm using SSR markers. Field Crop Res 127:64–70
Zhao Y, Williams R, Prakash CS, He G (2012) Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.) BMC Plant Biol 12:237

Section C) Aims and objectives of this study

Based on an extensive review of literature on winged bean past research, and identification of potential future breeding targets, the present investigation set the following aims and objectives:

1. To generate molecular tools for winged bean, and apply them for genetic diversity, breeding system analysis, and linkage analysis on winged bean:
    - Development of a validated set of Simple Sequence Repeat (SSR);
    - Discovery of Single Nucleotide Polymorphisms (SNPs) through genotyping-by-sequencing (GbS) technologies;
    - Generation of a genetic linkage map.

2. To understand growth and development in winged bean; to understand plant architecture, development, and yield-related traits, and to reveal how such traits correlate between each other:
    - Selection of parental genotypes, and generation of hybrids through controlled crosses;
    - Field assessment of F_1 and F_2 generation, and phenotypic analysis for traits related to plant architecture, development, and yield.

3. To reveal the genetic basis of traits underlying plant architecture, development, and yield components:
    - Genotyping of the F_2 mapping population using Diversity Array Technology sequencing (DArTseq);
    - Construction of genetic linkage map upon F_2 mapping population DArTseq data;
    - Performing Quantitative Trait Locus (QTL) analysis.

4. To assess the level of genetic diversity in germplasm from private and public collections, for selection of material to include in future improvement programmes:
    - Distance and frequency analyses of accessions coming from public and private collections using SNPs discovered through DArTseq.

Chapter 2 - Development of SSR markers in winged bean

Wong, Q., **Tanzi, A. S.**, Ho, W., Malla, S., Blythe, M., Karunaratne, A., Massawe, F., Mayes, S. (2017). Development of Gene-Based SSR Markers in Winged Bean (Psophocarpus tetragonolobus (L.) DC.) for Diversity Assessment. *Genes*, *8*(3), 100. https://doi.org/10.3390/genes8030100

Published research paper, presenting the development and validation of a set of SSR markers in winged bean.

The chapter follows the format used in the publication.

Communication

Development of gene-based SSR markers in winged bean [*Psophocarpus tetragonolobus* (L.) DC.] for diversity assessment

Quin Nee Wong[1], Alberto Stefano Tanzi [1,2], Wai Kuan Ho[1,2], Sunir Malla[3], Martin Blythe[3], Asha Karunaratne[2,4], Festo Massawe[1,2] and Sean Mayes[1,2,5]*

[1] Biotechnology Research Centre, School of Biosciences, Faculty of Science, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia; khyx4asi@nottingham.edu.my (A.S.T.); festo.massawe@nottingham.edu.my (F.M.)
[2] Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan; waikuan@cffresearch.org (W.K.H.); asha.karunaratne@cffresearch.org (A.K.)
[3] Deep Seq, Faculty of Medicine and Health Sciences, Queen's Medical Centre, University of Nottingham, Nottingham, NG7 2UH, U. K; sunir.malla@nottingham.ac.uk (S.M.); martinblythe@hotmail.com (M.B.)
[4] Department of Export Agriculture, Faculty of Agricultural Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, 70140 Sri Lanka
[5] School of Biosciences, Faculty of Science, University of Nottingham Sutton Bonington Campus, Sutton Bonington, Leicestershire, LE12 5RD, U.K.
* Correspondence: S.Ms., Sean.Mayes@nottingham.ac.uk; Tel.: +44 115 951 8062

Academic Editor: J. Peter W. Young

Received: 20 December 2016; Accepted: 3 March 2017; Published: 9 March 2017

Abstract: Winged bean (*Psophocarpus tetragonolobus*) is an herbaceous multipurpose legume grown in hot and humid countries as a pulse, vegetable (leaves and pods) or root tuber crop depending on local consumption preferences. In addition to its different nutrient-rich edible parts which could contribute to food and nutritional security, it is an efficient nitrogen fixer as a component of sustainable agricultural systems. Generating genetic resources and improved lines would help to accelerate the breeding improvement of this crop, as the lack of improved cultivars adapted to specific environments has been one of the limitations preventing wider use. A transcriptomic *de novo* assembly was constructed from four tissues; leaf, root, pod and reproductive tissues from Malaysian accessions, comprising of 198,554 contigs with a N50 of 1,462 bp. Of these, 138,958 (70.0%) could be annotated. Among 9,682 genic SSR repeat motifs identified (excluding monomer repeats), trinucleotide-repeats were the most abundant (4,855), followed by di-nucleotide (4,500) repeats. A total of 18 SSR markers targeting di- and tri-nucleotide repeats have been validated as polymorphic markers based on an initial assessment of nine genotypes originated from five countries. A cluster analysis revealed provisional clusters among this limited, yet diverse selection of germplasm. The developed assembly and validated genic SSRs in this study provide a foundation for a better understanding of the plant breeding system for the genetic improvement of winged bean.

Keywords: *Psophocarpus tetragonolobus* (L.) DC; winged bean; SSR marker; transcriptome

1. Introduction

Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) ($2n=2x=18$) is a tropical perennial vine species, classified in the family of Fabaceae and subfamily of Papilionoideae, cultivated mainly at a subsistence scale in hot and humid countries across India, Southeast Asia, and the Western Pacific islands, with a presence in a number of African countries, as well [1-5]. It is grown for its green pods, tuberous roots and mature seeds, all

of which have received attention for their nutritional content in the past, as comprehensively described in '*The Winged Bean – A high-protein crop for the tropics'* from the National Academy of Science in 1981 [1]. Initial interest was drawn to high crude protein levels in seeds, which are comparable to soybean [6-8]. Its vining nature and nitrogen fixation activity have seen it used as a cover crop and also incorporated into rotation or intercropping systems [9-11]. As such, winged bean could be a good candidate for diversifying diets to improve nutritional security, based on complex and more sustainable agricultural systems [12]. Despite its potential, winged bean has received limited research investment for developing molecular tools that can support breeding programmes, until recently. Recent reports include the development of inter-Simple Sequence Repeats (iSSRs) and Randomly Amplified Polymorphic DNA (RAPD) markers for genetic diversity and for clonal fidelity analyses and two small transcriptomic assemblies derived from a mix of leaf, bud and shoot of Sri Lankan accessions and leaf tissue from a Nigerian accession, respectively [13-17]. Given that winged bean is believed to be largely self-pollinated, heterozygosity would be expected to be low, although a formal assessment is needed and the species does produce large flowers, suggesting a contribution from insect pollination, as recorded by Erskine [18]. Thus, molecular breeding will facilitate utilisation of genetic resources in winged bean breeding, especially among accessions, through combining beneficial traits. Molecular markers that are tightly linked to important agronomic traits are a precondition for undertaking molecular breeding in plants. The genetic basis of traits in winged bean remains largely unexplored and to date there has not been any genetic linkage map reported for this crop, although controlled crosses have been reported [19-22].

In this study, we generated RNA-seq data from four tissues (leaf, root, reproductive tissues and pod) of six locally grown accessions, followed by the identification of SSR-containing sequences and validation of a subset of genic SSR markers. To our knowledge, this is the first application of within-species genic SSR markers in winged bean accessions. The data will help to begin the development of comprehensive genetic information and tools to facilitate future breeding programmes as well as allow the levels of natural inbreeding to be determined, to allow appropriate breeding schemes to be devised. The transcriptome will allow us to gain a better understanding of the phylogenetic relationships between winged bean and other leguminous and model plants.

2. Material and Methods

2.1 Plant material, RNA extraction, cDNA library construction and sequencing

A total of six locally grown winged bean accessions [two derived from Malaysian Agricultural Research and Development Institute (MARDI) and four from local planters] were grown from Aug to Dec 2012 at Lady Bird Farm, Broga, Semenyih, Malaysia (Latitude: 20 57′ °N, Longitude: 1010 50′6°: Altitude: 45 meters). RNA was extracted separately from leaf, root, pod and reproductive tissue (comprising of bud and flower) by pooling the respective tissues from all the six accessions. Extraction was performed from different tissue groups separately using TRIzol® Reagent (Thermo Fisher Scientific) followed by another round of purification using RNeasy MinElute Cleanup Kit (Qiagen) before library preparation.

Total RNA was measured using the Qubit RNA BR assay kit (ThermoFisher). A total of 5μg of RNA was used for enrichment of mRNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). RNA fragmentation was done using NEBNext Magnesium RNA Fragmentation Module (NEB). Illumina stranded whole transcriptome sequencing libraries were prepared through a dUTP approach using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). Libraries were gel purified using 2% E-Gel SizeSelect (ThermoFisher) and QC was performed using bioanalyser HS kit (Agilent biotechnologies). Quantification was done using qPCR (Kapa Biosystems). Equimolar amounts of barcoded libraries were mixed and subjected to 250 bp paired-end run using MiSeq V2 chemistry (Illumina) on Illumina MiSeq sequencing platform according to manufacturer's instruction.

For SSR marker development, a total of nine plants from five accessions, each one representing a geographical origin, from IITA genebank and MARDI, as listed in Table 1 below, were used. Two individuals

were used per accession, which were collected after a cycle of single seed descent purification from Jan to Jun 2013 at the Lady Bird Farm, except for the Malaysian line.

Table 3. Winged bean accessions (two individuals per origin except Malaysian line) used and their origins.

Individuals	Origin
Tpt53-9-8 Tpt53-9-10	Bangladesh
Tpt17-6-3 Tpt17-6-8	Indonesia
M3-3	Malaysia
Tpt10-7-5 Tpt10-7-7	Papua New Guinea
SLS319-10-3 SLS319-10-4	Sri Lanka

2.2 De novo transcriptome assembly and microsatellite identification

Adaptors and low quality reads (below Q20) were trimmed using Scythe (https://github.com/vsbuffalo/scythe) and Sickle (https://github.com/najoshi/sickle) [23], respectively, using default settings. Trimmed reads from all tissues were pooled to assemble a combined de novo assembly with Trinity version 2.2.0 pipeline (Release 2013-02-25) using the strand specificity option. Trinity assembled transcripts were annotated with Trinotate software suite version 1.1 (https://trinotate.github.io) [24], with a blast e-value threshold of 1 x 10-5 from NCBI-BLAST [25], HMMER/PFAM [26], SignalP [27], EMBL eggNOG [28] and GO [29] databases. The data is deposited in NCBI Sequence Read Archive (BioProject ID PRJNA374598) under the accession umber of SRP099538; SRR5252646 (root), SRR5252647 (reproductive tissue), SRR5252648 (pod) and SRR5252649 (leaf).

This was followed by the identification of microsatellites using MISA (http://pgrc.ipk-gatersleben.de/misa/), based on a minimum number of repeats of six for di-, five for tri-, tetra-, penta- and hexa-nucleotide repeat motif (monomer repeats were excluded) whilst the maximum number of bases interrupting two SSRs in a compound microsatellite was 100 [30].

2.3 Microsatellite markers development and scoring

Primer pairs were designed from sequences harbouring a minimum of 18-bases long microsatellites, i.e. minimum 9 and 6 repetitions for di- and tri-nucleotide motifs respectively (Supplementary Files, Table S1). Primer3 [31] and PrimerQuest (IDT) were used for oligo design, with the latter used whenever the first was not able to design with standard parameters for the given sequence. Where possible, two pairs of primers were designed for a single target region, so as to have an alternative if the first pair failed to amplify the target region. DNA was extracted from the leaf of nine genotypes (Table 1) using a modified CTAB method [32]. In addition, an RNase digestion step was added and 1 volume of isopropanol was used instead of 2/3 volume. Primer screening and optimisation was carried out in a three primer system for fluorescent labelling [33] using an equimolar mixture of genotypes. Each 20µl of PCR reaction consisted of 1X Buffer S, 200µM dNTPs, 0.02µM forward primer, 0.18µM M13-dye labelled primer, 0.2µM reverse primer, 20ng of DNA and 1U of Taq DNA Polymerase (Vivantis). The PCR was programmed for 3 min of initial denaturation at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 60°C and 1 min at 72°C with a 10 min final elongation. Initial polymorphism evaluation of primers was performed across all genotypes using a long capillary fragment analyser (AATI) with default parameters except for using 4µl of samples, 4kV of separation voltage for 180 min per run. Data

was then analysed with Advance Analytical PROSize® 2.0 v1.3.1.1 to identify potential SSR markers based on presence of polymorphic amplicons across genotypes.

PCR products of single genotypes from potentially polymorphic SSR markers were then separated using an ABI Genetic Analyser ABI3730XL using Peak Scanner v2 for scoring. Validated markers were subsequently characterised using Power Marker v3.5 [34] for major allelic frequency, alleles per marker, heterozygosity, and polymorphic information content (PIC).

2.4 Cluster analysis

A hierarchical cluster analysis was performed with the Dice (also known as Nei and Li) similarity coefficient and UPGMA algorithm in Genstat 18th Edition [35].

3. Results and discussion

3.1 Transcriptome assembly and in silico identification of microsatellites

A total of four libraries generated 12.77 million reads of cleaned 250bp paired ends (Table 2). The de novo assembly derived from all tissues produced a total of 198,554 contigs with an average size of 798bp and an N50 of 1462bp.

Table 2. Summary statistics for the *de novo* assembled transcriptome

Tissue	**Leaf**	**Pod**	**Reproductive tissue**	**Root**
Number of raw read/base (bp)	3,150,356 / 1,544,004,822	3,973,092 / 1,868,456,680	3,544,968 / 1,719,303,632	3,873,893 / 1,859,527,511
Numbers of trimmed read/base (bp)	3,113,502 / 1,438,258,180	3,157,832 / 1,301,766,113	3,199,527 / 1,431,024,141	3,303,324 / 1,461,908,151
Number of contigs/base (bp)	198,554 / 158,382,439			
Average contig size (bp)	798			
N50	1,462			

Out of 198,554 contigs, 138,958 (70.0%) could be annotated. Among them, 75,308 (54.2%), 69,172 (49.8%), 6,499 (4.7%), 60,040 (43.2%) and 70,069 (50.4%) were found in NCBI-BLAST, HMMER/PFAM, SignalP, EMBL eggNOG and GO databases, respectively, with no significant homology found from tmHMM on the prediction of transmembrane helices (Supplementary File, Table S2). Fig. 1 illustrates the abundance of transcripts classified based on gene ontology.

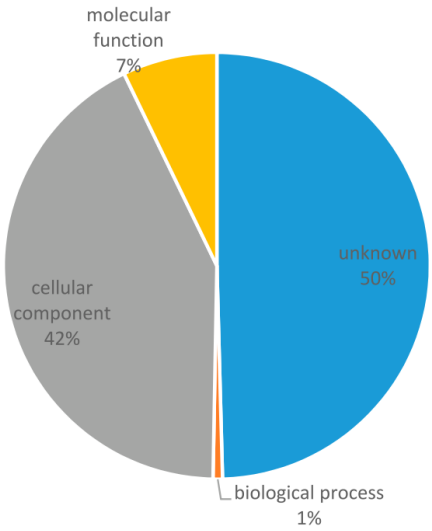


Figure 1. Distribution of first level gene ontology classification of the de *novo* assembly.

In this study, a total of 9,682 putative SSR repeat motifs were identified from 8,793 SSR containing sequences, which came from 4.4% of the total contig number in this assembly (Supplementary File, Table S3). On average, there was one SSR locus for every 16.4kbp of *de novo* assembly. After excluding mononucleotide motifs, trinucleotide repeats were the most abundant type (50.1%) (summarised in Table 3). This is consistent with Vatanparast *et al.*'s study [16], although hexamer motifs were not evaluated in this study. The most frequent dimer motifs were AG/GA/CT/TC type, followed AT/TA whereas for trimeric repeats, AAG/AGA/GAA/CTT/TCT/TTC were the most abundant (Fig. 2 & Table 4). Both observations on the most common di- and tri-nucleotide repeat motif are in agreement with the winged bean transcriptome from Vatanparast *et al.* as well as with the soybean, medicago and lotus EST-SSR summarised by Jayashree *et al.* [16,36].

Table 3. *In silico* identification of microsatellites from the mixed tissue assembly

Total number of sequences examined	198,554
Total size of examined sequences (bp)	158,382,439
Total number of identified SSRs	9,682
Number of SSR containing sequences	8,793
Number of sequences containing more than 1 SSR	780
Number of SSRs present in compound formation	352
Number of dimer-repeat	4,500
Number of trimer-repeat	4,855
Number of tetramer-repeat	279
Number of pentamer-repeat	48

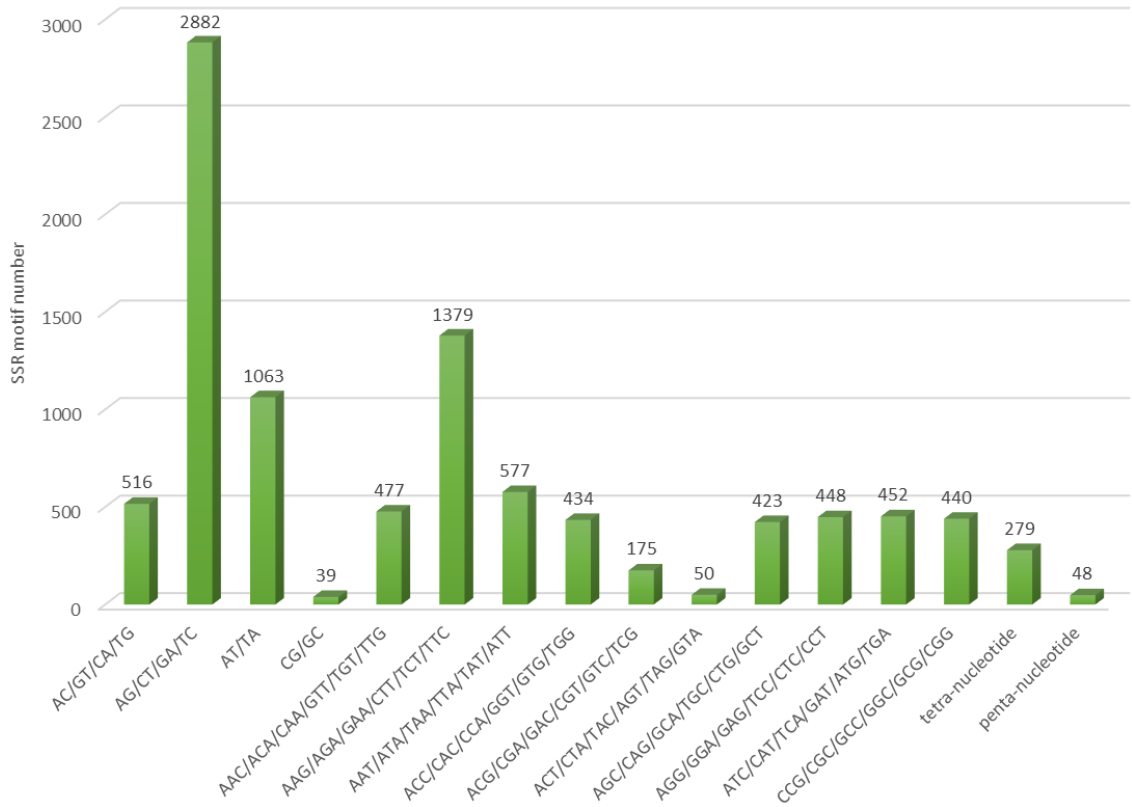


Figure 2. The number distribution of different microsatellite motif types identified.

Table 4. Frequency distribution of di- and tri-nucleotide motif repeat in this de novo assembly

	Number of Repeat Motif							Total	%
Di-nucleotide	5	6	7	8	9	10	>10		
AC/GT/CA/TG	-	256	138	63	37	12	10	516	11.5
AG/CT/GA/TC	-	995	543	330	391	434	189	2882	64.0
AT/TA	-	407	201	1167	113	107	68	1063	23.6
CG/GC	-	38	1	0	0	0	0	39	0.9
Total	-	1,696 (37.7%)	883 (19.6%)	560 (12.4%)	541 (12.0%)	553 (12.3%)	267 (5.9%)	4,500	
Tri-nucleotide									
AAC/ACA/CAA/GTT/TGT/TTG	279	131	50	17	0	0	0	477	9.8
AAG/AGA/GAA/CTT/TCT/TTC	612	405	351	11	0	0	0	1,379	28.4
AAT/ATA/TAA/TTA/TAT/ATT	305	145	112	15	0	0	0	577	11.9
ACC/CAC/CCA/GGT/GTG/TGG	307	62	55	10	0	0	0	434	8.9
ACG/CGA/GAC/CGT/GTC/TCG	90	66	12	7	0	0	0	175	3.6
ACT/CTA/TAC/AGT/TAG/GTA	36	8	3	3	0	0	0	50	1.0
AGC/CAG/GCA/TGC/CTG/GCT	271	105	38	9	0	0	0	423	8.7
AGG/GGA/GAG/TCC/CTC/CCT	247	115	75	11	0	0	0	448	9.2
ATC/CAT/TCA/GAT/ATG/TGA	311	83	24	34	0	0	0	452	9.3
CCG/CGC/GCC/GGC/GCG/CGG	247	130	55	8	0	0	0	440	9.1
Total	2,705 (55.7%)	1,250 (25.7%)	775 (16.0%)	125 (2.6%)	0	0	0	4,855	

3.2 Development of SSR markers and cluster analysis

A total of 56 (targeting 42 dimer-repeat regions) and 78 (targeting 53 trinucleotide SSR) primer pairs were designed. Subsequently, 20 dinucleotide SSR primers and 26 trinucleotide SSR primers gave good amplification products at the expected size. After polymorphism evaluation using all genotypes in this study, 18 validated SSR markers (8 for di-nucleotide and 10 for tri-nucleotide repeated motifs; Table S1) were scored and are summarised in Table 5. The low validation rate of polymorphic markers is likely to be partly due to the limited number of accessions screened, and should increase with more accessions covering a broader range of geographical origins. Residual heterozygosity could still be observed within each accession (shaded values in Table 5), even where a cycle of line purification in a controlled environment has been carried out, indicating that further cycles are needed to obtain homozygous lines, in particular for Tpt10, Tpt53, and M3. This data, along with the winged bean large flower size, also suggest that such a purification process may need to be carried out under an insect-proof enclosed environment. Using these markers, an average of 2.5 and 2.4 alleles per locus for di- and tri-nucleotide SSRs, respectively, was observed (Table 6). Individual PIC values varied from 0.16 to 0.67, which is comparable to recent legume studies in pigeopea [37], mungbean [38], and common bean [39], although lower than in cowpea [40] and bambara groundnut [41].

The cluster analysis from the SSR scores (Figure 3) showed a few clusters with the accessions originating from Papua New Guinea closely related to the Sri Lankan accession, but sharing the least similarity with the Malaysian and Indonesian materials, comparatively. To our knowledge, the genetic relationship between germplasm from Bangladesh and Malaysia are here investigated for the first time with molecular markers, and place the Bangladesh origin closer to the Sri Lankan and Papua New Guinean germplasm. Although the number of accessions used in this study is limited, they cover a reasonable range of germplasm from different origins.

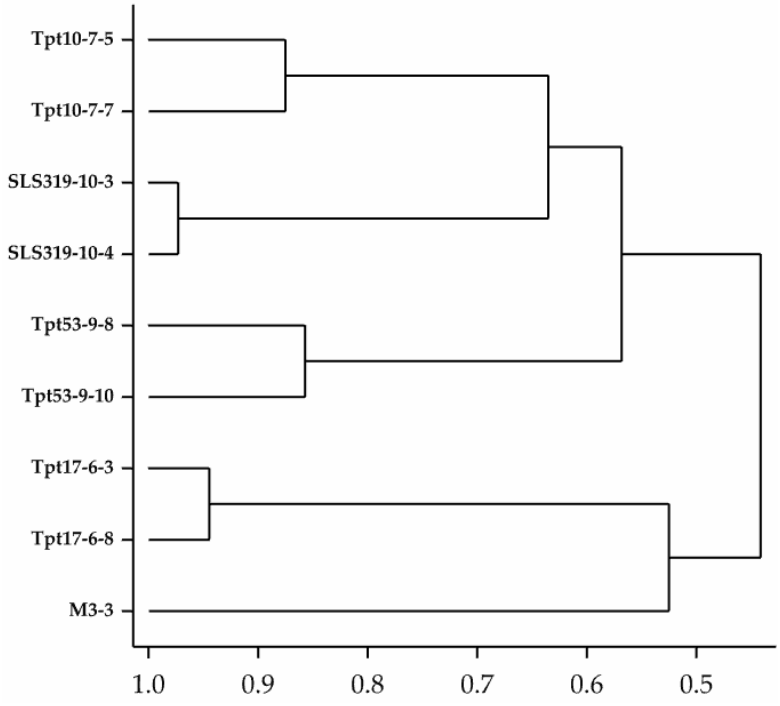


Figure 3. A dendrogram of the genetic relationship between genotypes from Papua New Guinea (Tpt10), Sri Lanka (SLS319), Bangladesh (Tpt53), Indonesia (Tpt17), and Malaysia (M3).

Table 5. Scores of 18 SSR markers from nine winged bean individuals.

Marker	Papua New Guinea		Indonesia		Bangladesh		Sri Lanka		Malaysia
	Tpt10-7-5	Tpt10-7-7	Tpt17-6-3	Tpt17-6-8	Tpt53-9-8	Tpt53-9-10	SLS319-10-3	SLS319-10-4	M3-3
P27.2	205	199/205	199	205	205	205	205	205	205
P43.2	199	199	195	195	197	199	199	199	195
Pt1.1	335	335	339	339	339	335/339	335	335	339
Pt10	226/228	228	226	226	228	228	228	228	228
Pt14	358	358	352	352	350	350	358	358	354
Pt24	219	217/219	217	217	219	219	219	219	217
Pt7.2	426/432	426	426	426	428	428	426	426	426/428
WB17	198	198	198	198	198	194/198	198	198	198
Pt53	315	309	315	315	309/315	309/315	312	312	315
Pt58	255/261	255/261	261	261	261	261	261	261	261
Pt65.1	273	273	267	267	267	267	267	267	267/273
Pt67.1	293	293	296	296	293	293	296	296	293/296
Pt68.1	226	226	229	229	226	223/226	223	223/226	226/235
Pt76.1	203	203	203	203	209	209	209	209	209
Pt78.1	306/309	306/309	306	306	309	309	306	306	309
Pt85.1	276/279	276/279	276	276	276	276	276	276	279
Pt93.1	266	266	272	272	266/272	272	266	266	276
Pt99.2	189/195	189/195	195	195	189	189	189	189	195

Table 6. A summary of data analysis of 18 SSR markers.

Marker	SSR Motif	Major Allele Frequency	No. of Alleles	Heterozygosity	PIC
P27.2	TA	0.83	2	0.11	0.24
P43.2	TA	0.56	3	0	0.49
Pt1.1	CT	0.5	2	0.11	0.38
Pt10	TC	0.72	2	0.11	0.32
Pt14	TG	0.44	4	0	0.64
Pt24	GT	0.61	2	0.11	0.36
Pt7.2	TC	0.67	3	0.22	0.4
WB17	GA	0.94	2	0.11	0.1
Average dimer SSR markers		0.66	2.5	0.1	0.37
Pt53	CGC	0.56	3	0.22	0.53
Pt58	TAG	0.89	2	0.22	0.18
Pt65.1	CAG	0.72	2	0.11	0.32
Pt67.1	AGA	0.5	2	0.11	0.38
Pt68.1	AAC	0.5	4	0.33	0.59
Pt76.1	CGC	0.56	2	0	0.37
Pt78.1	AAC	0.56	2	0.22	0.37
Pt85.1	GCG	0.78	2	0.22	0.29
Pt93.1	TGT	0.5	3	0.11	0.5
Pt99.2	TTC	0.56	2	0.22	0.37
Average trimer SSR marker		0.61	2.4	0.18	0.39

4. Conclusions

A set of validated functional winged bean genic-SSR markers is reported here for the first time, to our best knowledge. The reported residual heterozygosity across screened genotypes has suggested further investigation needs to be carried out on the rate of natural outcrossing in winged bean, in order to understand how genetic materials should be maintained, improved, and introduced into breeding programmes. The cluster analysis provides an initial insight into the potential for these markers to be used on a larger number of winged bean accessions, to carry out a more comprehensive diversity analysis with the evaluation of germplasm from genebanks and from commonly cultivated lines. Finally, this set of 18 microsatellite markers could also be used to contribute to genetic linkage maps in winged bean, with the integration of SNPs markers for higher density. Such a map would be the first backbone for linkage analysis and the genetic dissection of traits with agronomic importance in this legume.

Acknowledgments: This work was financially supported by Kouk Foundation Berhad, Crops For the Future (CFF) and the University of Nottingham Malaysia Campus (UNMC) through the CFF-UNMC Doctoral Training Partnership scheme. Authors would like to thank Victoria Wright from DeepSeq in coordinating. Seeds were contributed by the International Institute of Tropical Agriculture (IITA) and Malaysian Agricultural Research and Development Institute (MARDI). This paper is dedicated to the memory of Quin Nee Wong, who passed away in a tragic car accident on October 12, 2014.

Author Contributions: Q.N.W. and A.S.T. carried out planting and sampling work and, together with W.K.H., SSR marker development and validation. A.S.T. and W.K.H wrote the manuscript. S.M. performed library preparation for RNA-seq and M.B. performed bioinformatics analysis. A.K. contributed Sri Lanka genetic materials. F. M. helped to draft the manuscript and contributed to experimental design. S.Ms. conceived of the study, and participated in its design and coordination and revised the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. NAS. *The Winged Bean: High-Protein Crop for the Humid Tropics*, Second Ed.; National Academy Press: Washington, DC, USA, **1981.**
2. Klu, G.Y.P. Induced mutations for accelerated domestication-a case study of winged bean. *West African J. Appl. Ecol*. **2000,** *1*, 47–52.
3. Khan, T.N. Papua New Guinea: A centre of genetic diversity in winged bean (*Psophocarpus tetragonologus* (L.) Dc.). *Euphytica* **1976** *25*, 693–705.
4. Harder, D.K. Chromosome Counts in Psophocarpus. *Kew Bull.* **1992,** *47*, 529.
5. Harder, D. K.; & Smartt, J. Further evidence on the origin of the cultivated winged bean, *Psophocarpus tetragonolobus* (L.) DC. *Econ. Bot.* **1992,** *46*, 187–191.
6. Prakash, D.; Misra, P.N.; Misra, P.S. Amino acid profile of winged bean (*Psophocarpus tetragonolobus* (L.) DC): a rich source of vegetable protein. *Plant Foods Hum. Nutr.* **1987,** *37*, 261–4.
7. Kadam, S.S.; Salunkhe, D.K. Winged bean in human nutrition. *Crit. Rev. Food Sci. Nutr.* **1984** *21*, 1–40.
8. Okezie, B.O.; Martin, F.W. Chemical composition of dry seeds and fresh leaves of winged bean varieties grown in the U.S. and Puerto Rico. *J. Food Sci.* **1980,** *45*, 1045–1051.
9. Anugroho, F.; Kitou, M.; Kinjo, K.; Kobashigawa, N. Growth and Nutrient Accumulation of Winged Bean and Velvet Bean as Cover Crops in a Subtropical Region. *Plant Prod. Sci.* **2010** *13*, 360–366.
10. Banerjee, A.; Bagchi, D.K.; Si, L.K. Studies on the potential of winged bean as a multipurpose legume cover crop in tropical regions. *Exp. Agric.* **2008**, *20*, 297.
11. Hikam, S.; MacKown, C.T.; Poneleit, C.G.; Hildebrand, D.F. Growth and N accumulation in maize and winged bean as affected by N level and intercropping. *Ann. Bot.* **1991**, *68*, 17–22.
12. FAO Coping with climate change – the roles of genetic resources for food and agriculture; Rome, Italy **2015.**
13. Mohanty, C.S.; Verma, S.; Singh, V.; Khan, S.; Gaur, P.; Gupta, P.; Nizar, M.A.; Dikshit, N.; Pattanayak, R.; Shukla, A.; *et al*. Characterization of winged bean (*Psophocarpus tetragonolobus* (L .) DC .) based on molecular , chemical and physiological parameters. *Am. J. Mol. Biol.* **2013**, *3*, 187–197.

14. Koshy, E.P.P.; Alex, B.K.K.; John, P. Clonal fidelity studies on regenerants of *Psophocarpus tetragonolobus* (L.) DC. using RAPD markers. *Bioscan* **2013,** *8*, 763–766.
15. Chen, D.; Yi, X.; Yang, H.; Zhou, H.; Yu, Y.; Tian, Y.; Lu, X. Genetic diversity evaluation of winged bean (Psophocarpus tetragonolobus (L.) DC.) using inter-simple sequence repeat (ISSR). *Genetic Resources and Crop Evolution* **2015**, *62*(6), 823–828.
16. Vatanparast, M.; Shetty, P.; Chopra, R.; Doyle, J.J.; Sathyanarayana, N., Egan, A.N. Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). *Sci. Rep.* **2016**, *6*, 29070.
17. Chapman, M. A. Transcriptome sequencing and marker development for four underutilized legumes. *Appl. Plant Sci.* **2015**, *3*, 1400111.
18. Erskine, W. Measurements of the cross-pollination of winged bean in Papua New Guinea. *SABRAO Journal of Breeding and Genetics* **1980**, 12 (1), 11-14.
19. Eagleton, G. E. Evaluation of genetic resources in the winged bean (*Psophocarpus tetragonolobus* (L.) DC.) and their utilisation in the development of cultivars for higher latitudes. Ph.D. Dissertation, University of Western Australia, Perth Australia, **1983**.
20. De Silva, H. N.; Omran, A. Diallel analysis of yield and yield components of winged bean (*Psophocarpus tetragonolobus* (L.) D.C.). *J. Agric. Sci.* **1986**, *106*(3), 485.
21. Erskine, W.; Khan, T. N. Inheritance of pigmentation and pod shape in winged bean. *Euphytica* **1977**, *26*(3), 829–831.
22. Erskine, B. Y. W. (1981). Heritability and combining ability of vegetative and phenological characters of winged beans (*Psophocarpus tetragonolobus* (L.) D. C.). *J. Agric. Sci.* **1981**, *96*, 503–508.
23. Joshi, N. A.; Fass, J. N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.
24. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; *et al.* Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* **2011,** 29(7), 644-652.
25. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*(1), 421.
26. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **2014**; 42 (D1), D222-D230.
27. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, 8(10), 785–786.
28. Powell, S.; Forslund, K.; Szklarczyk, D.; Trachana, K.; Roth, A.; Huerta-Cepas, J.; Gabaldón, T.; Rattei, T.; Creevey, C.; Kuhn, M.; *et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* **2014**, 42(D1), D231-D239.
29. Gene Ontology Consortium. "The Gene Ontology (GO) database and informatics resource." *Nucleic acids res.* **2004**, 32(Suppl 1), D258-D261.
30. Thiel, T.; Michalek, W.; Varshney, R. K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. **2003**, 106(3), 411-422.
31. Koressaar, T.; Remm, M. Enhancements and modifications of primer design program Primer3. Bioinformatics 2007, 23, 1289–1291.
32. Doyle, J. DNA Protocols for Plants. In *Molecular Techniques in Taxonomy*; Hewitt, G. M.; Johnston, A. W. B.; Young, J. P. W., Eds.; Springer Berlin Heidelberg: Berlin Heidelberg, Germany, **1991**, pp. 283-293.
33. Schuelke, M. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **2000**, *18*, 233–234.
34. Liu, K.; Muse, S. V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*(9), 2128–2129.
35. VSN International. GenStat *for Windows* 18th Edition. VSN International, Hemel Hempstead, UK, **2015**.
36. Jayashree, B.; Punna, R; Prasad, P; Bantte, K; Hash, C. T.; Chandra, S.; Hoisington, D. A.; Varshney, R. K. A database of simple sequence repeats from cereal and legume expressed sequence tags mined *in silico*: survey and evaluation. *In Silico Biology* **2006**, 6, 0054.

37. Dutta, S.; Kumawat, G.; Singh, B.P.; Gupta, D.K.; Singh, S.; Dogra, V.; Gaikwad, K.; Sharma, T.R.; Raje, R.S.; Bandhopadhya, T.K.; *et al.* Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* **2011,** *11*, 17.
38. Gupta, S.K.; Bansal, R.; Gopalakrishna, T. Development and characterization of genic SSR markers for mungbean (*Vigna radiata* (L.) Wilczek). *Euphytica* **2014**,*195*, 245–258.
39. Blair, M.W.; Hurtado, N.; Chavarro, C.M.; Muñoz-Torres, M.C.; Giraldo, M.C.; Pedraza, F.; Tomkins, J.; Wing, R.; Varshney, R.; Graner, A.; *et al.* Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMc series. *BMC Plant Biol.* **2011**, *11*, 50.
40. Gupta, S.K.; Gopalakrishna, T. Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other Vigna species. *Genome* **2010**, *53*, 508–523.
41. Molosiwa, O.O.; Aliyu, S.; Stadler, F.; Mayes, K.; Massawe, F.; Kilian, A.; Mayes, S. SSR marker development, genetic diversity and population structure analysis of Bambara groundnut [*Vigna subterranean* (L.) Verdc.] landraces. *Genet. Resour. Crop Evol.* **2015**, *62(8)*, 1225-1243.

Supplementary Files

Table S1. The sequences of forward and reverse SSR-primers used in this study.

SSR Motif	Marker	Primer sequence 5′ -> 3′
TA	P27.2F	CCTCCCAGATCAACCACTACTTTC
	P27.2R	GGATAGGCATGGGTTCCTTTACTC
TA	P43.2F	GGAGTGTAACTAAGTATCGTGTATGG
	P43.2R	AGATGCTCTTGGTTTACATCATTTG
CT	Pt1.1F	GCACCTAACAACCAATCTAAC
	Pt1.1R	GATACAATCACCGTGTTGCTTCAA
TC	Pt10F	CCATTCACCACCGTATTCACTATTC
	Pt10R	CACAGAATCAGAAATCACGACAGG
TG	Pt14F	TAAACTCGGTAAGTGTGTTCTTGAG
	Pt14R	GAGACGCAAAGAGTAAGCAATAAG
GT	Pt24F	GAAAGAAAGAGGTTTGTTTGGAAGG
	Pt24R	CCTTCTATTTCTCTCAACCAGTCTTC
TC	Pt7.2F	CCCTCCTTCTTGTTTAGTAGTGTAA
	Pt7.2R	CCTGCAGTTCTGTTCTGTTTG
GA	WB17F	CATAATGGGTTTCATGGCAATGTG
	WB17R	GCAAAGTTTGTGAATAGCTCCATAAC
CGC	Pt53F	TTCAATTCCGACTTCTTAATCCATTCCCCGAA
	Pt53R	CTTCGTAGGTTTGATTTGAAGACGCCGAAG
TAG	Pt58F	TCTGAGGAGCTGAAAAGTTTGGCGCTTC
	Pt58R	TGAACTTGCATCATGGGTTGGATCAGTC
CAG	Pt65.1F	CCTTTCCTGCAACACAAACCAAAC
	Pt65.1R	GAGGTCCTTAATTTCCTCGAACAAGTG
AGA	Pt67.1F	CAGCCCGTGAGTAAGTTGCATTAG
	Pt67.1R	GCTCTTCTTCAACCTCTTCTTCTTCTTC
AAC	Pt68.1F	CTCCGCGAGAACCTTAATCAATCG
	Pt68.1R	GCACAGCATAGCTTCTACTTCTGG
CGC	Pt76.1F	CACTCCAATTTCAACCATGCCCTTT
	Pt76.1R	GAAGAACACGACCACCAGGATGAC
AAC	Pt78.1F	TAGCAATGAAGAAGTGGAGTACGGAATAG
	Pt78.1R	GCAATTTGGGAATGTGGTGGGTAA
GCG	Pt85.1F	CCTGAATGAGAAGCTGAAACGAAACG
	Pt85.1R	GAGGTGCCATGCGATACATATACTCC
TGT	Pt93.1F	CACCACACCATTCTCTTCCTCTTC
	Pt93.1R	CCCATTTCACTGATTCAAATCATCACC
TTC	Pt99.2F	TCATCTTTCCATGGCTTCCACTCC
	Pt99.2R	CAAGTGATTATGATGAGGGCTTTGTGC

Chapter 3 - Generation and phenotyping of an F_2 population

Tanzi AS, Ho WK, Massawe F, Mayes S (2019) Development and interaction between plant architecture and yield-related traits in winged bean (*Psophocarpus tetragonolobus* (L.) DC.). Euphytica 215:36. doi: 10.1007/s10681-019-2359-8

This chapter describes the generation and the field assessment of an F_2 population derived from one of the controlled crosses carried out between April and October 2016. The hybrids from such crosses, along with their parents, were phenotyped in field conditions between October 2016 and March 2017: based on this data (reported in appendix) only one cross was selected to be advanced towards F_2 field assessment for phenotypic analyses (presented in this chapter). The selection was based on morphological and developmental traits recorded across parental and hybrid genotypes in field conditions, with particular focus on selecting for individuals differing in branch number per plant, stem length, flowering time, and number of leaves per plant.

Supplementary information is reported in the appendix at the end of the chapter (in particular, correlations between all measured traits are reported in paragraph a.5)

Development and interaction between plant architecture and yield-related traits in winged bean (*Psophocarpus tetragonolobus* (L.) DC.)

Alberto Stefano Tanzi[1,2], Wai Kuan Ho[1,2], Festo Massawe[1,2], Sean Mayes[2,3]

[1]Biotechnology Research Centre, School of Biosciences, Faculty of Science, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[2]Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[3]School of Biosciences, Faculty of Science, University of Nottingham Sutton Bonington Campus, Sutton Bonington, Leicestershire LE12 5RD, UK

*Corresponding author: sean.mayes@nottingham.ac.uk

Keywords Winged bean; *Psophocarpus tetragonolobus*; plant breeding; plant development; plant architecture; yield

Abstract

Winged bean (*P. tetragonolobus* (L.) DC.) is a leguminous crop that could contribute towards food security in tropical areas, but whose growth and development is still poorly understood. In order to develop improved individuals for increased green pod and seed productivity, we investigated the factors involved in winged bean plant architecture, development, and their link to a number of yield-related traits. An F_2 population was generated from the cross between M3 and FP15 Malaysian accessions and assessed under field conditions in Malaysia. The results showed stem length to be mainly influenced by internode length (r_s= .80; p< .01), while multiple genes could be controlling the number of branches, with an average number of branches in the offspring above the highest parent value. The average length of branches appeared to influence the most the final number of pods per plant (r_s= .44; p< .001), while flowering showed potentially transgressive segregation towards earliness, without preventing the potential development of high pod-yielding individuals (r_s= -208; p= .056). Taken together, the results reported here shed light on the interaction between morphological, developmental, and yield-related

traits, defining potential targets for developing crop ideotypes to direct breeding programmes for this underutilised crop.

1 Introduction

The second of the Sustainable Development Goals by the United Nations is to end hunger, achieve food security, improved nutrition, and promote sustainable agriculture (UN 2015). To achieve this goal, food production systems have to accommodate the demand for sufficient, safe, and nutritious food by increasing production, diversity, and access. Despite the number of edible crops that could contribute, only a handful (wheat, *Triticum aestivum*; maize, *Zea mays*; rice, *Oryza sativa*; soybean, *Glycine max*) still represent the majority of food production, with a recorded global trend towards more homogeneous agricultural systems (Godfray et al. 2010; Khoury et al. 2014). Such a narrow number of species could undermine food security whenever yield expectations are not matched, due to ever-changing conditions and extreme weather events (Challinor et al. 2014; Ray et al. 2015). Furthermore, an increasing volume of research shows the positive effects of a diversified agriculture on yield stability, sustainability, socio-economics, and human health (Thrupp 2000; Jackson et al. 2007; Hajjar et al. 2008; Chappell and LaValle 2011; Padulosi et al. 2013; Powell et al. 2015; Dwivedi et al. 2016, 2017; Lachat et al. 2017), with a case made for increased vegetable and fruit production (Schreinemachers et al. 2018). If we want to reverse this global trend, we need to encourage and enable the cultivation of a wider range of crops in each geographical area. In this regard, underutilised species offer the advantage of being locally adapted, often to low input systems, with desirable traits to contribute towards a more resilient and diversified agriculture (Weil and Khalil 1986; Padulosi et al. 2013; Pellegrini and Tasciotti 2014; Chivenge et al. 2015; Adhikari et al. 2017).

In the context of Asia and the Pacific area, a leguminous species named "Winged Bean" (*Psophocarpus tetragonolobus* (L.) DC.) has been cultivated as a vegetable for its green pods, leaves, and tuberous roots, and also as a pulse (NAS 1975; Khan 1976; Lepcha et al. 2017). Like a number of other leguminous species, winged bean has intertwining vines, with a complex plant architecture. Shaped by a main shoot apical meristem and lateral branches, these latter grow according to the number and length of each phytomer (the repetitive unit formed by internode and node), and the fate of each axillary

meristem it harbours. In terms of plant architecture, little research has been carried out on winged bean, although there have been a series of experiments focused on the effects of combined temperature and day length regimes on flowering and vegetative biomass accumulation (Herath and Ormrod 1979; Anonymous 1982; Wong 1983; Schiavinato and Válio 1996a). A few investigations aimed at identify influences on plant structure have been published, by testing different physical supports and eventually concluding that trellis was needed in order to promote lateral branching and pod production (Anonymous 1981; Schiavinato and Válio 1996b). However, no investigation has so far described winged bean plant architecture through its component morphological traits (e.g. lateral branching, stem length, node number and internode length), neither has there been research on how these are controlled genetically and inherited. This is despite research in crops like common bean (*Phaseolus vulgaris*), soybean (*G. max*), faba bean (*Vicia faba*), and pea (*Pisum sativum*) addressing the underlying mechanisms involved in plant architecture changes during domestication, and their impact on productivity and field practices (Lester et al. 1997; Teixeira et al. 1999; Weeden 2007; Ando et al. 2007; Wang and Li 2008; Braun et al. 2012; Andrivon et al. 2013; Ávila et al. 2017; Sun et al. 2018).

Here, we present a study of winged bean plant architecture, carried out on an F_2 population obtained from morphologically contrasting parents: M3 and FP15. The aim is to provide a better understanding of the relationships between key morphological traits, their underlying genetic mechanisms, and to determine how variation in plant architecture could impact on pod and seed yield. This could identify the basis for future improvement programmes focused on obtaining high yielding and more amenable winged bean ideotypes, to promote the cultivation of this crop and support efforts to increase agrobiodiversity in tropical areas through adoption and development of endemic species.

2 Materials and methods

2.1 Plant material and controlled crosses

Winged bean is considered to have a cleistogamous floral system, which would imply autogamy, with self-pollination taking place before the large flowers open in the morning hours (Karikari 1972; Erskine and Bala 1976; Erskine 1980). Parental individuals were grown in controlled environment rooms (12h photoperiod, 27°/22° day/night temperature, 65%RH minimum) at Crops For the Future (CFF,

Semenyih, Malaysia), and crossed following the protocol developed by Erskine and Bala (Erskine and Bala 1976). Half-diallel crosses were performed between the male parent M3 (MARDI, Malaysia), and the emasculated flowers of FP15 (Malaysian, local accession). The male parent did not undergo any additional purification, while the female individual was derived from 2 rounds of purification by single seed descent (SSD). Using previously developed genic-SSR markers (Wong et al. 2017), single parents used during crossing were genotyped following the same protocol: M3 reported 17% of heterozygous loci, while FP15 showed no heterozygosity. From crossing to F_2 field assessment, parental individuals went through another 2 rounds of self-pollination and SSD purification. This should have further reduced any residual heterozygosity before phenotyping along with the F_2 population. The same SSR markers were used to validate F_1 hybrids obtained from the performed crosses.

2.2 Growing conditions

All plants assessed in field conditions were grown on ridges at the Crops For the Future - Field Research Centre (CFF-FRC, Semenyih, Malaysia) on trellis structures, each made by a net (2m tall; 1m width), with a 1.5m distance between consecutive structures, and 1m between each ridge. This ensured no competition between adjacent individuals. Seeds were scarified with sand paper in order to obtain more homogenous germination in bags. Plants were then transferred into the field 2 weeks after emergence (a.g.), when stems had reached about 15-20cm in height. Fertilizer (NPK 15:15:15) was applied 4 times during the growing season (5g pellet each time, per individual), while pesticide was sprayed approximately every 10 days to prevent pest damage of the main stems (Karate – Syngenta, Switzerland) at 0.5mL/L concentration.

For the first generation: eight F_1 hybrids, along with 5 replicates of each parental genotype, were assessed and left free to self-pollinate in a complete randomized design (CRD) from October 2016 to March 2017, in a sandy loam soil (pH 5.3). A single hybrid plant was randomly selected for collection of F_2 seeds. The F_2 segregating population (XB2) derived from this was then assessed between June 2017 and November 2017 in a complete randomized block design (CRBD) with three blocks, each with five replicates of each parent and thirty F_2 individuals randomly allocated across four ridges; additional

F_2 individuals were used as border plants. The soil had a sandy loam profile with pH 5.0, while day/night temperatures were 32±0.9°C and 23±0.3°C (on site weather station; DeltaT).

2.3 Traits recorded

2.3.1 Morphological traits (recorded weekly until 56 days after emergence)

Stem length (StL) was recorded for the main shoot apical meristem, removing individuals that were damaged and had lost apical dominance before the end of morphological measurements; *Weekly growth* (WGR) was calculated as the increase in each individual stem length since the previous measurement.

Number of nodes (NoN) were counted by starting from the soil to the first true leaves;

Internode length (InL) was recorded for 6^{th}, 10^{th}, 15^{th}, and 20^{th} internode;

Number of leaves per plant (LPP) was recorded excluding cotyledon leaves (unifoliate);

Leaf width and *leaf length* (LeW, LeL) were recorded at the largest leaf lamina section, and along the main vein to the tip respectively; measurements were taken on central leaflet on 3 fully expanded leaves per plant, at approximately 1 meter from the ground. Both parental, as well as progeny individuals had deltoid leaflet shape.

Branch number per plant (NoB) was recorded for branches at least 10cm in length and carrying at least 1 fully expanded leaf, for the first 10 nodes of the main stem;

Length of branch (LoB) was the average length of all branches longer than 10cm per plant, at week 8 after emergence; *total sum of branch length* (SLB) was calculated by addition of the single branches lengths.

2.3.2 Development traits

Days to first open flower (DtF) was recorded from the day of emergence (considered when stem length was 2cm) to the first open flower;

Pod maturing time (PMT) was recorded from first open flower to first harvesting of that mature pod ;

2.3.3 Yield-related traits

Harvested *mature pods per plant* (PPP);

Dry pod length (PoL), and fully-developed *seeds per pods* (SPP) were recorded using the first 15 harvested pods. Seeds weight was also recorded from each pod, to calculate *hundred-seeds weight* (HSW). For such traits, harvested mature pods were dried in a high-volume oven (Memmert, Germany) at 35°C for 4 days before measurements.

2.4 Statistical analysis

Descriptive statistics and data analysis were performed using IBM SPSS® Statistics v25. A Shapiro-Wilk test was performed to check for normal distribution of trait data (significance level at = .05), for each parental genotype and progeny population, along with Levene's test for homogeneity of variance. Spearman's bivariate correlation analysis was carried out rather than Pearson's, whenever data did not have a normal frequency distribution. For each Mann-Whitney U test we reported median (*Mdn*), test statistic (*U*), significance (*p*) and effect size (r= z-score/$\sqrt{}$ (number of observations)).

Two Principal Component Analysis have been performed with IBM SPSS® Statistics v25, and with Genstat v18.

3 Results

3.1 Plant traits

3.1.1 Stem growth

Stem length (StL), *number of nodes* (NoN), and *internode length* (InL), followed a normal frequency distribution in both parental genotypes (p> .05). At week 8 (56 days after emergence), the male M3 parent showed a significantly shorter stem compared to the female FP15, the latter appearing to grow faster and with longer internodes (Table 3.1). Such differences between parental phenotypes were also recorded since earlier stages (see Fig.3.1 and also supplementary material for more pictures of parental and F_2 individuals).

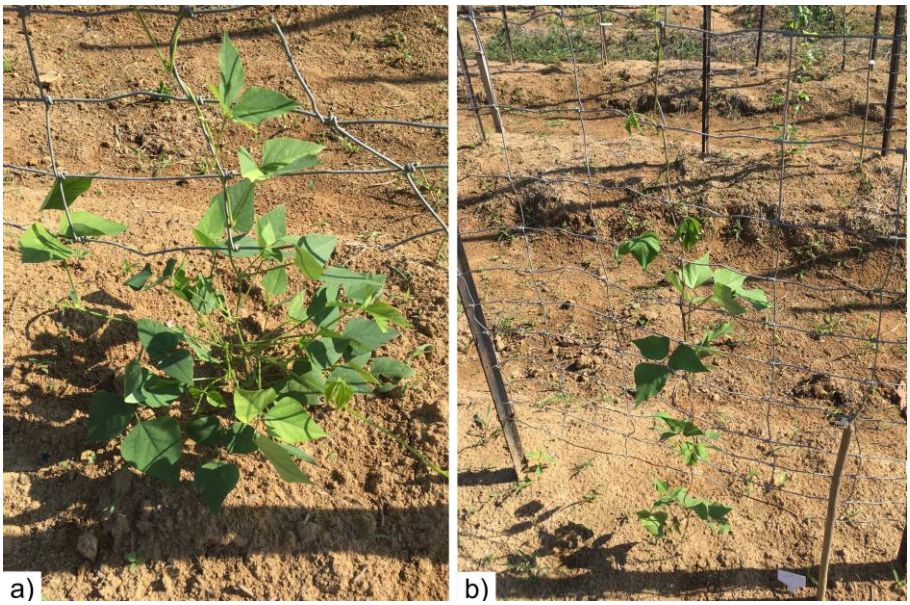


Fig. 3.1: Parental individuals at about 36 days after emergence. a) M3, with a *stem length* of 95cm, 3 lateral branches and 27 leaves; b) FP15, *stem length* of 125cm, no branches, and 15 leaves.

Table 3.1: Average of traits values related to plant height and growth, at 56 days after emergence, in parents (M3 and FP15) and segregating population (XB2). Letters (a, b) for parental values show statistical differences ($p < .05$) (see results section).

	Stem Length (StL) (cm)	**Number of Nodes (NoN)**	**Internode Length (InL) (cm)**	**Weekly Growth (WGR) (cm/week)**
M3	200.4±64.3 [a]	28.9±4.6 [a]	7.2±1.8 [a]	27.8±9.1 [a]
FP15	322.3±13.1 [b]	29.2±1.8 [a]	12.9±1.0 [b]	44.3±1.8 [b]
XB2(F_2)	270.2±67.5	30.3±3.3	9.9±2.65	37.4±9.7

StL and NoN in the F_2 showed a significant deviation from the normal ($p < .001$), while *internode length* (InL) had a normal frequency distribution (Fig. 3.2a). Spearman's correlation found *stem length* to be significantly ($p < .01$) correlated with both InL ($r_s = .80$) and NoN ($r_s = .58$), and the same results were obtained for *weekly growth* (WGR). InL appears to be the main factor that influences StL, explaining 60.1% of variation for the trait from the regression analysis (Fig. 3.2b). InL also showed the highest coefficient of variation (27%), followed by StL (25%), while NoN showed more limited variation between replicates (11%).

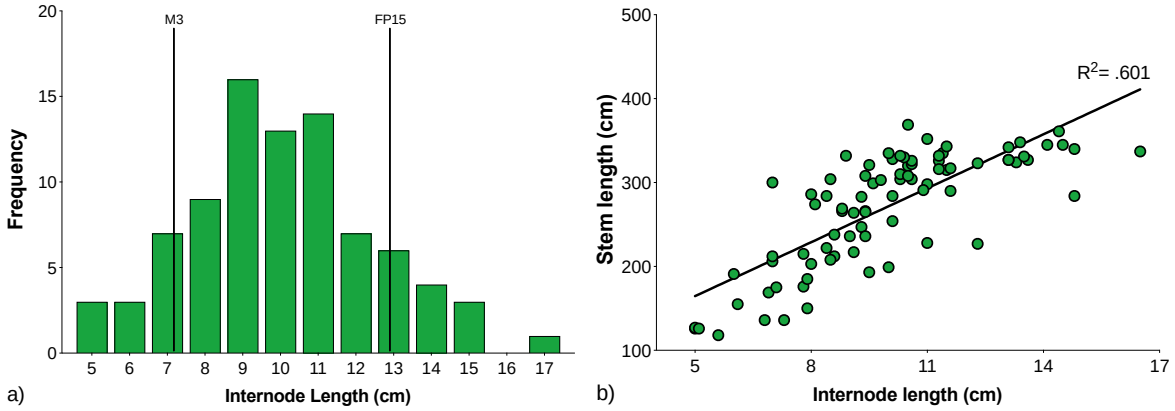


Fig. 3.2: a) Internode length frequency distribution across the F_2; vertical lines represent mean values for parental genotypes (M3 and FP15). b) Stem and internode length regression in F_2 (data at 56 days after emergence).

3.1.2 Secondary branches

The *branch number per plant* (NoB) showed a non-normal frequency distribution in both parents (p< .05), with the female FP15 (*Mdn*= .00) having significantly lower NoB than the male parent M3 (*Mdn*= 2.00) (U= 8.50; p= .001; r= -.80) (Table 3.2).

Also across the F_2, NoB showed a non-normal distribution (p< .05) (Fig. 3.3). The average number of branches per plant in the F_2 was above the highest parent (M3) (Table 3.2), possibly due also to individuals carrying 6 to 8 branches (above the maximum of 5 recorded in M3 individuals). NoB coefficient of variation was 62%, and the trait had a small but significant positive correlation with *number of nodes* (r_s= .29; p< .01).

Table 3.2: Lateral branches and their average lengths in the parental and F_2 population. Different letters (a, b) for parental values indicate statistically differences (see results section).

	Branches (NoB)	Length of Branch (LoB) (cm)	Sum of Branches (SLB) (cm)	Leaves per Plant (LPP)	Leaf Width (LeW) (cm)	Leaf Length (LeL) (cm)
M3	2.3±1.2 [a]	134.0±53.3	299.4±177.1	45.8±12.9 [a]	7.4±1.0 [a]	8.8±1.1 [a]
FP15	0.3±0.6 [b]	12.8±30.0 [*]	18.9±46.3 [*]	23.9±3.5 [b]	10.0±1.0 [b]	13.0±1.2 [b]
XB2(F_2)	2.6±1.6	88.2±51.2	248.7±166.9	44.6±13.2	8.2±1.2	9.6±1.5

*: only 2 individuals developed lateral branches.

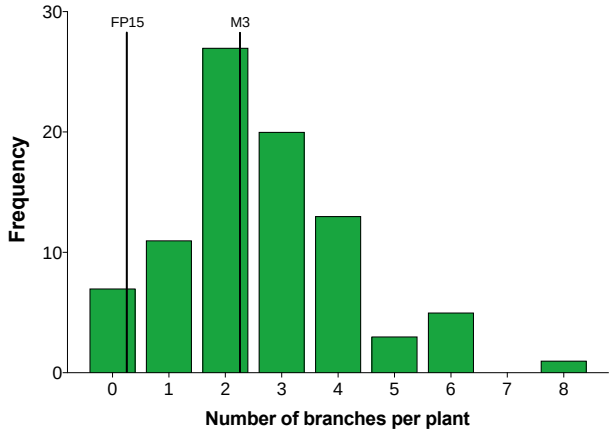


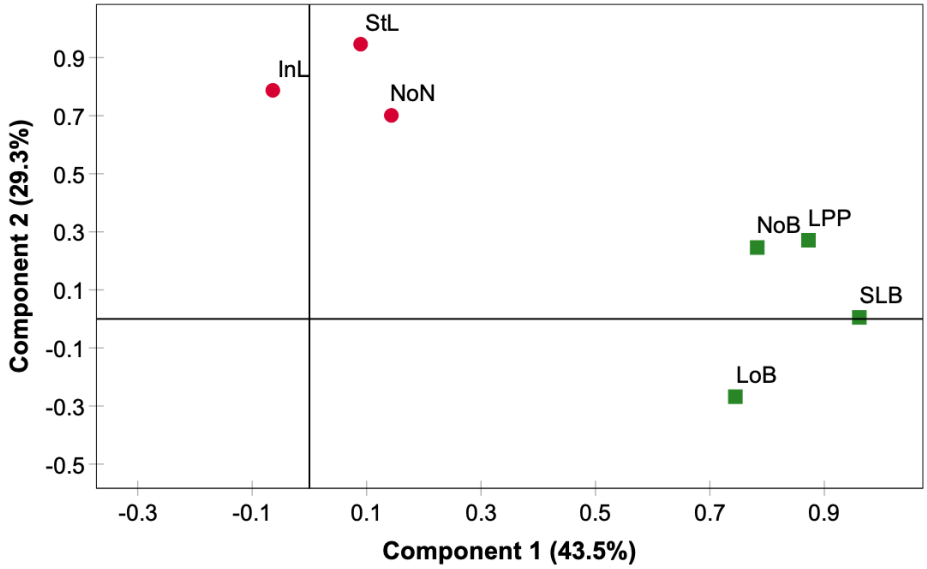
Fig. 3.3: Number of branches for the first 10 nodes (NoB) for the F_2 population at 56 days after emergence (2.6±1.6, N=84; Skewness: 0.657±0. 263). Vertical lines represent mean values for parental genotypes (FP15 and M3).

Between parents, and across the segregating population, there appeared to be differences in the average *length of branch* (LoB) and the *total sum of branch lengths* (SLB), the first showing a normal distribution (p> .05). From an earlier stage, M3 individuals displayed long branches between the first three nodes that overtook, in 43% of cases, the length to the apical meristem. Long branches were also observed in the F_2 population as well (see Table 3.2), although only 11% of the individuals generated branches which were longer than the initial SAM. In agreement with this observations, NoB and LoB were found to contribute to or correlate with the higher numbers of *leaves per plant* (LPP) (r_s= .794 and r= .386) (see also Fig. 3.8, in section 3.2 below), compared to *number of nodes* alone (r_s= .34 respectively; p< .01). Variation in *branch number per plant* did not translate into significant effects on LoB (r_s= .20; p= .058), while both traits contributed to the *total sum of branch lengths* for individuals (p< .01), with the first having a slightly larger effect (r_s= .77) than the second (r_s= .72). LoB and SLB reported respectively 58 and 68% coefficients of variation.

Parents showed differences in both LPP and leaf shape-related traits (*leaf width*, LeW; *leaf length* LeL): M3 (*Mdn*= 45.5) had significantly more leaves than FP15 (*Mdn* = 22.0) (U= 4.00; p< .001; r= -.81), but M3 also had narrower (*Mdn*= 7.6) and shorter leaves (*Mdn*= 9) than FP15 (*Mdn*= 10 and 13.3 for LeW and LeL respectively) (U= .50; p< .001; r= -.84 for LeW; U= 4.50; p< .001; r= -.78 for LeL respectively). Across the F_2 population, LPP was found to have a significant (p< .001) inverse correlation with both leaf width (r= -.36) and length (r= -.41).

3.1.3 Principal Component Analysis (PCA) for morphological traits

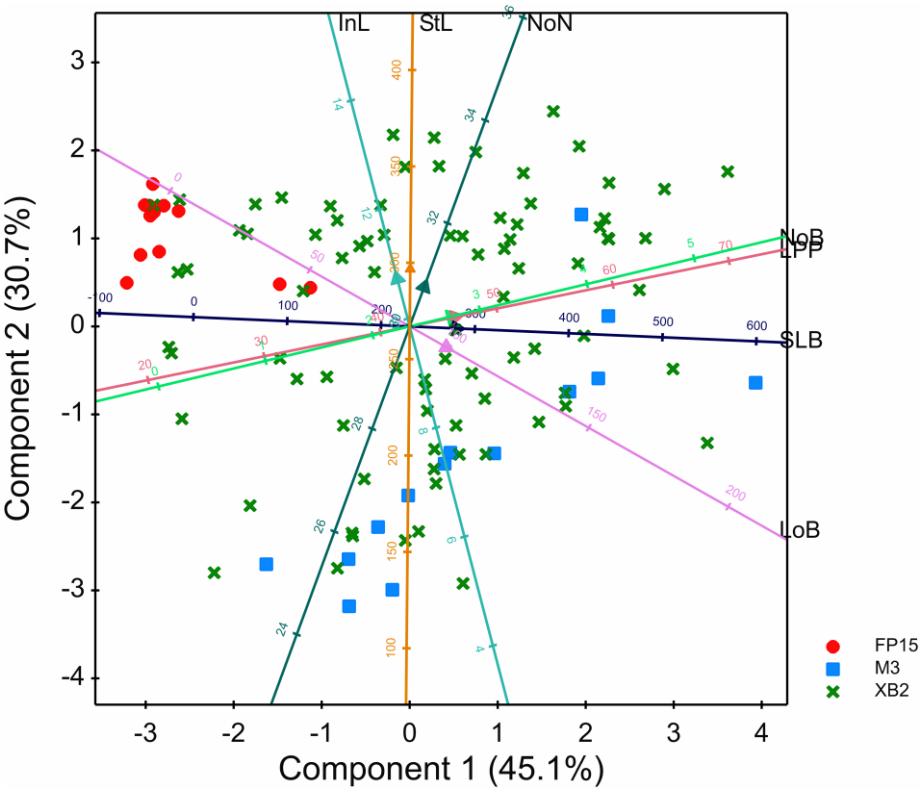
A principal component analysis was carried out in order to investigate whether there were factors that could group correlated morphological traits together. Two components explaining a cumulative variance of 72.8% were identified: one was associated with traits that define the main vertical growth (*stem length*, *number of nodes* and *internode length*), while the second was related with what determines how "bushy" individuals are (*number of branches*, *length of branch*, and *leaves per plant*). PCA and loading scores from factor rotation analysis using F_2 phenotypic data are reported in Fig. 3.4.



	C.1	C.2
StL		.946
InL		.787
NoN		.701
NoB	.783	
SLB	.961	
LoB	.744	
LPP	.872	

Fig. 3.4: PCA graph and loading scores for each component (C.1 and C.2) from the rotated component matrix (orthogonal varimax rotation), performed with IBM SPSS using F_2 population data. StL: *stem length*; InL: *internode length*; NoN: *number of nodes* NoB: *branch per plant*; SLB: *total sum of branch length*; LoB: *length of branch*; LPP: *number of leaves per plant*.

An additional PCA analysis mapped all the parental and F_2 individuals, again across two components (see Fig. 3.5). These latter explained a cumulative variance of 75.8%, with included phenotypic traits showing similar associations to each component as in the first analysis (see loading scores in Fig. 3.4).



	C.1	C.2
StL		.652
InL		.545
NoN		.445
NoB	.476	
SLB	.534	
LoB	.411	
LPP	.521	

Fig. 3.5: PCA graph and loading scores for each component (C.1 and C.2) from orthogonal varimax rotation, performed with Genstat v18 using parental (M3 and FP15) and F_2 (XB2) individuals' phenotypic data. StL: *stem length*; InL: *internode length*; NoN: *number of nodes* NoB: *branch per plant*; SLB: *total sum of branch length*; LoB: *length of branch*; LPP: *number of leaves per plant*.

3.1.4 Days from emergence to flowering, and pod development time

Days to first open flower (DtF) showed significantly earlier open flowers in the female FP15 parent (*Mdn*= 56.5) compared to the male parent (*Mdn*= 65.0) (*U*= .500; *p*< .01; *r*= -.847). The F_2 population mostly fell between the two parental averages (*Mdn*= 60.0), although there was possible transgressive segregation, with individuals flowering earlier (50 days) as well as later (78 days) than any parental individual (Fig. 3.6a). Trait distribution deviated from normal (*p*< .01), with high skewness towards late flowering (1.38, *SE*= .26). Flowering appears to be significantly (*p*< .01) negatively correlated with *stem length* (r_s= -.45), and *branch number per plant* (r_s= -.25).

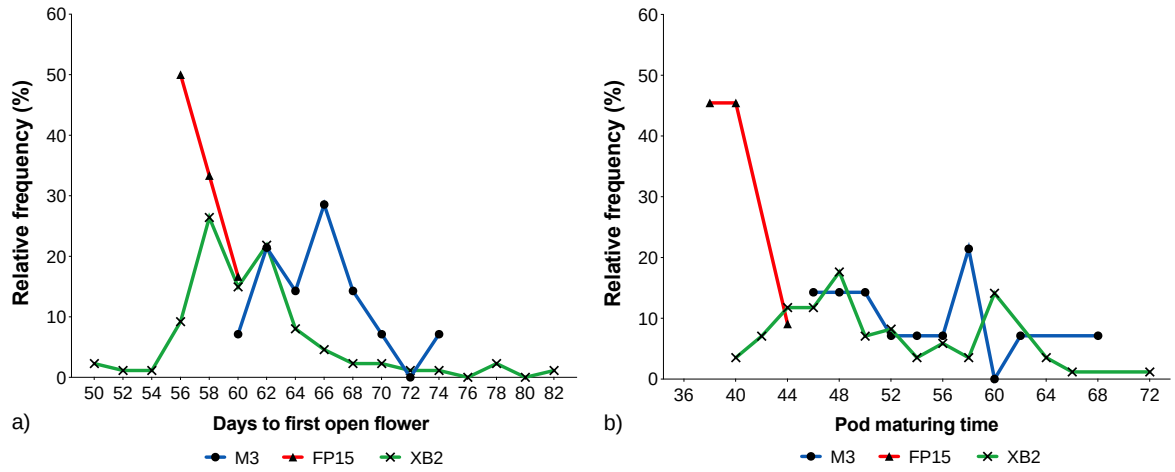


Fig. 3.6: *Days to first open flower* (Left) and *Pod maturing time* (Right) frequency distribution (%) for parental (M3; FP15) and F_2 (XB2) individuals expressed in days from first open flower to first harvested mature pod. Measurement points at joined to facilitate interpretation only.

Pod maturing time (PMT) was significantly different between parents, with the female parent FP15 producing mature pods in fewer days (38.9±1.8, *Mdn*= 39.0) than M3 (53.3±7.0; *Mdn*= 52.5) (*U*= .000; *p*< .01; *r*= -.85). The F_2 showed a non-normal frequency distribution, with 50.4±7.0 days (*Mdn*= 48.0) required to produce mature pods after flowering onset. Only 3.5% of individuals produced mature pods in 39-40 days, a maturing timing similar to FP15, and a positive correlation was found between DtF and PMT (r_s= .25; *p*< .01) (Fig. 3.6b).

3.2 Yield-related traits

Dry Pod Length (PoL), *Pods Per Plant* (PPP), *Seeds Per Pod* (SPP), and *Hundred-Seed Weight* (HSW) showed normal trait distributions in the parental genotypes (*p*> .05). FP15 had significantly longer and heavier pods, fewer in number, but each carrying more seeds compared to the male M3 parent (*p*< .01). On the other hand, HSW did not show a significant difference between the parents (*p*= .326) (Table 3.3).

Table 3.3: Harvest traits summary. Different letters (a, b) for parental values equals to statistically different values (see results section).

	Pod Length (PoL) (cm)	**Pods Per Plant (PPP)**	**Seeds Per Pod (SPP)**	**100-seed Weight (HSW) (g)**
M3	17.7±2.2 [a]	14.7±6.8 [a]	10.3±2.0 [a]	35.7±5.4
FP15	25.5±2.2 [b]	7.8±2.8 [b]	13.3±2.1 [b]	33.7±3.5
XB2(F_2)	18.4±2.4	14.4±7.3	8.0±2.5	39.6±5.7

All the mentioned harvest-related traits exhibited normal distribution in the F_2 segregating population, except for PPP ($p < .05$). Noteworthy was the presence of several undeveloped seeds within the F_2 pods: these had a normal shape, but their development had stopped, leaving them with a diameter less than 1mm. Pods, however, had an average length between the parental values (Table 3.3).

Pearson's test revealed a significant positive correlation between PoL and SPP ($r = .67$; $p < .001$), confirmed also by a regression analysis ($F_{(1,81)} = 66.01$; $p < .001$; $\beta = .67$) (Fig. 3.7).

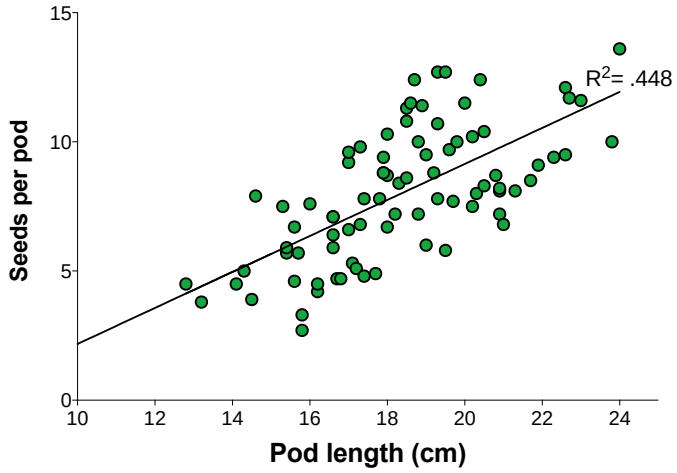


Fig. 3.7: Regression between fully-developed *seeds per pod* (SPP) and *dry pod length* (PoL).

Pods Per Plant had a significant ($p < .001$) positive correlation with *number of branches* ($r_s = .38$) (see also Fig. 3.8), and *length of branch* ($r_s = .44$). However, the effect of incremental branches on final PPP did not appear consistent, while it showed to increase more consistently the *number of leaves per plant* (Fig. 3.7). Among the 20 individuals with highest pod productivity (top 25% of F_2 population, "Q_3" in Fig. 3.8): 1 (5%) had a single branch, 3 (15%) had 2 branches, 8 (40%) carried only 3 branches, while just 4 individuals (20%) had 4 branches, and another 4 (20%) had 6 branches.

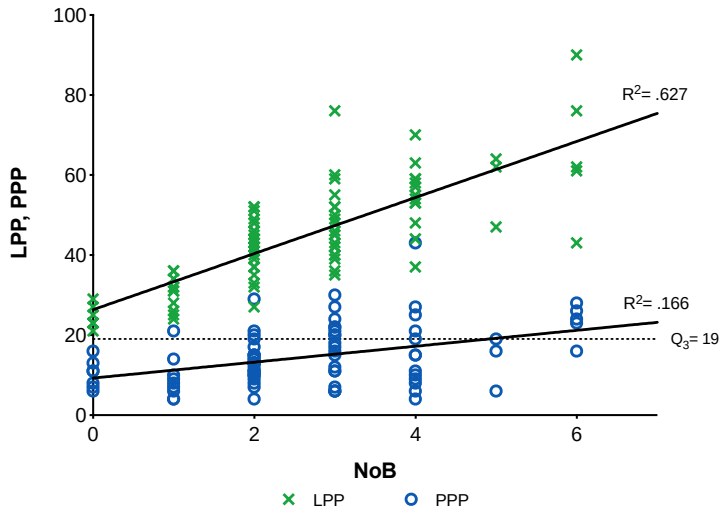


Fig. 3.8: Regression for *number of leaves per plant* (LPP) and *pods per plant* (PPP) against *branch number per plant* (NoB). The dotted line represents the third quartile (Q_3) splitting the highest 25% from the rest of data for PPP across the entire F_2 population.

On the other hand, all individuals with 3 NoB carried significantly less PPP (*Mdn*= 17) than those with 6 NoB (*Mdn*= 24) (*U*= 19.0; *p*< .05; *r*= -.41), and doubling the NoB brought an average 43% increase in PPP (see Table 3.4). This seemed supported by significantly higher SLB, which has in turn led to higher LPP values as well (see Table 3.4 for analysis results, and Fig. 3.8).

Table 3.3: Comparison for selected morphological traits between 3-branch and 6-branch individuals, with results from single Mann-Whitney U tests. NoB: *branch per plant*; LoB: *length of branch*; SLB: *total sum of branch length*; LPP: *number of leaves per plant*; PPP: *mature pods per plant*; SPP: *seeds per pod*.

	LoB (cm)	**SLB (cm)**	**LPP**	**PPP**	**SPP**
3 NoB 20 individuals	113.4±49.2	339.9±147.7	47.7±9.7	16.4±7.3	8.1±3.2
6 NoB 5 individuals	86.7±9.3	520.0±56.2	66.4±17.6	23.4±4.6	8.1±1.9
Change (%)	-24	53	39	43	-1
Mann-Whitney U test (p)	.209	.014*	.017*	.042*	.915
Effect size (r)		.49	.48	.41	

For the remaining yield-related traits, PPP had an inverse, but not significant correlation (r_s= -.21; *p*= .056) with SPP and HSW. HSW instead had a significant (*p*< .01) inverse relationship with the number of fully developed seeds per pod (*r*= -.28). The only significant correlation among yield-related and developmental traits was found between *Days to first open flower* and HSW (r_s= -.23; *p*< .05). Pearson's and Spearman's correlations between all measured traits have been reported in supplementary information (Table S.1).

4 Discussion

Winged bean has a plant architecture determined by the development and growth rate of stem phytomers, and the axillary meristem harboured on each of them. In the present cross, the recombinant population showed how *stem length* and *weekly growth* were correlated to *internode length*. Similar traits, here controlled by multiple genes, have already been a major target during the green revolution, with decreases in height that in some cases translated into increased fertility and yield in crops such as wheat (Harberd and Freeling 1989; Lester et al. 1997; Peng et al. 1999; Sasaki et al. 2002; Hedden 2003; de Saint Germain et al. 2013; Allen et al. 2018). Altering winged bean architecture through *internode length* could allow selection for plant height independently from the *number of nodes*. Given the positive correlation between this latter trait and *branch number per plant*, there is potential to breed for shorter individuals without affecting lateral growth, as is also supported by the two distinct components found in both PCA. When using phenotypic data from parental and F_2 individuals (Fig. 3.5), PCA also showed a pattern in agreement with the profiles emerged from the reported values, with FP15 having longer main stems and internodes, and with fewer and shorter branches that result in a less "bushy" individuals. In contrast, M3 showed shorter architecture, but with more and longer branches leading to higher LPP. The M3 individuals also showed a relatively larger phenotypic variation, perhaps underlied by higher genetic variation as compared to the female parent.

Branching, here described through number and length of lateral branches, is a mechanism controlled by multiple endogenous factors, and their interactions with the environment at different development stages. Across the F_2 population, *branch number per plant* appeared multigenic, with a mean above the parent carrying the highest number. The average value, above the highest parent (M3), appears in agreement with what was previously reported by Eagleton (Eagleton 1983) across 3 F_2s from 3 diallel crosses assessed in 2 different environments (Malaysia and Australia). In that study, the NoB trait reported the highest inheritance across the analysed traits (h^2= 65.5, estimate based on variance components using Jinks-Hayman diallel analysis of F_1 generation) along with a proposed additive gene effect with dominance directed towards higher branch number, and minimal environmental effect, later confirmed across three F_2 populations.

For the relationship between *branch number per plant* and *pods per plant*, the overall trend would suggest that more branches lead to more pods. However, the presence of high pod-yielding individuals with 2 to 3 branches points towards the possibility of having fewer branches and maintaining high *pods per plant*. Increasing the *length of branch*, actually, could bring a greater benefit to final pod number, as there was a stronger correlation between these two traits than between NoB and PPP. A possible explanation for these results could be the presence, on each branch, of the unreproductive nodes that are produced below the growing shoot meristem. This could become a significant limitation in shorter branches and result in vegetative biomass growth, without contributing to final pod development and yield. Therefore, a possible ideotype for winged bean could include a low number of lateral branches, but each with higher average length (*length of branch*, LoB). Such a phenotype would harbour more reproductive organs, while limiting vegetative biomass growth that follows from higher NoB. This could also reduce the number of leaves, a mechanism that could be compensated at least partially by the observed increase in leaf size. Furthermore, from the open flower stage to a growing pod stage there are traits, such as the rate of successful fertilization, or pod photosynthetic activity, that could also be targeted in breeding programmes to improve pod productivity, without having to rely on morphological traits alone. Examples for this could be the individual XB2-147 and XB2-211: the first having 3 lateral branches on average 138cm long, and producing 27 pods while having 40 leaves; the second, carrying 6 branches, on average 98cm long, and producing one additional pod (28 in total) but with an additional 21 leaves (61 in total). Another individual, XB-237, showed a relatively short structure (*stem length* of 127cm), 3 branches on average 101cm long, producing 22 pods with 52 leaves. Selection, and further investigations, could look into how to breed for individuals with a high pod productivity and curbed vegetative growth.

In the present study were also reported individuals having branches departing from low nodes and able to grow above the main stem, with axillary meristems behaving as if there was a loss of the main shoot apical dominance. This phenomenon, possibly inherited from the M3 parent, was only previously reported in winged bean individuals where apical dominance was probably lost due to horizontal, unsupported growth of the main shoot (Schiavinato and Válio 1996b). In the present case it could be rather due to changes in regulatory mechanisms acting on axillary meristem growth. These could

include gene expression regulatory networks, hormones, secondary messengers, or acropetal signals from root system (Foo et al. 2001; Symons et al. 2002; Mason et al. 2014; Wang 2014; Teichmann and Muhr 2015), and further studies could elucidate the underlying mechanism involved here.

Days to first open flower appears to be polygenic, with the F_2 population average closer to the later-flowering parent M3. This trait, however, showed transgressive segregation in both directions, with individuals flowering earlier and later than the earliest (FP15) and latest (M3) parental genotype respectively. The negative correlations reported between DtF and *branch number per plant* would encourage selection of early flowering individuals, which are unaffected in structure and capable of yielding a great number of pods, and seed yield. At the same time, early flowering individuals were also shown to produce mature pods earlier (*pod maturing time*), eventually shortening the time from emergence to harvest. An example could be XB2-75, reporting a DtF value of 55, PMT of 47 days, 2 lateral branches and producing 21 pods about 21cm long. Some of the F_2 individuals with potentially useful combinations of morphological, developmental, and yield-related traits are reported in Table 4.2. Future studies could investigate such combinations in later generations, and further elucidate the effects on green pod and seed yield.

Table 4.1: F_2 individuals with potentially useful combinations of morphological, developmental, and yield-related traits. Parental values (M3 and FP15) have been reported as reference. InL(StL*): internode length* (*stem length*); NoB (LoB): *branch per plant (length of branch)*; LPP: *number of leaves per plant*; DtF: *days to first open flower*; PMT: *pod maturing time*; PPP: *mature pods per plant*; PoL: *dry pod length*.

	InL (StL) (cm)	**NoB (LoB) (cm)**	**LPP**	**DtF (days)**	**PMT (days)**	**PPP**	**PoL (cm)**
M3	7.2±.1.8 (200.4±64.3)[a]	2.4±1.2 (134±53.3) [a]	45.8±12.9 [a]	65.1±3.3 [a]	53.3±7.0 [a]	14.7±6.8 [a]	17.7±2.2 [a]
FP15	12.9±1.0 (322.3±13.1)[b]	0.3±0.6 (12.8±30.0) [b]	23.9±3.5 [b]	56.9±1.4 [b]	38.9±1.8 [b]	7.8±2.8 [b]	25.5±2.2 [b]
XB2-75	10.4 (330)	2 (167.5)	52	55	47	21	20.9
XB2-147	11.3 (326)	3 (138.7)	40	60	42	27	20.9
XB2-237	5 (127)	3 (101.3)	45	64	52	22	20.2

*: only 2 individuals developed lateral branches.

Pod length and *seeds per pod* were found to be positively correlated, indicating that *pod length* could be used as a direct target trait for increasing seed yield. This correlation is reported although the F_2 showed a relatively low *seeds per pod* mean, below the lower parent (M3) despite having similar *pod*

lengths. Such an observation is consistent with the previous F_1 hybrids (5.5±0.9 *seeds per pod*), while *hundred-seed weight* appeared consistently higher in both generations, following the negative correlation between these two traits. The mechanism controlling the development of pod morphology could be, at least partially, independent from the final number of fully developed seeds inside the pod itself. Undeveloped seeds appeared, indeed, as if arrested early during their filling, perhaps after the switch to filial control over seed development (Weber et al. 2005).

Finally, the major component of seed yield, here PPP, SPP, and HSW usually have negative correlations between each other, likely because they act as competing sink for dry matter partitioning (Cai et al. 2016). In the present study such correlations were found as well, although only HSW and SPP reported significant correlation values. An increase in *pods per plant* and *seeds per pod* could be the way towards higher pod and seed-yielding winged bean individuals, while further studies could focus on assimilate partitioning during pod filling.

5 Conclusions

Winged bean has received limited focus for the improvement of plant architecture, despite its broad cultivation and potential benefits in Asia and the Pacific areas food systems. The population reported here demonstrated that there is the possibility to improve and select for traits that most effect the cultivation of this leguminous species. Shorter individuals, with fewer but longer lateral branches, yielding pods and seed in less time might be obtained through selection without having these traits affecting each other negatively. Further studies can focus on the mechanisms involved in the release of dormant axillary meristems, different planting densities, and investigate more in depth the correlation between yield-related traits and final harvest.

Acknowledgements

This work was financially supported by Crops For the Future (CFF), and the University of Nottingham Malaysia Campus (UNMC) through the CFF-UNMC Doctoral Training Partnership scheme. Seeds were obtained from the Malaysian Agricultural Research and Development Institute (MARDI), and by our collaborators in Malaysia. The authors would like to thank Dr Graham Eagleton for helping in the

design of the trellis structure, and the team at the Field Research Centre (CFF-FRC) for helping in setting up the experimental plots.

Compliance with ethical standards

Conflict of interest. The authors declare that they have no conflict of interest**.**

References

Adhikari L, Hussain A, Rasul G (2017) Tapping the potential of neglected and underutilized food crops for sustainable nutrition security in the mountains of Pakistan and Nepal. Sustain 9:. doi: 10.3390/su9020291

Allen LH, Zhang L, Boote KJ, Hauser BA (2018) Elevated temperature intensity, timing, and duration of exposure affect soybean internode elongation, mainstem node number, and pod number per plant. Crop J 6:148–161. doi: 10.1016/j.cj.2017.10.005

Ando K, Grumet R, Terpstra K, Kelly JD (2007) Manipulation of plant architecture to enhance crop disease control. CAB Rev Perspect Agric Vet Sci Nutr Nat Resour 2:1053–63. doi: 10.1079/PAVSNNR20072026

Andrivon D, Giorgetti C, Baranger A, et al (2013) Defining and designing plant architectural ideotypes to control epidemics? Eur J Plant Pathol 135:611–617. doi: 10.1007/s10658-012-0126-y

Anonymous (1982) The Winged Bean Flyer (4 No.1). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Anonymous (1981) The Winged Bean Flyer (3 No.2). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Ávila CM, Ruiz-Rodríguez MD, Cruz-Izquierdo S, et al (2017) Identification of plant architecture and yield-related QTL in Vicia faba L. Mol Breed 37:88. doi: 10.1007/s11032-017-0688-7

Braun N, de Saint Germain A, Pillot J-P, et al (2012) The Pea TCP Transcription Factor PsBRC1 Acts Downstream of Strigolactones to Control Shoot Branching. PLANT Physiol 158:225–238. doi: 10.1104/pp.111.182725

Cai G, Yang QQ, Chen H, et al (2016) Genetic dissection of plant architecture and yield-related traits in Brassica napus. Sci Rep 6:21625. doi: 10.1038/srep21625

Challinor AJ, Watson J, Lobell DB, et al (2014) A meta-analysis of crop yield under climate change and adaptation. Nat Clim Chang 4:287–291. doi: 10.1038/nclimate2153

Chappell MJ, LaValle LA (2011) Food security and biodiversity: Can we have both? An agroecological analysis. Agric Human Values 28:3–26. doi: 10.1007/s10460-009-9251-4

Chivenge P, Mabhaudhi T, Modi AT, Mafongoya P (2015) The potential role of neglected and underutilised crop species as future crops under water scarce conditions in Sub-Saharan Africa. Int J Environ Res Public Health 12:5685–5711. doi: 10.3390/ijerph120605685

de Saint Germain A, Ligerot Y, Dun EA, et al (2013) Strigolactones Stimulate Internode Elongation Independently of Gibberellins. Plant Physiol 163:1012–1025. doi: 10.1104/pp.113.220541

Dwivedi SL, Ceccarelli S, Blair MW, et al (2016) Landrace Germplasm for Improving Yield and Abiotic Stress Adaptation. Trends Plant Sci 21:31–42. doi: 10.1016/j.tplants.2015.10.012

Dwivedi SL, Lammerts van Bueren ET, Ceccarelli S, et al (2017) Diversifying Food Systems in the Pursuit of Sustainable Food Production and Healthy Diets. Trends Plant Sci 22:842–856. doi: 10.1016/j.tplants.2017.06.011

Eagleton GE (1983) Evaluation of genetic resources in the winged bean (*Psophocarpus tetragonolobus* (L.) DC.) and their utilisation in the development of cultivars for higher latitudes. University of Western Australia

Erskine W (1980) Measurements of the cross-pollination of winged bean in Papua New Guinea. SABRAO J. 12:11–14

Erskine W, Bala AA (1976) Crossing technique in winged bean. Trop. Grain Legum. Bullettine 6:32–35

Foo E, Turnbull CG, Beveridge CA (2001) Long-Distance Signaling and the Control of Branching in the rms1 Mutant of Pea. Plant Physiol 126:203–209. doi: 10.1104/pp.126.1.203
Godfray HCJ, Beddington JR, Crute IR, et al (2010) Food Security: The Challenge of Feeding 9 Billion People. Science (80-) 327:812–818. doi: 10.1126/science.1185383
Hajjar R, Jarvis DI, Gemmill-Herren B (2008) The utility of crop genetic diversity in maintaining ecosystem services. Agric Ecosyst Environ 123:261–270. doi: 10.1016/j.agee.2007.08.003
Harberd NP, Freeling M (1989) Genetics of dominant gibberellin-insensitive dwarfism in maize. Genetics 121:827–838
Hedden P (2003) The genes of the Green Revolution. Trends Genet 19:5–9. doi: 10.1016/S0168-9525(02)00009-4
Herath HMW, Ormrod DP (1979) Effects of Temperature and Photoperiod on Winged Beans [*Psophocarpus tetragonolobus* (L.) D.C.]. Ann Bot 43:729–736. doi: 10.1093/oxfordjournals.aob.a085686
Jackson LE, Pascual U, Hodgkin T (2007) Utilizing and conserving agrobiodiversity in agricultural landscapes. Agric Ecosyst Environ 121:196–210. doi: 10.1016/j.agee.2006.12.017
Karikari SK (1972) Pollination requirements of winged beans (*Psophocarpus* spp. Neck) in Ghana. Ghana J Agric Sci 5:235–239
Khan TN (1976) Papua New Guinea: A centre of genetic diversity in winged bean (*Psophocarpus tetragonologus* (L.) D c.). Euphytica 25:693–705. doi: 10.1007/BF00041608
Khoury CK, Bjorkman AD, Dempewolf H, et al (2014) Increasing homogeneity in global food supplies and the implications for food security. Proc Natl Acad Sci 111:4001–4006. doi: 10.1073/pnas.1313490111
Lachat C, Raneri JE, Smith KW, et al (2017) Dietary species richness as a measure of food biodiversity and nutritional quality of diets. Proc Natl Acad Sci 115:201709194. doi: 10.1073/pnas.1709194115
Lepcha P, Egan AN, Doyle JJ, Sathyanarayana N (2017) A Review on Current Status and Future Prospects of Winged Bean (*Psophocarpus tetragonolobus*) in Tropical Agriculture. Plant Foods Hum Nutr 72:225–235. doi: 10.1007/s11130-017-0627-0
Lester DR, Ross JJ, Davies PJ, Reid JB (1997) Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell 9:1435–43. doi: 10.1105/tpc.9.8.1435
Mason MG, Ross JJ, Babst BA, et al (2014) Sugar demand, not auxin, is the initial regulator of apical dominance. Proc Natl Acad Sci 111:6092–6097. doi: 10.1073/pnas.1322045111
NAS (1975) The Winged Bean: A High-Protein Crop for the Humid Tropics. National Academy of Sciences, Washington, D.C.
Padulosi S, Thompson J, Rudebjer P (2013) Fighting poverty, hunger and malnutrition with neglected and underutilized species: needs, challenges and the way forward. Biodiversity International, Rome
Pellegrini L, Tasciotti L (2014) Crop diversification, dietary diversity and agricultural income: Empirical evidence from eight developing countries. Can J Dev Stud 35:211–227. doi: 10.1080/02255189.2014.898580
Peng J, Richards DEDE, Hartley NMNM, et al (1999) "Green revolution" genes encode mutant gibberellin response modulators. Nature 400:256–261. doi: 10.1038/22307
Powell B, Thilsted SH, Ickowitz A, et al (2015) Improving diets with wild and cultivated biodiversity from across the landscape. Food Secur 7:535–554. doi: 10.1007/s12571-015-0466-5
Ray DK, Gerber JS, MacDonald GK, West PC (2015) Climate variation explains a third of global crop yield variability. Nat Commun 6:1–9. doi: 10.1038/ncomms6989
Sasaki A, Ashikari M, Ueguchi-Tanaka M, et al (2002) Green revolution: A mutant gibberellin-synthesis gene in rice. Nature 416:701–702. doi: 10.1038/416701a
Schiavinato MA, Válio IFM (1996a) Influence of photoperiod and temperature on the development of Winged Bean plants. RBrasFisiolVeg 8:105–110
Schiavinato MA, Válio IFM (1996b) Influence of staking on development of Winged Bean plants. RBrasFisiolVeg 8:99–103
Schreinemachers P, Simmons EB, Wopereis MCS (2018) Tapping the economic and nutritional power of vegetables. Glob Food Sec 16:36–45. doi: 10.1016/j.gfs.2017.09.005
Sun Z, Su C, Yun J, et al (2018) Genetic improvement of the shoot architecture and yield in soya bean

plants via the manipulation of GmmiR156b. Plant Biotechnol J 0–1. doi: 10.1111/pbi.12946
Symons GM, Ross JJ, Murfet IC (2002) The bushy pea mutant is IAA-deficient. Physiol Plant 116:389–397. doi: 10.1034/j.1399-3054.2002.1160315.x
Teichmann T, Muhr M (2015) Shaping plant architecture. Front Plant Sci 6:1–18. doi: 10.3389/fpls.2015.00233
Teixeira FF, Ramalho MAP, De Fátima Barbosa Abreu Â (1999) Genetic control of plant architecture in the common bean (*Phaseolus vulgaris* L.). Genet Mol Biol 22:577–582. doi: 10.1590/S1415-47571999000400019
Thrupp LA (2000) Linking Agricultural Biodiversity and Food Security: the Valuable Role of Agrobiodiversity for Sustainable Agriculture. Int Aff 76:283–297. doi: 10.1111/1468-2346.00133
UN (2015) Transforming our world: the 2030 agenda for sustainable development (A/RES/70/1). United Nations
Wang J-W (2014) Regulation of flowering time by the miR156-mediated age pathway. J Exp Bot 65:4723–4730. doi: 10.1093/jxb/eru246
Wang Y, Li J (2008) Molecular Basis of Plant Architecture. Annu Rev Plant Biol 59:253–279. doi: 10.1146/annurev.arplant.59.032607.092902
Weber H, Borisjuk L, Wobus U (2005) Molecular Physiology of Legume Seed Development. Annu Rev Plant Biol 56:253–279. doi: 10.1146/annurev.arplant.56.032604.144201
Weeden NF (2007) Genetic changes accompanying the domestication of Pisum sativum: Is there a common genetic basis to the “domestication syndrome” for legumes? Ann Bot 100:1017–1025. doi: 10.1093/aob/mcm122
Weil RR, Khalil NA (1986) Salinity Tolerance of Winged Bean as Compared to that of Soybean. Agron J 78:67. doi: 10.2134/agronj1986.00021962007800010015x
Wong KC (1983) Effects of daylength, temperature, light intensity and applied growth substances on the growth, flowering and tuberization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.). Universiti Putra Malaysia
Wong Q, Tanzi AS, Ho W, et al (2017) Development of Gene-Based SSR Markers in Winged Bean (Psophocarpus tetragonolobus (L.) DC.) for Diversity Assessment. Genes (Basel) 8:100. doi: 10.3390/genes8030100

Appendix – Chapter 3

a.1 Controlled crosses

Crosses were performed in controlled environment at the Crops For the Future (CFF) facility in Semenyih, Malaysia (April-October 2016). Pots, with diameter of 40cm and 32cm height, were placed about 10cm from each other, with a 2m height single stalk placed in each of them. The two chambers were set at maximum 27°C during the day, and 22°C during night, with a 12h photoperiod (light on). The maximum light intensity was set at 1000µmol/m²s, starting from 300µmol/m²s at 7AM in the morning, gradually increase till reaching maximum intensity at 11AM. This was kept constant until 3PM, when light intensity was gradually decreased till 7PM, when lights were set to switch off. Humidity was kept above 65%RH.

Crosses were performed following the technique developed by Erskine and Bala (1976). The authors obtained a maximum success fertilization rate of 57.9% in shade house conditions, against 20.9% in field conditions, measured as the number of crossed flowers eventually producing seeds. Here, we obtained 45 and 53%, respectively in Pop.1 and 2, of crossed flowers eventually producing pods that have been harvested. Pod set (i.e. crossed flowers that develop a pod) was 64 and 66%.

Table a.1: Controlled crosses success rates across the two crossing populations (Pop. 1 and 2), consisting each of a common male (M3) and 4 female parents. Crosses were performed in controlled environment between April and October 2016. Pod set was calculated upon number of observed pods 5-6cm long on crossed flowers. Pod harvest is the total number of pods harvested from crossed flowers.

		Crosses Performed	Pod Set (%)	Pods Growing/ Pods set 4 weeks after crossing (%)	Pod Harvest (%)
			M3-3b (male)		
Pop. 1	FP15-10-3-4	16	12 (75)	7 (58.3)	7 (44)
	TPt10-7-5-2	25	20 (80)	17 (85)	17 (68)
	TPt17-6-3-4	33	25 (75.8)	20 (80)	20 (61)
	TPt53-9-8-3	36	13 (36.1)	6 (46.2)	6 (17)
	Total	110	70 (63.6)	50 (71)	50 (45)
			M3-8b (male)		
Pop. 2	FP15-10-3-2	15	10 (66.7)	9 (90)	9 (60)
	TPt10-7-5-1	29	20 (69.0)	19 (95)	19 (66)
	TPt17-6-3-10	23	19 (82.6)	17 (89.5)	17 (74)
	TPt53-9-8-4	39	21 (53.8)	11 (52.4)	11 (28)
	Total	106	70 (66.0)	56 (78.6)	56 (53)

a.2 Selection of parental combination to establish the F_2 population

Eight replicates for all 5 parental genotypes, and for the 4 hybrid combinations were grown in a complete randomised design (CRD) between October 2016 and March 2017. Ridges and structures were arranged in the same way as described in the material and methods section in chapter 3. Unfortunately, a bacterial infection started to appear between week 6 and 7, with symptoms including wilting of leaves without preliminary yellowing (Figure a.1 left), epinasty, followed by complete drying and death of leaf tissues and the whole plant in one to three days. A whitish exudate was revealed through bacterial streaming test. Bacteria oozing from washed stems (Figure a.1 centre) were collected and grown on nonselective media (pseudomonas agar base, Thermo Fisher Scientific) following procedure described at p.23 in ICRISAT (1995). Staining of the colonies resulted in a gram-negative rod shaped bacteria (Figure a.1 right), corresponding profile of *Pseudomonas/ Ralstonia* spp.

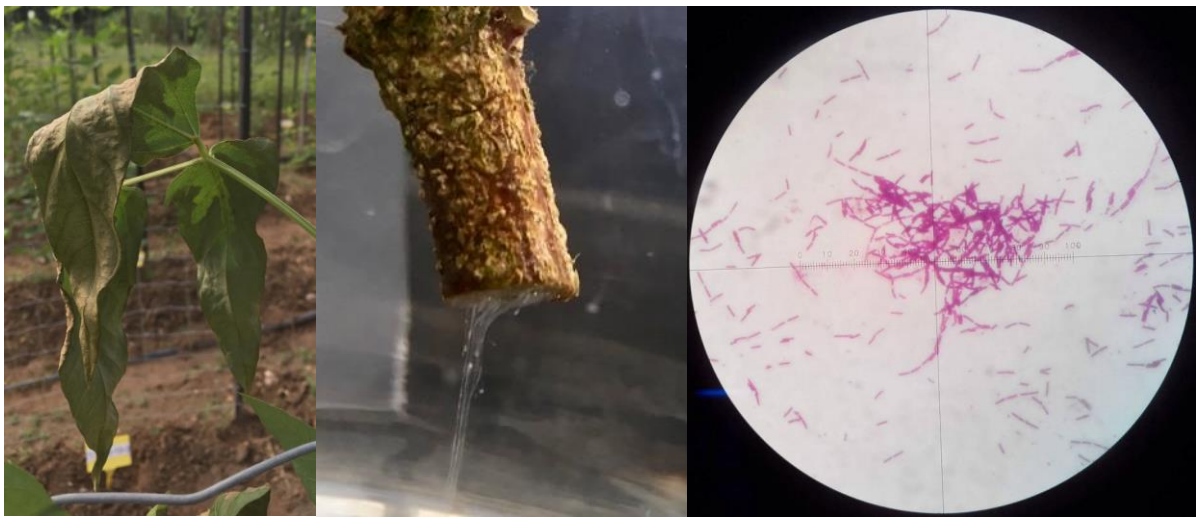


Figure a. 1: (Left) Wilting leaf in infected plant; (Centre) Oozing bacteria from cut stem; (Right)

Due to this infection, a number of individuals were lost during the experiment, limiting the analysis and interpretation of the results by using only 3 to 4 replicates for each parental and hybrid genotype. Table a.2 presents results obtained for the measured traits across replicates. A consistent criteria for measuring *branch number per plant* (NoB), as the one used during the phenotyping of the F_2 individuals, was not developed at the time of the F_1 assessment. In particular, counting did not distinguish between short (1-2cm), non-growing axillary buds, and fully growing branches. Hence measurements for this trait could not be used and have not been reported.

From this data, the most promising parental genotypes were selected to develop the F_2 population. Based on the differences observed across multiple traits including morphological (StL, *6th Internode*, *InL678*, NoN, LPP), yield-related (PPP, SPP), and developmental (DtF, PMT), along with the observed behaviour of their offspring, the parental genotypes M3 and FP15, and their hybrids, were selected to establish the F_2 population.

Table a. 2: Average (± SD) of the measured traits during F_1 field phenotyping. M3 values have been reported each time grouped with a different genotyped used as females, and the respective hybrid.

Genotype	StL	6th Internode	InL678*	NoN	LPP	PPP	SPP	PoL	HSW	DtF	PMT
M3	261 ± 47.1	6.7 ± 0.6	8 ± 1.3	36 ± 2	103.7 ± 8.4	62 ± 10.4	11.4 ± 0.4	21.2 ± 0.6	42.3 ± 2.1	54 ± 1	53 ± 2
FP15	321.7 ± 30.7	13.3 ± 0.6	11 ± 0.9	27.3 ± 1.5	59.7 ± 6.7	43 ± 21.2	15.5 ± 1.4	29 ± 0.6	39.4 ± 0.9	48.3 ± 2.1	45 ± 11.3
XFP	353 ± 33.8	10.5 ± 4.5	10.5 ± 4.1	34.5 ± 1	78.5 ± 17.4	66 ± 55.5	5.8 ± 1.5	21.3 ± 1.6	42.2 ± 8.1	48.5 ± 1	45.5 ± 5.2
M3	261 ± 47.1	6.7 ± 0.6	8 ± 1.3	36 ± 2	103.7 ± 8.4	62 ± 10.4	11.4 ± 0.4	21.2 ± 0.6	42.3 ± 2.1	54 ± 1	53 ± 2
TPt10	353.5 ± 32.6	6.6 ± 4.2	9 ± 4.5	34.3 ± 1	90 ± 19.2	81 ± 19.3	12.1 ± 0.6	19.3 ± 0.3	32.9 ± 2	47.5 ± 7.9	57.5 ± 3.1
XT10	316.8 ± 50.3	5 ± 2.4	7.8 ± 2.6	34 ± 2.6	115.3 ± 25.2	72 ± 28.8	13.5 ± 1	21 ± 1.2	35.6 ± 4.8	49.8 ± 2.2	53.3 ± 4.6
M3	261 ± 47.1	6.7 ± 0.6	8 ± 1.3	36 ± 2	103.7 ± 8.4	62 ± 10.4	11.4 ± 0.4	21.2 ± 0.6	42.3 ± 2.1	54 ± 1	53 ± 2
TPt17	298.3 ± 25.6	12.7 ± 3.1	12.7 ± 2.6	31 ± 2.6	82.3 ± 6.8	46	10.3	17.5	25.2	56.7 ± 2.5	45 ± 1.4
XT17	302.3 ± 18.9	13.7 ± 1.5	13.9 ± 1.3	28.7 ± 0.6	129.3 ± 28.6	42 ± 30.5	11.9 ± 1.2	20.6 ± 1.1	28.4 ± 6.8	54 ± 1.7	46 ± 0
M3	261 ± 47.1	6.7 ± 0.6	8 ± 1.3	36 ± 2	103.7 ± 8.4	62 ± 10.4	11.4 ± 0.4	21.2 ± 0.6	42.3 ± 2.1	54 ± 1	53 ± 2
TPt53	323.7 ± 6.5	7.7 ± 4.7	10.9 ± 3.5	29.3 ± 0.6	66.7 ± 15.5	63	5.4	22.2	41.1	55.7 ± 8.3	56
XT53	298.5 ± 20.5	4.5 ± 4.5	5.4 ± 4.1	32.3 ± 2.2	48 ± 16.1	33 ± 26.1	12.5 ± 1.3	24.6 ± 2	35.1 ± 11.6	58 ± 7.8	44.5 ± 4.1

*average measured across 6th, 7th, 8th internode length.

a.3 SSR markers genotyping of parents, and of selected hybrids

All parental genotypes were profiled with previously developed SSR markers (Wong et al. 2017): results are reported in the table below.

Table a.2: SSR fingerprinting for parental plants (population 2 from controlled crosses). (*)polymorphic locus between male and female parent, with homozygous alleles within each of them.

	ABI Scoring for Parental Genotypes (Population 2)				
Marker	**M3-8b**	**TPt10-7-5-1**	**TPt17-6-3-10**	**TPt53-9-8-4**	**FP15-10-3-2**
P27.2	205	205	199*	205	205
Pt65.1	273	273	267*	267*	267*
Pt67.1	293	293	296*	293	293
Pt76.1	203/209	203	203	209	209
Pt58	261	255/261	261	261	261
Pt78.1	294/306	306/309	306	309	306
Pt7.2	426	426/432	426	428*	426
Pt10	228	228	226*	228	228
Pt93.1	266	266	272*	266/272	266
Pt14	354	358*	352*	350*	358*
Pt99.2	195	189/195	195	189*	189*
Pt24	217	219*	217	219*	219*
Pt68.1	226	226	229*	226	223*
Pt85.1	279	276/279	276*	276*	276*
Pt53	315	315	315	309/315	312*
WB17	198	198	198	198	198
P43.2	195	199*	195	197*	199*
Pt1.1	331/339	335	339	339	335

From these, a hierarchical cluster analysis was performed with the Dice (i.e. Nei and Li) similarity coefficient. A dendrogram was generated using unweighted pair-group method with arithmetic mean (UPGMA) algorithm (see Figure a.2).

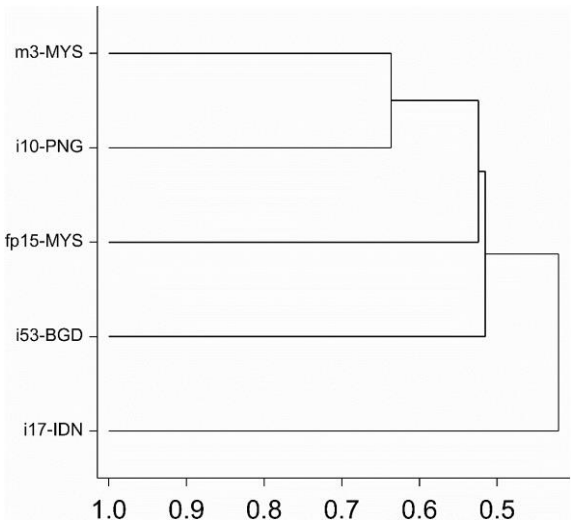


Figure a. 2: UPGMA dendrogram of the genetic relationship between genotypes used in crosses: M3-8b (m3), TPt10-7-5-1 (i10), TPt17-6-3-10 (i17), TPt53-9-8-4 (i53), FP15-10-3-2 (fp15).

In order to genotype hybrid individuals, 3 markers able to undergo multiplexing analysis were chosen. These were selected to be polymorphic between parental genotypes whereas homozygous within each of the two. PCR products from selected markers upon XFP hybrid genotypes were then separated using an ABI Genetic Analyser ABI3730XL (see chapter 2 for protocol): scoring has been summarised in Table a.2. Profiles of the putative hybrids corresponded to the expected ones for individuals derived from M3-8b X FP15-10-3-2 crosses, thus validating the crosses performed.

Table a.3: ABI scoring for M3-8b × FP15-10-3-2 hybrids genotypes (F_1).

		F_1 individuals							
Marker	**Expected profile**	**XFP_A1**	**XFP_A2**	**XFP_B1**	**XFP_B2**	**XFP_B3**	**XFP_C1**	**XFP_C2**	**XFP_C3**
Pt68.1 (AAC)	223/226	223/226	223/226	223/226	223/226	223/226	223/226	223/226	223/226
Pt53 (CGC)	312/315	312/315	312/315	-	312/315	312/315	-	312/315	312/315
Pt14 (TG)	354/358	354/358	354/358	354/358	354/358	354/358	354/358	354/358	354/358

a.4 Total branch number per plant (NoBTOT) across F_2 population

Total branch number per plant (NoBTOT) along the main stem, and its frequency distribution, for both parental (M3, FP15) and F_2 (XB2) individuals are reported in Figure a.3 and a.4. Only 29 F_2 individuals showed branches at upper nodes, while the correlation between NoBTOT and PPP was still significant, but lower than NoB-PPP: r_s= .280 (p< .01), instead on r_s= .35 (p< .001). The decreased effect size for the correlation might suggest that additional branches departing from higher nodes might not contribute as much as branches from basal nodes (i.e. those within first 10 nodes, NoB). None of the M3 individuals reported branches at higher nodes, while only one FP15 individual reported a branch on node 14 (see Figure a.3).

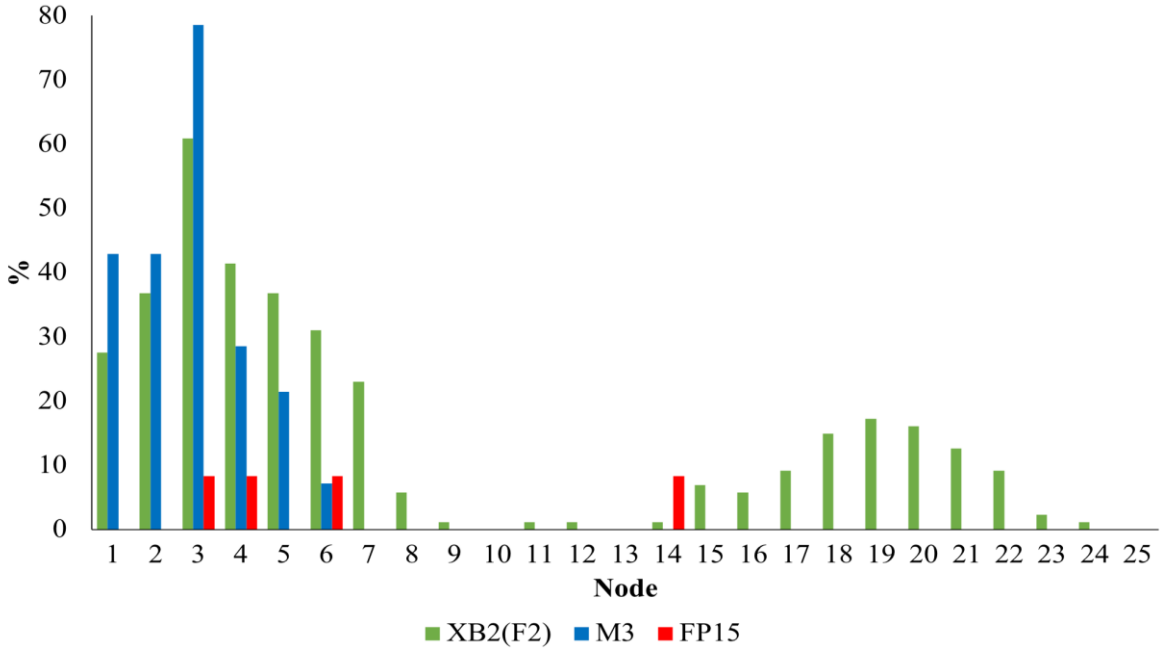


Figure a. 3: Distribution of branches across nodes along the main stem.

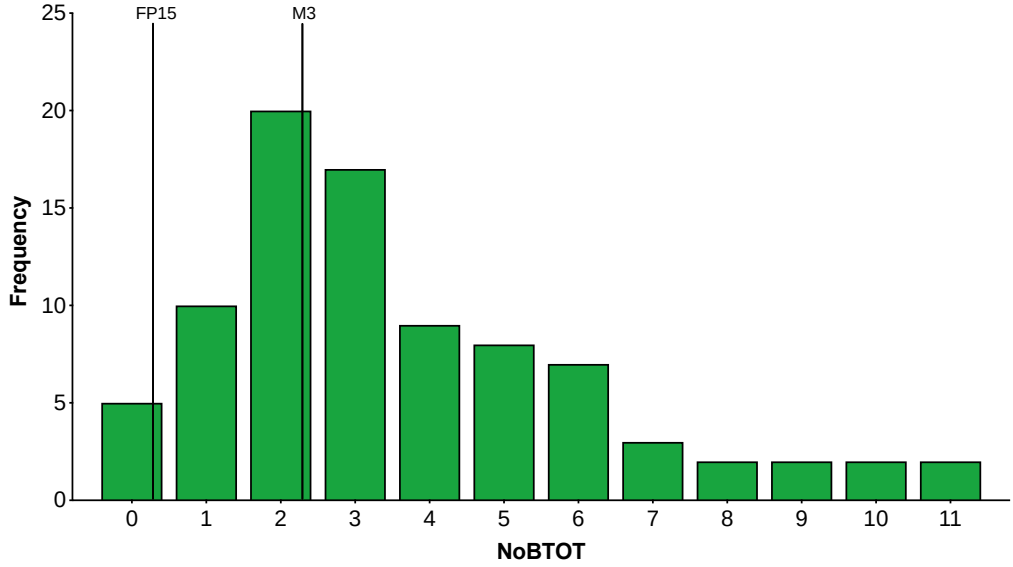


Figure a. 4: Histogram for phenotypic distribution of total number of branch per plant along the main stem of F_2 individuals.

a.5 Correlations across all traits from F_2 phenotypic analysis.

Table a. 5: Parametric (Pearson's) trait correlations from F_2 phenotypic analysis.

Pearson's	StL	WGR	InL	NoN	LPP	LeW	LeL	NoB	LoB	SLB	DtF	PMT	PPP	PoL	SPP	HSW
StL	1	.923**	.775**	.602**	.300**	0.025	-0.082	0.199	-0.064	0.117	-.318**	-0.127	.344**	-0.025	-0.075	0.083
WGR	.923**	1	.618**	.647**	0.167	-0.019	-0.061	0.116	-0.168	0.008	-.224*	0.024	0.177	-0.06	-0.102	0.127
InL	.775**	.618**	1	0.146	-0.092	0.094	0.01	-0.038	-0.05	-0.065	-.311**	-0.053	0.156	-0.078	-.241*	0.13
NoN	.602**	.647**	0.146	1	.283**	-0.06	-0.074	.255*	-0.184	0.107	-0.072	-0.004	.230*	-0.055	0.05	0.137
LPP	.300**	0.167	-0.092	.283**	1	-.451**	-.453**	.776**	.386**	.747**	-0.202	0.109	.451**	-.245*	0.012	-0.025
LeW	0.025	-0.019	0.094	-0.06	-.451**	1	.901**	-.466**	0.168	-0.149	-0.015	-0.05	0.124	.236*	-0.017	.284*
LeL	-0.082	-0.061	0.01	-0.074	-.453**	.901**	1	-.441**	0.143	-0.16	-0.027	-0.011	0.1	.294**	-0.008	.270*
NoB	0.199	0.116	-0.038	.255*	.776**	-.466**	-.441**	1	0.194	.714**	-.249*	0.202	.341**	-0.21	-0.089	0.046
LoB	-0.064	-0.168	-0.05	-0.184	.386**	0.168	0.143	0.194	1	.732**	-0.062	0.123	.374**	-0.138	-0.072	0.016
SLB	0.117	0.008	-0.065	0.107	.747**	-0.149	-0.16	.714**	.732**	1	-.232*	0.136	.516**	-.227*	-0.09	0.081
DtF	-.318**	-.224*	-.311**	-0.072	-0.202	-0.015	-0.027	-.249*	-0.062	-.232*	1	-0.017	-0.207	0.102	0.166	-0.157
PMT	-0.127	0.024	-0.053	-0.004	0.109	-0.05	-0.011	0.202	0.123	0.136	-0.017	1	-0.014	-0.103	0.043	-0.121
PPP	.344**	0.177	0.156	.230*	.451**	0.124	0.1	.341**	.374**	.516**	-0.207	-0.014	1	-0.042	-0.124	.220*
PoL	-0.025	-0.06	-0.078	-0.055	-.245*	.236*	.294**	-0.21	-0.138	-.227*	0.102	-0.103	-0.042	1	.670**	0.201
SPP	-0.075	-0.102	-.241*	0.05	0.012	-0.017	-0.008	-0.089	-0.072	-0.09	0.166	0.043	-0.124	.670**	1	-.284**
HSW	0.083	0.127	0.13	0.137	-0.025	.284*	.270*	0.046	0.016	0.081	-0.157	-0.121	.220*	0.201	-.284**	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table a. 6: Non-parametric (Spearman's) trait correlations from F_2 phenotypic analysis.

Spearman's	StL	WGR	InL	NoN	LPP	LeW	LeL	NoB	LoB	SLB	DtF	PMT	PPP	PoL	SPP	HSW
StL	1	.988**	.801**	.576**	.296**	-0.068	-0.153	.231*	-0.137	0.077	-.451**	-0.118	.329**	-0.047	-0.112	0.13
WGR	.988**	1	.719**	.622**	.279**	-0.093	-0.147	0.201	-0.181	0.035	-.411**	-0.038	.249*	-0.062	-0.133	0.152
InL	.801**	.719**	1	0.19	0.059	-0.019	-0.07	0.021	-0.074	-0.034	-.423**	-0.148	0.201	-0.066	-.228*	0.167
NoN	.576**	.622**	0.19	1	.336**	-0.112	-0.124	.291**	-0.159	0.118	-0.176	-0.026	.270*	-0.023	0.099	0.116
LPP	.296**	.279**	0.059	.336**	1	-.363**	-.412**	.794**	.407**	.777**	-0.206	0.181	.488**	-.279*	-0.022	0.02
LeW	-0.068	-0.093	-0.019	-0.112	-.363**	1	.891**	-.403**	0.212	-0.127	-0.122	-0.045	0.131	.240*	-0.015	0.214
LeL	-0.153	-0.147	-0.07	-0.124	-.412**	.891**	1	-.426**	0.199	-0.148	-0.146	-0.045	0.11	.312**	-0.002	0.211
NoB	.231*	0.201	0.021	.291**	.794**	-.403**	-.426**	1	0.204	.765**	-.250*	.234*	.346**	-.252*	-0.116	0.12
LoB	-0.137	-0.181	-0.074	-0.159	.407**	0.212	0.199	0.204	1	.720**	-0.084	0.145	.443**	-0.092	-0.036	-0.005
SLB	0.077	0.035	-0.034	0.118	.777**	-0.127	-0.148	.765**	.720**	1	-.221*	0.193	.528**	-.233*	-0.097	0.089
DtF	-.451**	-.411**	-.423**	-0.176	-0.206	-0.122	-0.146	-.250*	-0.084	-.221*	1	0.054	-0.208	0.094	0.205	-.231*
PMT	-0.118	-0.038	-0.148	-0.026	0.181	-0.045	-0.045	.234*	0.145	0.193	0.054	1	-0.024	-0.084	0.077	-0.089
PPP	.329**	.249*	0.201	.270*	.488**	0.131	0.11	.346**	.443**	.528**	-0.208	-0.024	1	-0.092	-0.133	0.169
PoL	-0.047	-0.062	-0.066	-0.023	-.279*	.240*	.312**	-.252*	-0.092	-.233*	0.094	-0.084	-0.092	1	.678**	0.167
SPP	-0.112	-0.133	-.228*	0.099	-0.022	-0.015	-0.002	-0.116	-0.036	-0.097	0.205	0.077	-0.133	.678**	1	-.288**
HSW	0.13	0.152	0.167	0.116	0.02	0.214	0.211	0.12	-0.005	0.089	-.231*	-0.089	0.169	0.167	-.288**	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

a.6 Chlorophyll concentration and observed top leaves colour

Between 20 to 30 days after emergence, in M3 individuals and across the F_2 population, the first 3 to 4 top open leaves displayed yellow or pale green colouration (Figure a.5). Newly developed leaves in the following weeks also showed the same colouration, while the older tissues slowly recovered to their "natural" (i.e. green) colour. Such phenotype ceased completely at 50 days after emergence. The same phenotype was observed in the originally crossed M3-8b parent (see Figure a.5), but not in F_1 individuals.

Figure a.5: Yellow leaves on male parent M3-8b at week 4 after emergence, during controlled crosses. Both top unopen and open leaves are visible, while older tissues display greener leaves.

Chlorophyll measurements (estimated using SPAD-502, Konica Minolta, Japan) were taken on the central leaflets of the first top 3 open leaves (3 measurements per leaflet), at 28 days after emergence. Visual scoring was also performed by labelling the average observed colour as "green", "pale green", or "yellow". A "green" visual score was estimated to be 34.52±2.35, "pale green" 23.26±5.37, and "yellow" 6.93±4.74. Frequency for SPAD values showed a non-normal distribution across the F_2 ($p > .05$). A Kruskal-Wallis test followed by single Mann-Whitney U tests (using a Bonferroni correction) revealed significant differences between green and pale green ($U = 29.00$; $p < .001$; $r = -.84$); pale green and yellow ($U = 13.00$; $p < .001$; $r = -.53$); and between green and yellow ($U = .00$; $p < .001$; $r = -.61$). Looking at the frequencies by genotype, at week 4 a.g. all FP15

individuals displayed green leaves and 35.6±2.1 chlorophyll content. In stark contrast, male M3 parent displayed individuals with pale green (79%) or yellow (21%) top leaves, with a lower chlorophyll mean content (14.6±8.4). The segregating population reported 41% of the individuals with green, 53% with pale-green, and 6% with yellow leaves, and an overall 27.4±7.8 chlorophyll content. These data, and the observation in the M3 individuals, would suggest the presence of more than one gene involved in chlorophyll biosynthesis that confer observed pale green/ yellow leaves phenotypes. Data of visual scorings and average chlorophyll content for each genotype is reported in the boxplot in figure a.6 below.

In the F_2 population, no significant correlation was found between SPAD content and StL, NoB, WGR, or HSW, suggesting that such trait did not affect growth in plants with yellow leaves phenotype.

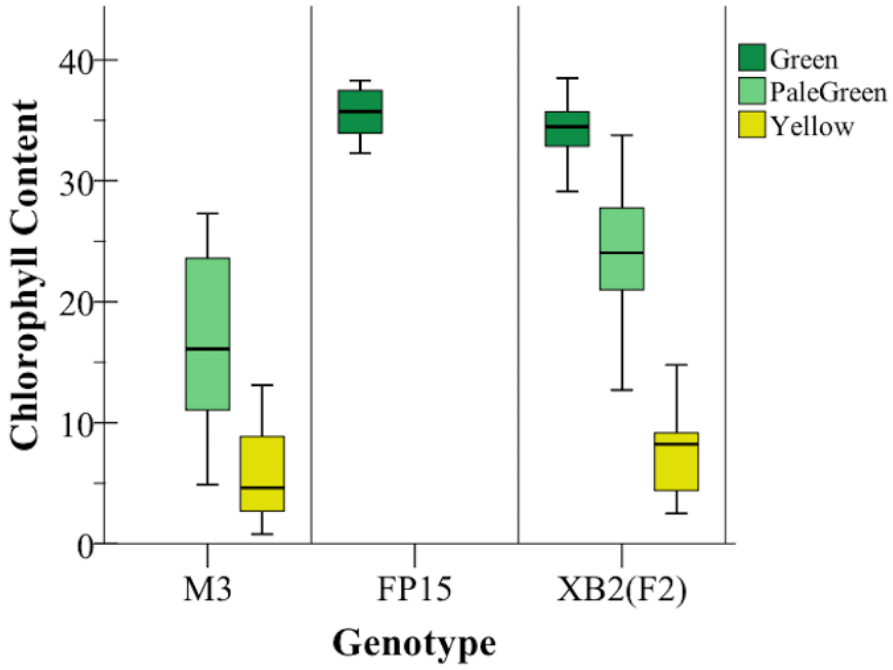


Figure a.6: Boxplot for visual scoring (Green, PaleGreen, Yellow top leaves) and average SPAD values for top 3 leaves in parental genotypes (M3; FP15) and F_2 population (XB2(F2)).

Chapter 4 - Construction of a linkage map in winged bean (*Psophocarpus tetragonolobus* (L.) DC.) from high-throughput genotyping of a biparental mapping population

Abstract

Only in recent years have efforts been made to develop tools such as molecular markers in winged bean (*Psophocarpus tetragonolobus* (L.) DC.; 2*n*= 2x= 18). So far these have been useful for initial genetic diversity assessment of germplasm. However, a linkage map has not been developed, despite the support that it could provide for more critical analysis such as dissection of agronomically important traits, marker-assisted selection, and transferring of information from other well-studied species. Therefore, the objective of the present study was to construct a genetic linkage map in a segregating F_2 population derived from the biparental cross of two winged bean accessions (M3 × FP15). Nine linkage groups (LGs), presumably corresponding to the 9 chromosomes of winged bean, were obtained at LOD scores of up to 30, with an average of 199 single nucleotide polymorphism (SNP) markers each (1937 SNP in total). Analysing the SNPs revealed about 75% of segregation distortion (SD) (at $p< 0.1$) across LGs, corrected to 54% after false-discovery rate (FDR) analysis was carried out, with further analysis showing that the distortion was primarily in favour of alleles derived from the male parent (M3). Mapping therefore proceeded by generating an ordered map with only non-distorted markers, to serve as framework for re-introducing and ordering of the distorted markers. Eventually, a genetic linkage map with a total length of 1171.6cM, with 395 spaced SNPs (3.1cM/SNP), and an overall 34% SD was obtained. This could be used for initial QTL analyses using previously recorded phenotypic traits.

1 Introduction

Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) is an underutilised legume crop grown predominantly on a small scale in hot and humid countries, in the Asia and South Pacific areas. It is grown for its immature pods, tuberous roots, and leaves (Stephenson et al. 1979; NAS 1981; Eagleton 1999; Lepcha et al. 2017), and it is taxonomically classified under the Fabaceae family, Papilionoideae subfamily. Its diploid genome (2*n* = 2x= 18) is divided among three pairs of short and six pairs of long chromosomes (Harder 1992) and was recently estimated around 1.22Gbp/C in size (Vatanparast et al. 2016), while previous estimation reported a

haploid genome size of 782Mbp (Bennett and Smith 1976). Molecular tools have been recently reported in winged bean, in particular molecular markers like Random Amplification of Polymorphic DNA (RAPD) and Inter-Simple Sequence Repeat (ISSR) (Koshy et al. 2013; Mohanty et al. 2013; Chen et al. 2015), or Simple Sequence Repeat (SSR) (Chapman 2015; Vatanparast et al. 2016; Wong et al. 2017). Yet, so far, no linkage map has been published. Genetic maps are the cornerstone upon which to start trait dissection, marker assisted selection (MAS), and identification of genomic regions that underlie traits of interest (Yu et al. 2004; Leonforte et al. 2013; Cai et al. 2016). Linkage maps could even lead to the discovery of syntenic regions in comparative studies for transferring knowledge between species (Hecht 2005; McClean et al. 2010; Ho et al. 2017), to the mapping of QTLs on orthologous chromosomes (Tian et al. 2010), and to the investigation of reproductive barriers within and between species/subspecies (Liu et al. 2007; Reflinur et al. 2014; Chandnani et al. 2017). The construction of a linkage map from a winged bean population would therefore support investigation in traits related to crop morphology and yield, such as internode length, seed number per pod, or days to flowering, as recently shown also in crops like *Brassica napus* (Cai et al. 2016), *Brassica rapa* (Huang et al. 2017), *Glycine max* (Xu et al. 2013; Sonah et al. 2015), *Vicia faba* (Catt et al. 2017), and other legume crops (Weller and Ortega 2015).

Thanks to decreasing costs in high-throughput sequencing, and the ability to discover large number of molecular markers at once, genotyping-by-sequencing (GbS) has enabled analysis in otherwise unknown crops. Single Nucleotide Polymorphisms (SNPs) are particularly useful and based on this technology, with their low mutation rate, abundance, and distribution across the entire genome making them widely employed into high-resolution mapping (Collard et al. 2005; Shirasawa et al. 2010; Ren et al. 2015; Cheng et al. 2017; Saxena et al. 2017; Huang et al. 2017). In the present study we report the first linkage map in winged bean, constructed using an F_2 mapping population (Chapter 3), obtained from high-density SNP discovery using the DArTseq technology. The obtained linkage map could be employed for QTL analysis in winged bean, using the phenotypic traits previously recorded during the F_2 field assessment (Chapter 3).

2 Material and methods

2.1 Plant material

An F_2 mapping population of 221 individuals was obtained from the cross M3 × FP15: phenotypic analysis has been reported earlier for a subset of this population, along with the corresponding parents M3 (retrieved from Malaysian Agricultural Research and Development, MARDI) and FP15 (collected in Malaysia). Crosses were performed in a controlled environment facility, while F_1 and F_2 individuals were grown in field conditions and left free to self-pollinate at the Field Research Centre of Crops For the Future (CFF, Semenyih, Malaysia), in the years 2017 and 2018 respectively (see Chapter 3 for details on growing conditions in controlled environment and field experiment).

2.2 DNA extraction, and high-throughput genotyping

Leaves samples from F_2 individuals were collected six weeks after sowing, and dried completely with drying beads (Rhino Research Group, Phichit, Thailand). Genomic DNA was extracted using the TissueLyzer II and DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) according to manufacturer instruction, with only minor modifications, which included lysis steps in 1.5mL tubes. Extracted genomic DNA integrity and concentration were assessed by 1% agarose gel run in TAE buffer 80V (5µL of sample loaded) and NanoDrop 1000 (Thermo Scientific, NH, USA). Quality of DNA was checked by *Hin*dIII-HF restriction enzyme reaction with 100ng DNA in 1X CutSmart® buffer (New England BioLabs, Ipswich, MA) (37°C, 1h), followed by gel electrophoresis of the digested DNA. A total of 2µg of genomic DNA of each F_2 and parental individual were sent to Diversity Array Technology Pty. Ltd. (Canberra, Australia) for DArTseq genotyping. In brief, this latter consisted of a genomic complexity reduction by means of restriction enzymes (RE) combinations (a rare-cutting enzyme *PstI* and a more common-cutting enzyme *MseI*) that digest genomic DNA, followed by ligation of adapters, PCR amplification of fragments digested by both RE, and their sequencing on Illumina Hiseq 2500. Processing of generated sequences, and SNP calling were then carried out within the DArT analytical pipeline (Kilian et al. 2012).

2.3 Markers selection and genetic linkage map construction

Markers were filtered in order to select only for homozygous loci within each parent, and polymorphic between them. All loci with one or more missing values in either parental of F_2 genotypes were removed. Markers that

were included into JoinMap were also used to carry out genetic distance analysis, and a Principal Coordinate Analysis (PCoA) in GenAlex v6.5 (Peakall and Smouse 2012) to assess the F_2 population distance and distribution compared to parental genotypes.

Linkage mapping was carried out with JoinMap® v5 (Van Ooijen 2017). Markers were first divided into linkage groups (LG) obtained by testing at a logarithm of odds (LOD) value between 1 and 50 using the "grouping (tree)" function, and SNPs were tested against expected Mendelian segregation ratio (1:2:1) with the χ^2 analyses integrated in the "Locus genotype frequency" function in JoinMap. This latter offered the possibility to look for segregation distortion at a 0.1 threshold.

Initially, the maximum likelihood (ML) algorithm was used for mapping of LGs starting with more than 100 markers, in order to deal with the computational load. Then, the regression mapping algorithm was used with default parameters: recombination frequency $\leq$0.40, LOD value $\geq$1.0, goodness-of-fit jump threshold for removal of loci equal to 5.0, and rippling performed after every marker was added into the map. The Haldane's function was used to convert recombination frequency into map distances (cM), considering only the second round of mapping (i.e. without any 'forced' markers) for the optimum number of markers to include into the map. Markers were checked for nearest neighbour (NN) fit values and mean χ^2 contribution for the goodness-of-fit within each LG map.

2.4 Segregation distortion analysis

Markers included into the construction of the linkage map were analysed for their segregation patterns. Initially, the χ^2 test in JoinMap was used to check for zygotic segregation distortion (zygotic SD) among loci: i.e. a distortion from the expected Mendelian ratio 1:2:1 for *aa*, *ab*, and *bb* genotypes, with *a* representing the allele from the male parent M3, and *b* the one from the female parent FP15. Then, in order to investigate whether particular alleles (paternal or maternal) were favoured by segregation distortion among these markers, allelic frequencies were tested for deviation from the expected *a*:b = 1:1 Mendelian ratio. First, these were calculated at each locus according to the formulae:

$$p = \frac{(homozygous\ individuals\ (aa)\ \times 2) + heterozygous\ individuals\ (ab)}{n}$$

$$q = \frac{(homozygous\ individuals\ (bb)\ \times 2) + heterozygous\ individuals\ (ab)}{n}$$

Where *p* is the frequency of the allele *a* (from male parent M3), *q* is the frequency of the allele *b* (from female parent FP15), and *n* is the total number of F_2 genotypes assessed (i.e. 221). Once calculated, χ^2 test was carried for allelic SD, according to the formula from Li and colleagues (2011):

$$\chi^2_{(1)} = \frac{(2np - n)^2 + (2nq - n)^2}{n} \quad (\alpha = 0.05)$$

3 Results

3.1 DArTseq markers, and population analysis

From DArTseq data, 2467 loci were homozygous within each parental genotype, and polymorphic between them. From these, 530 were removed due to missing values in one or more F_2 individuals, leaving only markers with a call rate value of 1. The remaining 1937 SNPs markers were imported into JoinMap® 5 for the construction of the linkage map. Fifty-four percent of these corresponded to transitions (Ts) (517 A/G, 523 C/T, 1040 in total), while remaining 46% were transversions (Tv) (205 A/T, 233 A/C, 219 G/T, 240 G/C, 897 in total), for a Ts:Tv ratio of 1.16.

Frequency analysis revealed that heterozygosity in the F_2 was close to the expected value of 0.5 (0.491). Treating the parental and the F_2 individuals as distinct populations, a Nei's genetic identity matrix show different similarities among them: in particular, the F_2 was reported to be more similar to M3 (0.84) than to FP15 (0.58). This was in agreement with the pattern observed in the PCoA graph (see Figure a.1, in appendix) where the F_2 was distributed between parental values, but closer to the male M3 parent.

3.2 Analysis of segregation distortion

Out the 1937 SNPs imported for mapping, 144 were considered identical based on F_2 segregation patterns and therefore were excluded from the mapping process, leaving 1793 markers. Based upon these, the χ^2 test in JoinMap to test for zygotic segregation distortion, identified significant deviation from the expected Mendelian

ratio: out of the 1793 loci, 1314 (73%) showed zygotic segregation distortion at a significance level of 0.1, and almost 70% of them at a significance level of .0005 (Table 3.1). Out of the 1314 loci showing zygotic SD, 1246 (95%) also showed allelic SD. Among these, looking at which allele had the higher frequency revealed that in 99% of the cases the distortion was towards the alleles coming from the male parent M3 (allele *a*) (Table 3.1).

Table 3.1: Total SNPs and SNPs showing significant segregation distortion (SD). Significance was calculated with chi-square (χ^2) test at significance threshold 0.1.

SD			Total Loci	1793	-
			Total Zygotic SD	1314	73%
	Zygotic SD	SD significance ($p<$)	at 0.1	59	4.5%
			between 0.05 and 0.0001	1255	95.5%
		χ^2 test for ratios	Distortion from [*aa*:(*ab*+*bb*)]=1:3	1154	
			of which, distortion from *aa*:*ab*=1:2	1122	
			of which *aa* > (*ab*/2)	1114	
	Allelic SD		Total Zygotic SD	1314	
			Zygotic AND Allelic SD ($p< 0.05$)	1246	95%
			Towards *a* (male) allele	1241	99%
			Towards *b* (female) allele	5	1%

SD = segregation distortion

What the "χ^2 test for ratios" section indicates, is that out of the zygotic SD loci 1154 showed significant deviation from the ratio 1:3 between number of *aa* individuals and *ab*+*bb* individuals. Looking into these, another test revealed that in most of the cases (1122, 97%) there was a deviation between *aa* and *ab* expected 1:2 ratio. Among these, 1114 showed that the number of *aa* individuals was more than the expected (i.e. half the number of *ab* individuals). Altogether, these would confirm that the segregation distortion was in favour of the alleles from the male parent M3, and in particular this seems supported by an increase in homozygous loci *aa* over *bb* and *ab* genotypes.

A false-discovery rate (FDR) analysis was carried out, using a Benjamini-Hochberg procedure with a false discovery rate set at 5%, in order to correct for false positive segregation distortion markers that could arise from multiple χ^2 tests in JoinMap. The percentage of zygotic SD loci dropped from 73 to 54% (969 rather than 1314, out of the total 1793 loci included), with 99% of them showing again a higher frequency of the M3-inherited allele *a* over the allele *b*.

Zygotic SD markers showed no particular preference for transition or transversion, keeping a Ts:Tv ratio at 1.23, compared to the 1.16 across all SNP between the two parental genotypes.

3.3 Genetic linkage maps

From the 1793 unique SNPs, 9 LGs with an average of 199 SNPs each were obtained through the "grouping (tree)" function in JoinMap, at a LOD value of 30. The percentages of the zygotic SD (at a conservative significance level of 0.1) calculated in JoinMap ranged between 57.3 and 90.3%, for an average of 75% across groups (see Table 3.2).

Table 3.2: Linkage Groups (LGs), and SNP included, generated from "grouping tree" function in JoinMap using LOD value of 30. SD: segregation distortion.

	LG1	**LG2**	**LG3**	**LG4**	**LG5**	**LG6**	**LG7**	**LG8**	**LG9**
SNPs	328	247	234	222	218	171	165	140	67
% of SD SNPs	71.6	68.8	83.8	72.5	57.3	76.0	90.3	73.6	67.2
Non-SD SNPs	93	77	38	61	93	41	16	37	22

In order to proceed, all distorted SNPs were removed before the beginning of maximum likelihood and regression mapping. Those showing negative distance or NN fit values ≥ 4 were removed after each round of mapping, without removing consecutive markers at once. Eventually, the nine LGs had a range between 151.7 and 14.8cM, for a total length of 747.1cM, with 253 SNPs in total giving about 3cM/SNP (see table 3.2). Once obtained, this first map represented the framework into which SD markers could be included, while maintaining the order of the non-distorted markers.

For each LG, distorted markers were re-introduced one or two at a time, followed by regression mapping at each round. Markers with all the different significant levels of deviation from the expected zygotic ratio (1:2:1) were included (i.e. from those with $p < .1$, to $p < .0001$).

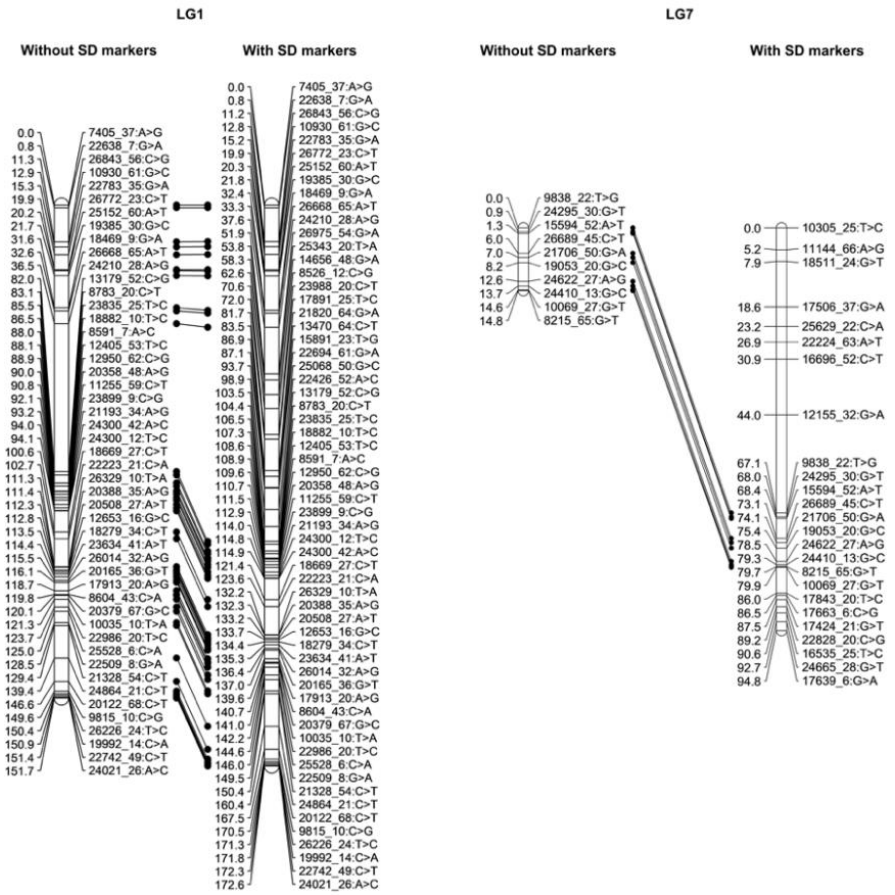


Figure 3.1: LG1 (left) and LG7 (right) linkage maps obtained without and with segregation distortion (SD) markers. Connecting lines show the shifting position for each marker from the first map to the second. Markers without these lines are the SD ones, i.e. not present during the initial mapping.

Addition of SD markers was followed by comparisons between initial maps (without SD markers) and newly obtained maps (with SD markers). This was carried out in order to check whether the introduced distorted markers were altering the order of non-SD markers from the framework LG maps (see as example, Figure 3.1). Eventually, this second mapping resulted in a map with a total length of 1171.6cM, with 395 SNPs (3cM/SNP) (see full map after addition of SD markers in Figure 3.2).

A summary with details of the two maps, without and with SD markers, is reported in the Table 3.3 below.

Table 3.3: Summary of the two maps, obtained without and with segregation distortion (SD) markers, with LG derived at LOD 30.

	Without - with SD markers									
	length (cM)		**SNPs**		**SD SNPs (%)**		**cM/SNP**		**unmapped SNPs (%)**	
LG1	151.7	172.6	49	61	-	12 (20)	3.1	2.8	279 (85)	267 (81)
LG2	111.3	176.8	34	66	-	32 (49)	3.3	2.7	213 (86)	181 (73)
LG3	126.6	160.5	25	42	-	17 (40)	5.1	3.8	209 (89)	192 (82)
LG4	84.0	102.5	35	49	-	14 (29)	2.4	2.1	187 (84)	173 (78)
LG5	104.0	171.0	37	54	-	10 (19)	2.8	3.2	181 (83)	164 (75)
LG6	48.8	89.9	28	42	-	14 (33)	1.7	2.1	143 (84)	129 (75)
LG7	14.8	94.8	10	25	-	15 (60)	1.5	3.8	156 (94)	141 (85)
LG8	42.8	94.1	20	33	-	13 (39)	2.1	2.9	120 (86)	107 (76)
LG9	63.1	109.4	15	23	-	8 (35)	4.2	4.8	52 (78)	44 (66)
Average	83.0	130.2	28.1	43.9	-	13.0 (36)	2.9	3.1	-	-
Total	747.1	1171.6	253	395	-	117	-	-	1540 (86)	1398 (78)

In LG5, additional non-SD markers fitted using the regression algorithm, when distorted markers were included. All included SD markers were the most significant ($p < .0001$) deviated from the expected Mendelian ratio. Although SNPs that deviated less-significantly from the expected ratio were brought back into the mapping process, these remained unmapped.

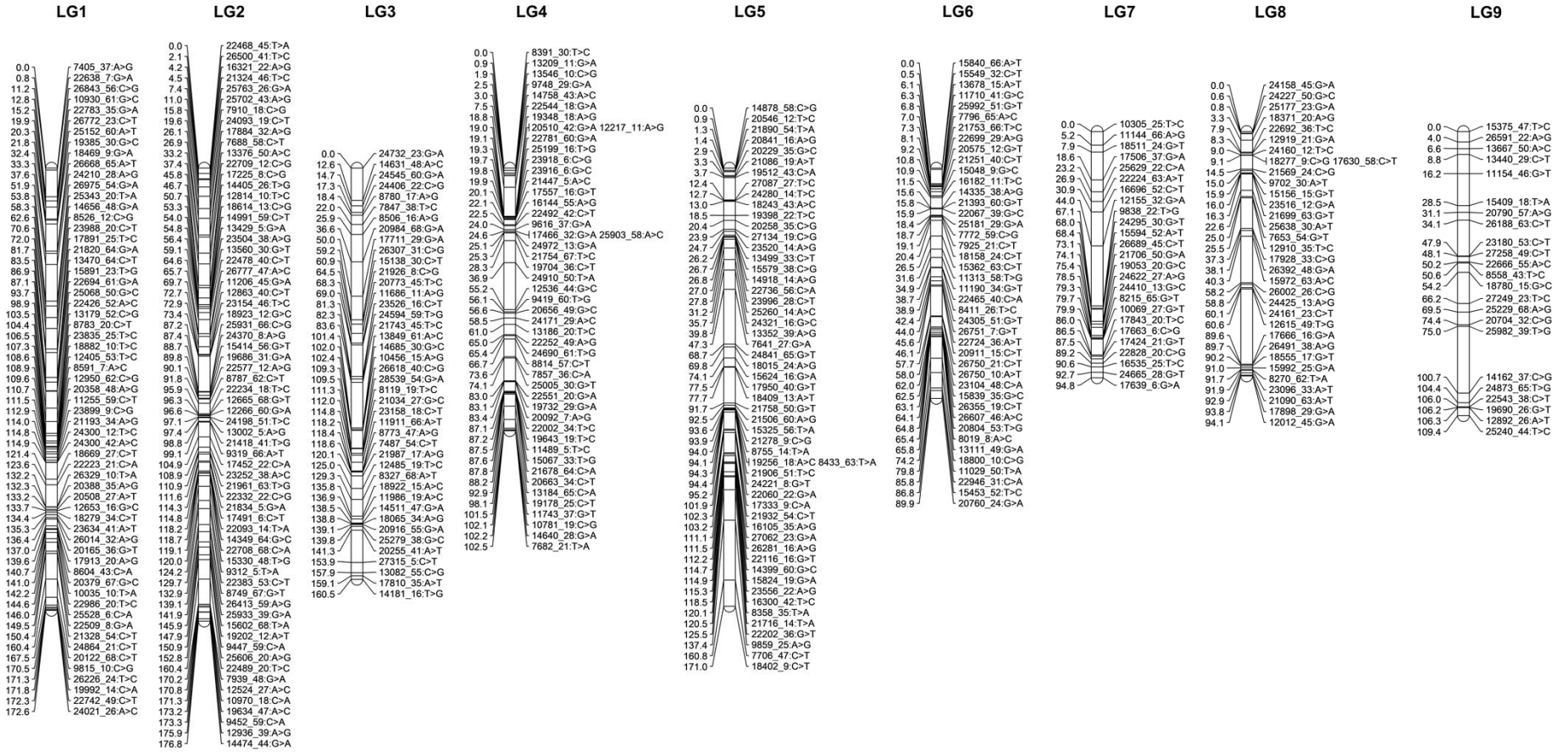


Figure 3.2: SNP-based genetic linkage map for winged bean, from M3 × FP15. Map distances are shown on the left side, while SNP names are on the right side of each LG. Identically positioned loci, if present, are reported next to each other.

4 Discussion

Generating genetic linkage maps could support trait dissection, marker assisted selection, knowledge transfer from well-studied species, and analysis of reproductive barriers. Here we reported the development of a high number of SNP markers between the parents M3 and FP15 along with genotyping of the derived mapping population, resulting into the first linkage map in winged bean. DArTseq was able to reveal 2467 SNPs between the two parents, 1937 of which were used for generating the linkage map. These latter had a Ts:Tv ratio of 1.16, lower than earlier studies in winged bean. Vatanparast and colleagues (2016) reported a ratio of 2.22 in their SNP discovery, based upon their combined transcriptome data from Sri Lankan accessions, and the transcriptome data from Chapman (2015) using Ibadan Local-1 from Nigeria (possibly TPt1, but retrieved from U.S. Department of Agriculture). Transversions are rarer than transitions given the required chemical change in the ring structure of the nitrogenous base, and have higher chances to generate non-synonymous polymorphisms in amino acid sequences, or to alter the function of transcription factor DNA-binding sequences (Yadav et al. 2015; Guo et al. 2017). Selective pressure is therefore likely to act more on them than on transformations, especially when they occur in transcribed regions. Hence, it is likely that the higher Ts:Tv ratio reported by Vatanparast and colleagues might be due to their analysis using transcript data, while the ratio reported here could have been lowered by looking also at the genomic regions where transversions have no effect on the phenotype. However, considering studies with mapping population and markers obtained across the genome, rather than focusing on transcriptomic data, the value reported here is also lower than has been reported between related species of oil palm (*Elaeis guineensis* and *Elaeis oleifera*, ~1.5) (Ting et al. 2014), chickpea (*Cicer arietinum* and *Cicer reticulatum*, 1.7) (Gaur et al. 2012), and between cultivar/lines in pea (*Pisum sativum*, 2.1) (Leonforte et al. 2013), barley (*Hordeum vulgare*, 2.4), sorghum (*Sorghum bicolor*, 1.9) (Ji et al. 2017), and soybean (*G. max*, ~2) (Yadav et al. 2015). Considering the probable lower mutation rate at which transversions should occur, the ratio reported here might suggest that parental genomes are derived from genetic pools that have been diverged significantly, perhaps due to selective pressure.

Applying population analysis to the progeny and parental genotypes could reveal the overall genetic behaviour in our population. From the identity matrix and PCoA analysis it is possible to observe a deviation of the

genetic similarity of the F_2 towards the M3 parent, although the F_1 individual remained hybrid between the two parents. This was a first indication of the observed amount of segregation distortion later revealed during map construction.

Segregation distortion is a phenomenon in common between plant and animal eukaryotic cells, whose mechanisms are yet to be fully elucidated. At a biological level, the segregation distortion might be caused by factors affecting gametogenesis, or gamete and/or zygote fitness (Falconer and Mackay 1989; Pardo-Manuel de Villena and Sapienza 2001; Fishman and Willis 2004; Bravo Núñez et al. 2018). For example, a class of factors called "meiotic drivers" are able to promote their segregation during gametes formation, eventually increasing their copy number among them, or are able to reduce the viability of other gametes lacking specific alleles (Bravo Núñez et al. 2018). Such mechanisms driven by so-called "selfish" sequences might be implicated in shaping genome sizes, structure, and diversity (Sandler and Novitski 1957; Buckler et al. 1999; Burt and Trivers 2006). Another factor, which could lead to SD in progeny derived from controlled crosses, is genetic divergence between parental genotypes, as shown in particular within interspecific crosses in rice (Pham and Bougerol 1993; Xu et al. 1997; Reflinur et al. 2014), soybean (Liu et al. 2007), maize (Wang et al. 2012), and other species (Fishman and Willis 2004; Chandnani et al. 2017). In these crosses, crossing-over might be affected, for example, by inverted or translocated regions, leaving a portion of gametes without a complete genetic set, and eventually sterile (Liu et al. 2007). Regarding zygotic fitness, abnormal segregation might be ascribed to, for example, zygotic abortion due to allelic dosage insufficiencies, or in general to unbalanced parental contribution in the endosperm development and function (Pham and Bougerol 1993; Reflinur et al. 2014; Lafon-Placette and Köhler 2016). All these interlinked mechanisms are responsible for setting hybridization barriers, and for leading species evolution.

At a genomic level, if a particular locus is responsible for causing SD, namely SDL, then surrounding linked loci will deviate from the expected ratio as well. This will result in distorted markers grouped in clusters as a segregation distortion region, namely SDR, the extent of which depends on the number of SDL and the number of loci linked to each them. SDRs eventually show alleles with a frequency above the expected Mendelian value, as shown in both animal (Pardo-Manuel de Villena and Sapienza 2001; Larracuente and Presgraves

2012) and plant linkage maps (Xu et al. 1997; Liu et al. 2007; Li et al. 2010, 2011, 2015; Zhou et al. 2014; Ji et al. 2017).

In the present study, SD markers were found across all the obtained LGs, rather than confined to a few, and the second mapping step (i.e. re-introduction of SD markers) showed these to be grouped in clusters within each LG, probably corresponding to SDRs. High segregation distortion was observed also in a map built upon RILs, from a peanut (*A. hypogaea*) intraspecific biparental cross, with about 40% of markers showing SD ($p <$ .05) (Zhou et al. 2014). Similar levels of SD were also found in alfalfa (*M. sativa*) using SSRs (Li et al. 2011), rice (*O. sativa*, subsp. *japonica* and *indica*) interspecific crosses (Reflinur et al. 2014), and watermelon (*C. lanatus*) (Ren et al. 2015).

The winged bean progeny from the controlled cross reported here showed *seeds per pod* to be consistently lower than the lowest parent (M3), in both F_1 and F_2. This was ascribed to undeveloped seeds observed within pods, while the length of the pod structure remained intermediate between parental values. Such an observation might suggest zygotic abortion due to allelic dosage alterations during endosperm development, perhaps shortly after seed maturing switches from maternal to zygotic control (Weber et al. 2005; Baroux and Grossniklaus 2015). Hence, selection at the zygotic level could be responsible for the observed SD, due perhaps to heterospecific genomic interaction between parental genomes, or incomplete genetic material within fused gametes. The analysis of allelic and zygotic SD confirmed that most of the distorted alleles were towards the male parent M3, recalling the pattern observed in the PCoA, with a high number of homozygous *aa*. This could indicate the lack of cytological effects, related to maternal control during seed development, given the male bias and suggests gametophytic factors. On the other hand, high rates of aborted seeds could imply the occurrence of post-fertilization mechanisms, perhaps zygotic selection. In a more complicated scenario, both the types of selection might be in place, but only further studies could elucidate the underlying causes. For example, reciprocal backcrosses in both paternal and maternal direction could be used investigate which is the source of SD and the whether there are organelle effects (i.e. which parent is responsible for the SD, and whether this is transmitted by male and/or female gametes). Another option might be to compare the current M3 × FP15 population with another F_2 derived from a cross between the male parent M3 and another

winged bean accession used as female, with progeny showing a *seeds per pod* trait not significantly lower than the lowest parent. If in that case SD shows the same levels as in the first population, it would suggest that it might be due to nuclear mechanisms and that M3 could be the possible source of SD. The comparison could also support the identification of conserved SDRs, and therefore the identification of candidate SDLs causing distortion (Lu et al. 2002). On the other hand, in case of low or no SD, FP15 could be appointed as the source of SD in the current cross.

Finally, at a linkage mapping level, SD could cause incorrect calculations of recombination frequencies and ordering of markers along the map. In the present case, it could be possible that segregation distortion towards one parental allele could lead to reduced frequencies and spurious linkages (i.e. the loci show less recombination because of the recombinant genotypes have been negatively selected, not because recombination did not occur) as previously shown in *B. napus* (Cloutier et al. 1997). The approach of initially excluding SD markers to obtain an ordered framework map should have allowed us to avoid as much as possible errors in map order, although to date there is no standard way to carry this out (Xian-Liang et al. 2006; Li et al. 2010).

5 Conclusions

The present study has reported a first linkage map for winged bean, from a biparental F_2 population. The amount of segregation distortion observed suggests the presence of some reproductive barriers influencing equal allelic inheritance. A two-step approach has been adopted to map construction and should have allowed us to obtain the 'correct' marker order and include some of the SD markers, in order to carry out initial QTL analysis for traits of interest recorded in the assessed F_2 (see Chapter 3).

References

Baroux C, Grossniklaus U (2015) The Maternal-to-Zygotic Transition in Flowering Plants. Evidence, Mechanisms, and Plasticity. Curr Top Dev Biol 113:351–371. doi: 10.1016/bs.ctdb.2015.06.005

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc London Ser B Biol Sci 274:227–274

Bravo Núñez MA, Nuckolls NL, Zanders SE (2018) Genetic Villains: Killer Meiotic Drivers. Trends Genet 34:. doi: 10.1016/j.tig.2018.02.003

Buckler ES, Phelps-Durr TL, Buckler CS, et al (1999) Meiotic drive of chromosomal knobs reshaped the maize genome. Genetics 153:415–26

Burt A, Trivers R (2006) Genes in conflict - The biology of selfish genetic elements, 1st edn. The Belknap Press of

Harvard University Press, Cambridge, Massachusetts (US); London (UK)
Cai G, Yang QQ, Chen H, et al (2016) Genetic dissection of plant architecture and yield-related traits in Brassica napus. Sci Rep 6:21625. doi: 10.1038/srep21625
Catt SC, Braich S, Kaur S, Paull JG (2017) QTL detection for flowering time in faba bean and the responses to ambient temperature and photoperiod. Euphytica 213:125. doi: 10.1007/s10681-017-1910-8
Chandnani R, Wang B, Draye X, et al (2017) Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species. Theor Appl Genet 130:2219–2230. doi: 10.1007/s00122-017-2952-y
Chapman MA (2015) Transcriptome Sequencing and Marker Development for Four Underutilized Legumes. Appl Plant Sci 3:1400111. doi: 10.3732/apps.1400111
Chen D, Yi X, Yang H, et al (2015) Genetic diversity evaluation of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) using inter-simple sequence repeat (ISSR). Genet Resour Crop Evol 62:823–828. doi: 10.1007/s10722-015-0261-3
Cheng A, Chai HH, Ho WK, et al (2017) Molecular Marker Technology for Genetic Improvement of Underutilised Crops. In: Abdullah SNA, Ho CL, Wagstaff C (eds) Crop Improvement - Sustainability Through Leading-Edge Technology. Springer International Publishing, pp 47–70
Cloutier S, Cappadocia M, Landry BS (1997) Analysis of RFLP mapping inaccuracy in Brassica napus L. TAG Theor Appl Genet 95:83–91. doi: 10.1007/s001220050535
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. doi: 10.1007/s10681-005-1681-5
Eagleton G (1999) Winged Bean in Myanmar, Revisted. Econ Bot 53:342–352
Falconer DS, Mackay TF (1989) Introduction to Quantitative Genetics Third Edition
Fishman L, Willis JH (2004) A Novel Meiotic Drive Locus Almost Completely Distorts Segregation in Mimulus (Monkeyflower) Hybrids. Genetics 169:347–353. doi: 10.1534/genetics.104.032789
Gaur R, Azam S, Jeena G, et al (2012) High-Throughput SNP Discovery and Genotyping for Constructing a Saturated Linkage Map of Chickpea (Cicer arietinum L.). DNA Res 19:357–373. doi: 10.1093/dnares/dss018
Guo C, McDowell IC, Nodzenski M, et al (2017) Transversions have larger regulatory effects than transitions. BMC Genomics 18:394. doi: 10.1186/s12864-017-3785-4
Harder DK (1992) Chromosome Counts in *Psophocarpus*. Kew Bull 47:529. doi: 10.2307/4110581
Hecht V (2005) Conservation of Arabidopsis Flowering Genes in Model Legumes. Plant Physiol 137:1420–1434. doi: 10.1104/pp.104.057018
Ho WK, Chai HH, Kendabie P, et al (2017) Integrating genetic maps in bambara groundnut [*Vigna subterranea* (L) Verdc.] and their syntenic relationships among closely related legumes. BMC Genomics 18:1–9. doi: 10.1186/s12864-016-3393-8
Huang L, Yang Y, Zhang F, Cao J (2017) A genome-wide SNP-based genetic map and QTL mapping for agronomic traits in Chinese cabbage. Sci Rep 7:46305. doi: 10.1038/srep46305
Ji G, Zhang Q, Du R, et al (2017) Construction of a high-density genetic map using specific-locus amplified fragments in sorghum. BMC Genomics 18:51. doi: 10.1186/s12864-016-3430-7
Kilian A, Wenzl P, Huttner E, et al (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. In: Pompanon F, Bonin A (eds) Data Production and Analysis in Population Genomics. Humana Press, Totowa, NJ, pp 67–89
Koshy EP, Alex BK, John P (2013) Clonal fidelity studies on regenerants of *Psophocarpus tetragonolobus* (L.) DC. using RAPD markers. The Bioscan 8:763–766
Lafon-Placette C, Köhler C (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. Mol Ecol 25:2620–2629. doi: 10.1111/mec.13552
Larracuente AM, Presgraves DC (2012) The selfish Segregation Distorter gene complex of Drosophila melanogaster. Genetics
Leonforte A, Sudheesh S, Cogan NO, 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 Biol 13:161. doi: 10.1186/1471-2229-13-161
Lepcha P, Egan AN, Doyle JJ, Sathyanarayana N (2017) A Review on Current Status and Future Prospects of Winged Bean (*Psophocarpus tetragonolobus*) in Tropical Agriculture. Plant Foods Hum Nutr 72:225–235. doi: 10.1007/s11130-017-0627-0
Li C, Bai G, Chao S, Wang Z (2015) A High-Density SNP and SSR Consensus Map Reveals Segregation Distortion Regions in Wheat. Biomed Res Int 2015:1–10. doi: 10.1155/2015/830618
Li H, Kilian A, Zhou M, et al (2010) Construction of a high-density composite map and comparative mapping of segregation distortion regions in barley. Mol Genet Genomics 284:319–331. doi: 10.1007/s00438-010-0570-3
Li X, Wang X, Wei Y, Brummer EC (2011) Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. Theor Appl Genet 123:667–679. doi: 10.1007/s00122-011-1617-5

Liu B, Fujita T, Yan ZH, et al (2007) QTL mapping of domestication-related traits in soybean (*Glycine max*). Ann Bot 100:1027–1038. doi: 10.1093/aob/mcm149

Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. TAG Theor Appl Genet 105:622–628. doi: 10.1007/s00122-002-0970-9

McClean PE, Mamidi S, McConnell M, et al (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC Genomics 11:184. doi: 10.1186/1471-2164-11-184

Mohanty CS, Verma S, Singh V, et al (2013) Characterization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) based on molecular , chemical and physiological parameters. Am J Mol Biol 3:187–197. doi: 10.4236/ajmb.2013.34025

NAS (1981) The Winged Bean: A High-Protein Crop for the Humid Tropics, 2nd edn. National Academy of Sciences, Washington, D.C.

Pardo-Manuel de Villena F, Sapienza C (2001) Nonrandom segregation during meiosis: the unfairness of females. Mamm Genome 12:331–339. doi: 10.1007/s003350040003

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28:2537–9. doi: 10.1093/bioinformatics/bts460

Pham J-L, Bougerol B (1993) Abnormal segregations in crosses between two cultivated rice species. Heredity (Edinb) 70:466–471. doi: 10.1038/hdy.1993.68

Reflinur, Kim B, Jang SM, et al (2014) Analysis of segregation distortion and its relationship to hybrid barriers in rice. Rice 7:3. doi: 10.1186/s12284-014-0003-8

Ren R, Ray R, Li P, et al (2015) Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. Mol Genet Genomics 290:1457–1470. doi: 10.1007/s00438-015-0997-7

Sandler L, Novitski E (1957) Meiotic Drive as an Evolutionary Force. Am Nat 91:105–110. doi: 10.1086/281969

Saxena RK, Singh VK, Kale SM, et al (2017) Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea. Sci Rep 7:1–11. doi: 10.1038/s41598-017-01537-2

Shirasawa K, Isobe S, Hirakawa H, et al (2010) SNP discovery and linkage map construction in cultivated tomato. DNA Res 17:381–391. doi: 10.1093/dnares/dsq024

Sonah H, O'Donoughue L, Cober E, et al (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol J 13:211–221. doi: 10.1111/pbi.12249

Stephenson RA, Kesavan V, Claydon A, et al (1979) Studies on tuber production in winged bean (*Psophocarpus tetragonolobus* (L.) DC.). Proceedings of the 5th International Symposium on Tropical Root and Tuber Crops. Laguna, Philippines

Tian Z, Wang X, Lee R, et al (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci 107:8563–8568. doi: 10.1073/pnas.1000088107

Ting NC, Jansen J, Mayes S, et al (2014) High density SNP and SSR-based genetic maps of two independent oil palm hybrids. BMC Genomics 15:1–11. doi: 10.1186/1471-2164-15-309

Van Ooijen JW (2017) JoinMap ® 5, Software for the calculation of genetic linkage maps in experimental populations of diploid species

Vatanparast M, Shetty P, Chopra R, et al (2016) Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Sci Rep 6:1–14. doi: 10.1038/srep29070

Wang G, He QQ, Xu ZK, Song RT (2012) High segregation distortion in maize B73 x teosinte crosses. Genet Mol Res 11:693–706. doi: 10.4238/2012.March.19.3

Weber H, Borisjuk L, Wobus U (2005) Molecular Physiology of Legume Seed Development. Annu Rev Plant Biol 56:253–279. doi: 10.1146/annurev.arplant.56.032604.144201

Weller JL, Ortega R (2015) Genetic control of flowering time in legumes. Front Plant Sci 6:207. doi: 10.3389/fpls.2015.00207

Wong Q, Tanzi AS, Ho W, et al (2017) Development of Gene-Based SSR Markers in Winged Bean (Psophocarpus tetragonolobus (L.) DC.) for Diversity Assessment. Genes (Basel) 8:100. doi: 10.3390/genes8030100

Xian-Liang S, Xue-Zhen S, Tian-Zhen Z (2006) Segregation distortion and its effect on genetic mapping in plants. Chinese J Agric Biotechnol 3:163–169. doi: 10.1079/CJB2006110

Xu M, Xu Z, Liu B, et al (2013) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol 13:91. doi: 10.1186/1471-2229-13-91

Xu Y, Zhu L, Xiao J, et al (1997) Chromosomal regions associated with segregation distortion of molecular markers in F 2 , backcross, doubled haploid, and recombinant inbred populations in rice (Oryza sativa L.). Mol Gen Genet MGG 253:535–545. doi: 10.1007/s004380050355

Yadav CB, Bhareti P, Muthamilarasan M, et al (2015) Genome-Wide SNP Identification and Characterization in Two Soybean Cultivars with Contrasting Mungbean Yellow Mosaic India Virus Disease Resistance Traits. PLoS One 10:e0123897. doi: 10.1371/journal.pone.0123897

Yu K, Park SJ, Zhang B, et al (2004) An SSR marker in the nitrate reductase gene of common bean is tightly linked to a

major gene conferring resistance to common bacterial blight. Euphytica 138:89–95. doi: 10.1023/B:EUPH.0000047077.75285.60

Zhou X, Xia Y, Ren X, et al (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15:1–14. doi: 10.1186/1471-2164-15-351

Appendix – Chapter 4

a.1 PCoA of parental, F_1, and F_2 individuals using SNPs

Selected SNP markers were used to check the parental, F_1, and F_2 populations through Principal Coordinate Analysis (PCoA). The graph showed in Figure a.1

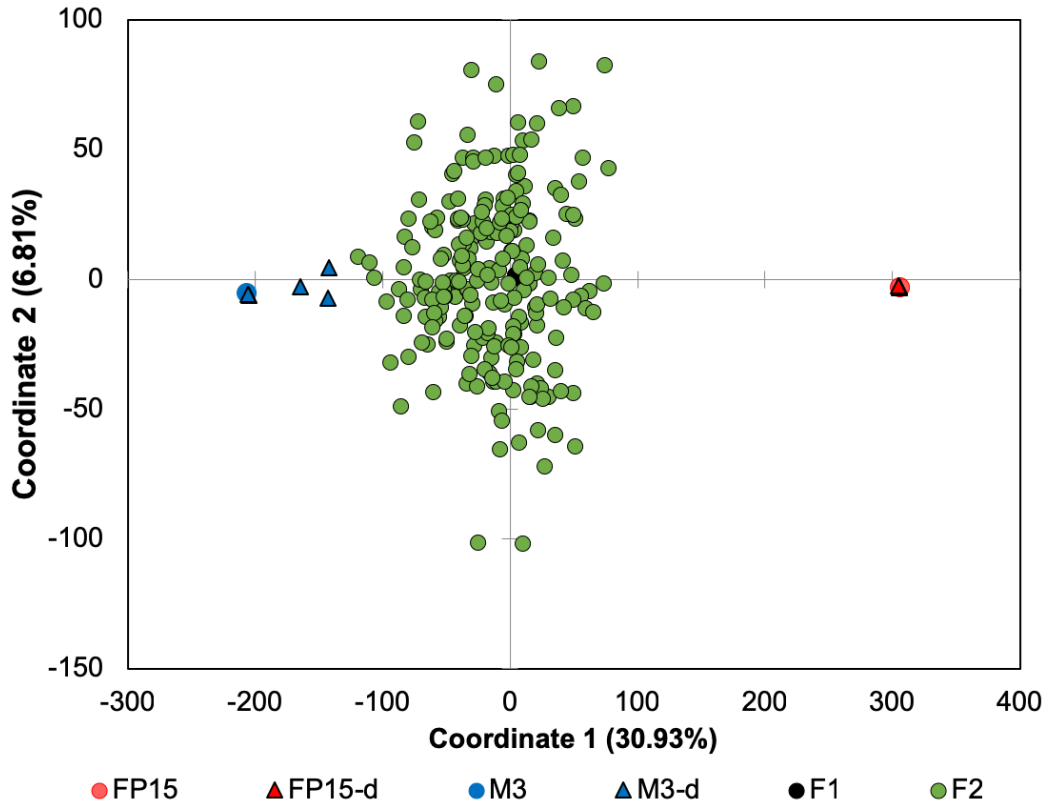


Figure a.1: PCoA with selected SNP markers from DArTseq used for mapping. M3, M3-derived (M3-d), FP15, FP15-derived (FP15-d), F_1, and F_2 individuals are mapped using 2 coordinates with a cumulative explained variance of 37.74%.

Chapter 5 - QTL mapping of plant architecture, yield, development, and physiological related traits in winged bean (*Psophocarpus tetragonolobus* (L.) DC.)

Abstract

Obtaining information about the number and location of genomic regions underlying traits of agronomical interest is crucial for breeding improvement programmes. The current study was conducted to identify QTLs linked to some of these traits in winged bean. From phenotypic data recorded in a biparental F_2 mapping population, QTL analysis was carried out on two genetic linkage maps. The first one included clustered segregation distortion markers, while the second map excluded such regions. According to Interval Mapping (IM) results, 5 segregating QTLs and 8 putative QTLs were found distributed across 9 traits. These included the discovery of markers linked to genomic regions that underlie *branch number per plant*, *length of branch*, *dry pod length*, and *chlorophyll content*. Such QTLs were in agreement with the subsequent Multiple-QTL Model (MQM) analysis. The information generated here provide a first insight into the genetic architecture of morphological, yield, developmental, and physiological traits in winged bean. This could support the implementation of marker-assisted-breeding, and isolation of candidate genes for traits of interest in winged bean improvement programmes.

1 Introduction

Winged bean (*Psophocarpus tetragonolobus* (L.) DC.; 2*n*=2x=18) is a leguminous underutilized species grown in hot and humid countries in Asia and Pacific areas, usually on a small-scale, as a vegetable crop and pulse. Edible parts such as green pods, tubers, and mature seeds have desirable characteristics (Kortt and Caldwell 1984; Kadam et al. 1984; Prakash et al. 1987), and the ability to fix nitrogen with a wide range of *Rhizobium* (Harding et al. 1978; Hildebrand et al. 1981; Iruthayathas et al. 1985) would encourage a wider use of winged bean in agriculture, for dietary diversification and more sustainable agriculture. However, the uncontrolled branching habit, indeterminate growth, and unstable productivity make large-scale utilization (and mechanisation) of this crop difficult. Earlier research tried to tackle this problems, for example investigating traits of agronomic relevance, like number of pods per plant (Arumugan and Perera 1989) or effects of vertical support on growth and branching (Schiavinato and Válio 1996), but this was before

molecular genetic analysis. From a recent phenotypic analysis on a F_2 population (Chapter 3), it appears that a trade-off could be achieved between reduced vegetative growth and increased pods per plant, for example by controlling the number of branches per plant. In particular, individuals with fewer but longer branches could maintain high pod productivity, while potentially reducing the vegetative biomass (i.e. leaves per plant). In order to obtain such individuals, it is necessary to dissect traits related to plant architecture, yield, and development, along with the isolation of markers able to select for desired phenotypes.

Such approaches have been part of the processes aimed to transform wild species into improved crops, leading to alterations in a common set of domestication-related traits, or "domestication syndrome" (Koinange et al. 1996; Weeden 2007; Doust 2007; Teichmann and Muhr 2015). Although this process dates long back in human history, only in the past few decades it has been possible to investigate the genetic basis of these traits, aided by advances in DNA markers development, genetic linkage mapping and Qualitative Trait Loci (QTL) analysis. Such molecular approaches have led to the dissection, for example, of *plant height*, and *internode length* in soybean (*Glycine max*) (Liu et al. 2007); or *number of branches per plant*, *pod length*, *pods per plant*, and *seeds per pods* in rapeseed (*Brassica napus*) (Cai et al. 2016) and faba bean (*Vicia Faba*) (Ávila et al. 2017). Such studies identified genomic regions controlling these traits, and provided molecular tool that could assist selection of individuals with desired characteristics. Similar investigations could greatly benefit crops that have suffered lack of research efforts in this direction (Varshney et al. 2009; Mayes et al. 2012), such as winged bean.

In this study we have carried out QTL analysis using data from an earlier phenotyped F_2 population, and the SNP-based genetic linkage map constructed upon it (see Chapters 3 and 4). The results provide novel information about the genetics of plant architecture, yield, and development related traits in winged bean, some of which have been proposed as breeding targets for the improvement of this tropical crop (see Chapter 1, literature review).

2 Material and Methods

2.1 Genetic linkage map

Genetic linkage mapping was carried out using an F_2 winged bean mapping population of 221 genotyped individuals, derived from the M3 × FP15 cross (see Chapter 4). Two maps were eventually constructed, the

first defined as a "framework" map without any segregation distorted loci, while the second "complete" map was obtained by adding distorted SNPs back into the framework map (for a full description of the mapping process, see Chapter 4). Eventually, the framework map had a length of 747.1cM with 253 SNPs included (about 2.9cM inter-marker distance), while the complete was map was 1171.6cM in length, with 395 SNPs (about 3.1cM intermarker distance, average 36% SD across the 9 LGs). First the complete, then the framework map have been used for QTL analysis.

2.2 Phenotypic data

Phenotypic data was included from the subset of 87 F_2 individuals, phenotyped for traits related to morphology, development, and yield (see Chapter 3). Included traits and their details are reported in Table 2.1.

Table 2.1: Phenotypic traits included in QTL analysis, and their phenotypic distribution in the F_2. For more details about the traits, see description below the table. ND: normal phenotypic distribution; Non-ND: non-normal phenotypic distribution.

	Trait	Abbreviation	F_2 distribution
Morphological	Internode length	InL	ND
	Number of nodes	NoN	Non-ND
	Stem length	StL	Non-ND
	Branch number per plant	NoB	Non-ND
	Total branch number per plant	NoBTOT	Non-ND
	Length of branch	LoB	ND
Yield	Seeds per pod	SPP	N
	Pods per plant	PPP	Non-ND
	Dry pod length	PoL	ND
	Hundred-seed weight	HSW	ND
Developmental	Days to first open flower	DtF	Non-ND
	Pod maturing time	PMT	Non-ND
Physiological	Chlorophyll concentration	SPAD	Non-ND

The *chlorophyll concentration* was measure with a SPAD 502DL (Spectrum Technologies Inc., IL, US): the average was obtained from the middle leaflets of the top 3 fully opened leaves, each measured 3 times, at 28 days after emergence. This was carried out following the observation of yellower leaves in the top portion of the stem in the parent M3 (across all growing cycles), and in the F_2. Visual scoring was linked to different values of chlorophyll concentration. All individuals reverted to complete green leaves between 40 to 50 days after emergence, before flowering onset (see Chapter 3, appendix). Traits were measured as reported earlier (Chapter 3, material and methods) except NoBTOT trait which represents the number of branches along the entire stem, rather than only within the first 10 nodes (i.e. NoB) (see Chapter 3, appendix).

2.3 QTL analysis

Two QTL analyses were carried out with MAPQTL 6 (van Ooijen, 2009) using the same phenotypic data with the two different density linkage maps, but using a common approach. For each trait, a permutation test (PT) was performed with 10000 permutations in order to obtain the genome-wide (GW) significance threshold ($\alpha = 0.05$) for the logarithm of odds (LOD) to use in downstream analysis (Churchill and Doerge 1994; Van Ooijen 2009). The non-parametric Kruskal-Wallis (KW) test was then performed to identify significant association between single markers and traits individually, leaving map groups and order information out. From here, only markers that reported a test value (K*) above 10 and significance lower than 0.01 were considered further. Then, Interval Mapping (IM) analyses noted loci up to 1 LOD below the GW threshold as putative QTL candidates, while those equal or above GW LOD were considered as detected segregating QTLs. For Multiple-QTL Model (MQM) analysis, potential cofactors were chosen based on both KW and IM results: these and adjacent loci (at a maximum of 5cM distance in both directions) were selected for Automatic Cofactor Selection (ACS) analysis. Retained cofactors were then included sequentially, rather than all at once, into the MQM analysis. Cofactors selected in ACS were removed whenever they did not return a LOD score above the genome-wide threshold in MQM results, while whenever new significant loci were found, these were included as cofactors (after ACS validation) and the MQM analysis was repeated. For IM, ACS, and MQM analyses a mixture model algorithm was used, while a P-value of 0.02 was used for ACS. Additive and dominance effects were also reported, as calculated by MapQTL software.

Detected QTLs explaining 10% or more of phenotypic variance (VE) were considered as major QTLs (Mallikarjuna et al. 2017; Wang et al. 2017). Confidence intervals were constructed by 2-LOD support interval at either side of the detected QTL (Van Ooijen 2009), with known markers immediately adjacent at either side of this map region taken as flanking markers. QTL location and confidence intervals were visualised on LGs using MapChart 2.32 (Voorrips 2002).

3 Results

Genome-wide (GW) LOD threshold ranged from 3.8 to 3.9 and 3.4 to 3.6 when PT was carried out using the complete map and the framework map, respectively. For each trait, it has been reported the whole analysis

using the complete map, starting with KW test, IM, and MQM, and percentage of explained phenotypic variance (%VE). All traits in IM analysis had a number of iterations needed to reach tolerance equal to 4, while during MQM the value decreased to 2. A summary of the analyses for all traits is presented in Table 3.1, while putative and significant QTLs from IM analysis are presented in Figure 3.1.

3.1 Morphological traits

Internode length **(InL; ND)**: KW test suggested two significant (p= 0.005) markers, on LG 5 and LG3, although both showed a LOD-score ≤ GW-1 in the interval mapping analysis (2.59 and 2.10 respectively). MQM analysis, however, confirmed the one on LG5 (marker 24841_65, at 68.66cM) with 6.16 LOD, and reported a putative QTL at 46.77cM (6.54 LOD). This was less than 1cM apart from the marker 7641_27 (47.26cM), although the latter showed a LOD score of 2.63. Another putative QTL linked to 12485_19 on LG3 (3.09 LOD, at 124.97cM) was also reported. The three QTLs accounted respectively for 22.7, 12.2, and 10.5% of phenotypic variance (45.5% altogether) from MQM results, while the two markers from IM explained a cumulative 23.6% of variance.

Number of nodes **(NoN; Non-ND)**: Trait-marker association and IM analysis both showed a QTL on LG4 (marker 19348_18, at 18.77cM) explaining 12.9% of variance, with a significance level at 0.005 but a LOD score of 2.57 (GW= 3.9). MQM reported a significant LOD score for the same QTL, and for another one linked to a marker on LG5 at 26.22cm (13499_33), accounting for 17.5 and 17% of variance, respectively.

Stem length **(StL, Non-ND)**: The non-parametric test showed a significant (p= 0.0005) link between marker 24841_65 on LG5 and this trait, with IM confirming a segregating QTL in the same position (LOD score of 4.38). The QTL explained 21.6% of the phenotypic variation and relatively high dominance effects (54.9), and it was also the same (non-significant) marker found in *internode length*.

MQM analysis agreed with this result, while adding a putative QTL (3.43 LOD) on LG3, represented by the marker 19348_18, the same reported for *number of nodes*.

***Branch number per plant* (NoB, Non-ND)** and ***Total branch number per plant* (NoBTOT)**: for NoB, the IM analysis showed only a putative QTL candidate on LG2 (15602_68, 3.47 LOD score, 14.16cM). This was showed to be a significant QTL in IM analysis for NoBTOT (4.43 LOD score), with the addition of a putative QTL on LG1 (marker 22986_20), for a cumulative 38.9%VE.

Automatic Cofactor Selection confirmed 15602_86 for LG2, and changed from marker 22986_20 to 10035_10 for the QTL on LG1 (from position 144.62 to 142.23cM). MQM analysis reported as significant both these two QTLs, accounting for 19.2 and 15.1% of phenotypic variance.

***Length of Branch* (LoB; ND)**: IM analysis showed a putative QTL (13849_61) on LG3 (KW test significant at p= .001, 3.60 LOD) explaining 17.5% of variance and relatively high additive and dominance effect (25.8; -30.6 respectively). MQM analysis reported the same QTL, but with a significant LOD score (4.85) and explaining 20.3%VE (29 and -32 of additive and dominance effect respectively). Such analysis also showed a putative QTL on LG8, linked to 15972_63 at 53.78cM (3.46 LOD, 13.9%VE).

3.2 Yield-related traits

***Pods per plant* (PPP; Non-ND)**: KW test revealed 7847_39 at 22.04cM on LG3 significantly (p= 0.005) linked with PPP. Interval mapping and MQM analysis showed similar results, with 7847_39 showing a LOD score of 2.99, explaining 15.5% of variance (5.1 and 8.6 of additive and dominance effect respectively).

***Seeds per pod* (SPP; ND)**: Putative QTL candidates were found on two linkage groups. KW test found 18511_24 at 7.92cM on LG7 (p= 0.001), and 15840_66 at 0.00cM on LG6 (p= 0.005), with IM confirming them as putative QTL candidates (3.42 and 3,15 LOD respectively), explaining altogether 33.7% of phenotypic variance. MQM was consistent with IM analysis, although the QTL on LG6 fell below the GW-1 threshold, despite ACS retaining the marker.

***Dry pod length* (PoL; ND)**: Marker-trait analysis found 24171_29 (LG4) and 11154_46 (LG9) significantly linked to this trait, with K* values of 20.06 and 13.49 at p= 0.0001 and 0.005 respectively. Results from IM

showed the first to be a significant QTL (4.57 LOD score), while the second to be a putative one (2.98 LOD score). MQM, however, returned only 24171_29, explaining 22.6% of variance.

Hundred-seed weight **(HSW; ND)**: KW test revealed 8814_54 at 66.75cM on LG4 to be significantly (K*= 10.46 at *p*= .01) linked to seed weight. The same locus showed a LOD score lower than the GW-1 threshold. In contrast, MQM analysis reported the same marker to be a significant QTL (3.86 LOD score, 17.4%VE), along with another putative QTL on the same LG at 98.11cM (3.22 LOD score, 14.3%VE).

3.3 Developmental and physiological traits

Days to first open flower **(DtF; Non-ND)**: A single marker on LG9 was linked to DtF (*p*= 0.01), although IM did not reveal any putative or significant QTL for this trait.

Pod maturing time **(PMT; Non-ND)**: 22543_38 (LG9, 106cM) was found significantly linked to pod maturing time (*p*= 0.005). IM results showed a putative QTL linked to the same marker, with LOD score of 3.05 and explaining 15.4% of variance. MQM reported the same results showing also a QTL on LG8, although the LOD score was below the threshold for a putative one.

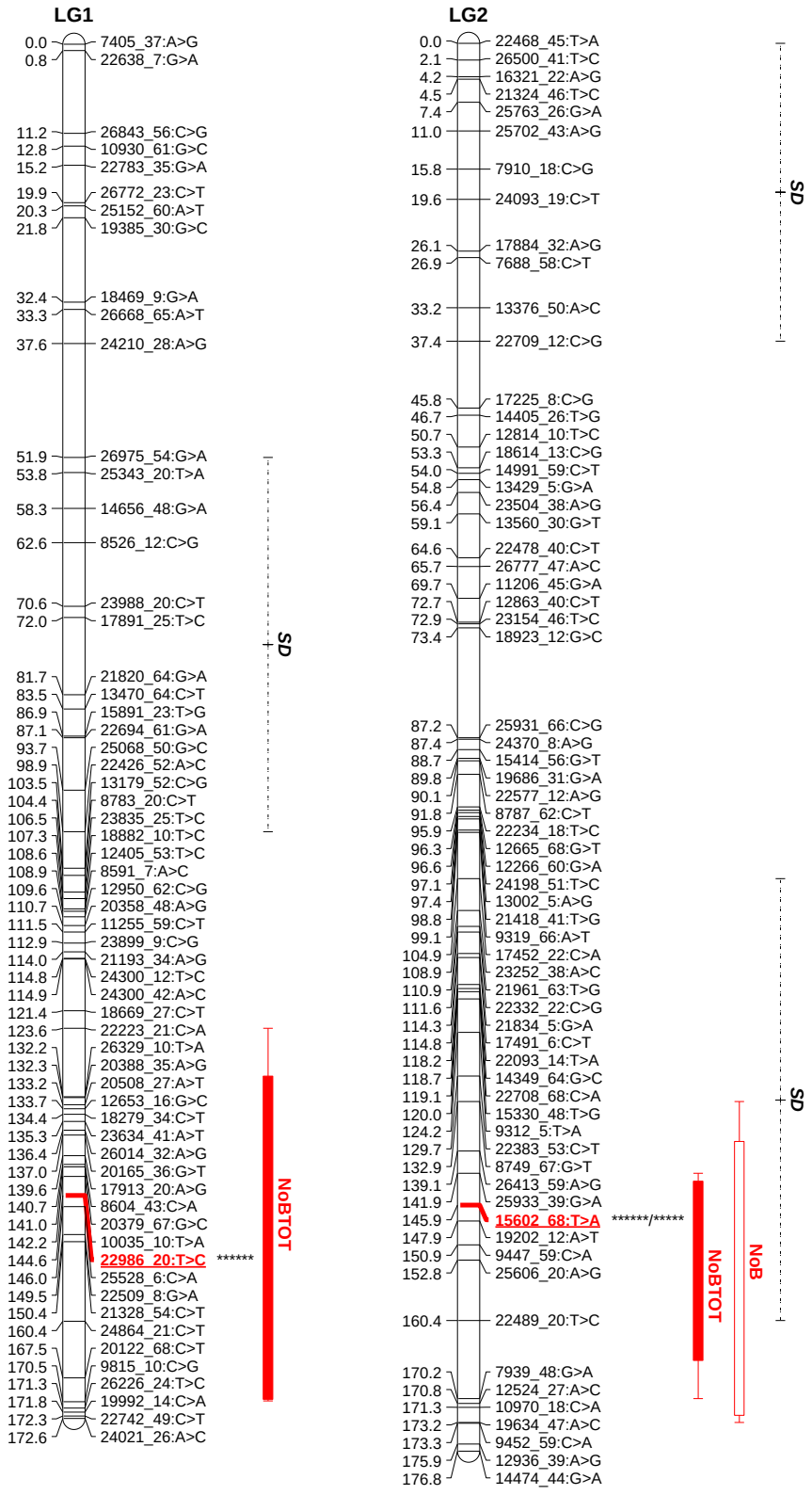
Chlorophyll concentration **(SPAD; Non-ND)**: KW test showed extensive regions on LG4 and LG5 with markers significantly linked to this trait. The two most significant (*p*= 0.0001) markers, 19643_19 on LG5 at 35.74cM and 27087_27 on LG4 at 87.21cM, were reported also in the IM results, which confirmed the presence of significant QTLs linked to them. Together, the two QTLs could explain 70.1% of phenotypic variance. MQM analysis was consistent with the findings from interval mapping.

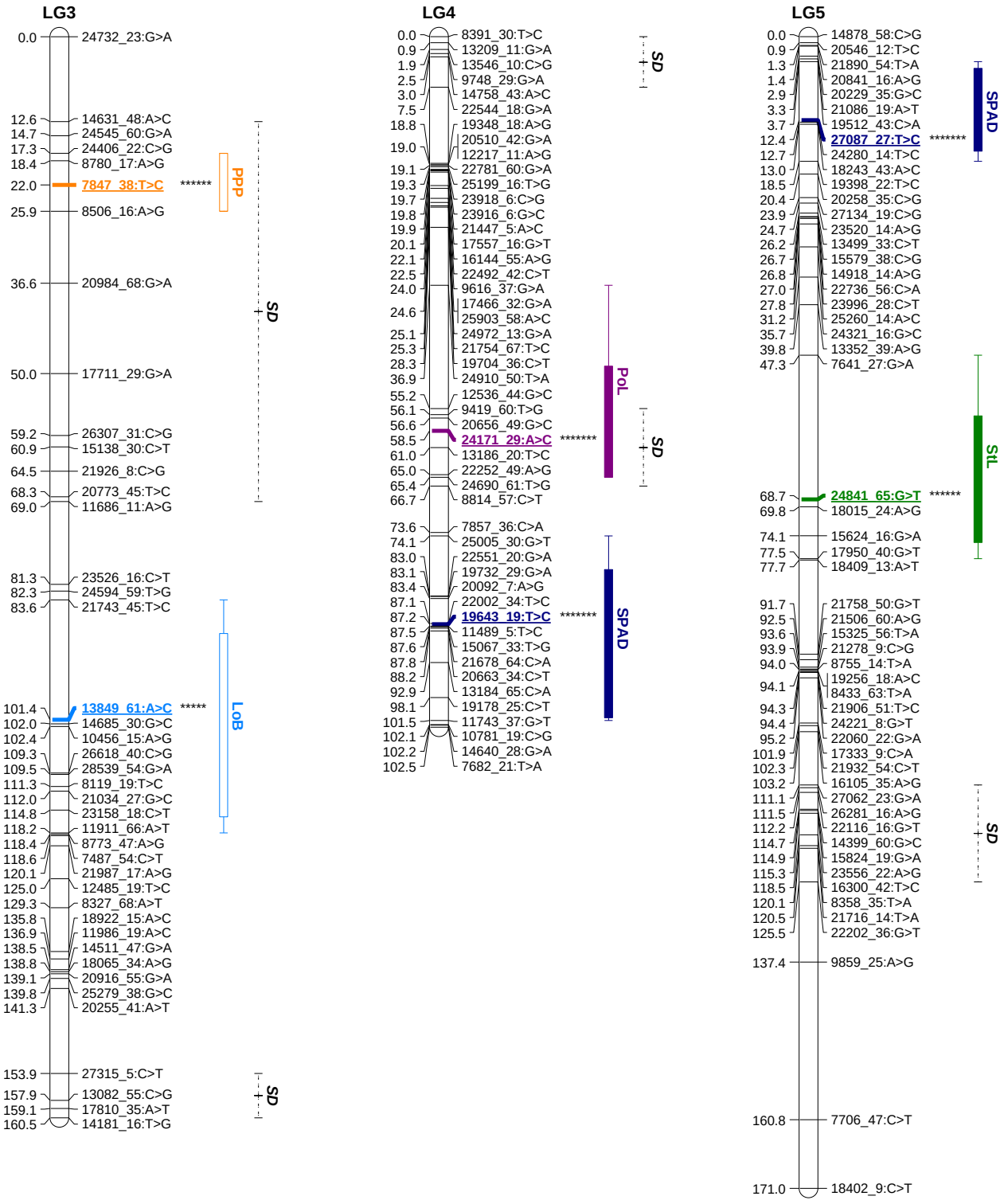
Table 3.1: Summary of KW test, Interval Mapping (IM), and Multiple-QTL Model (MQM) analyses for all included phenotypic traits.

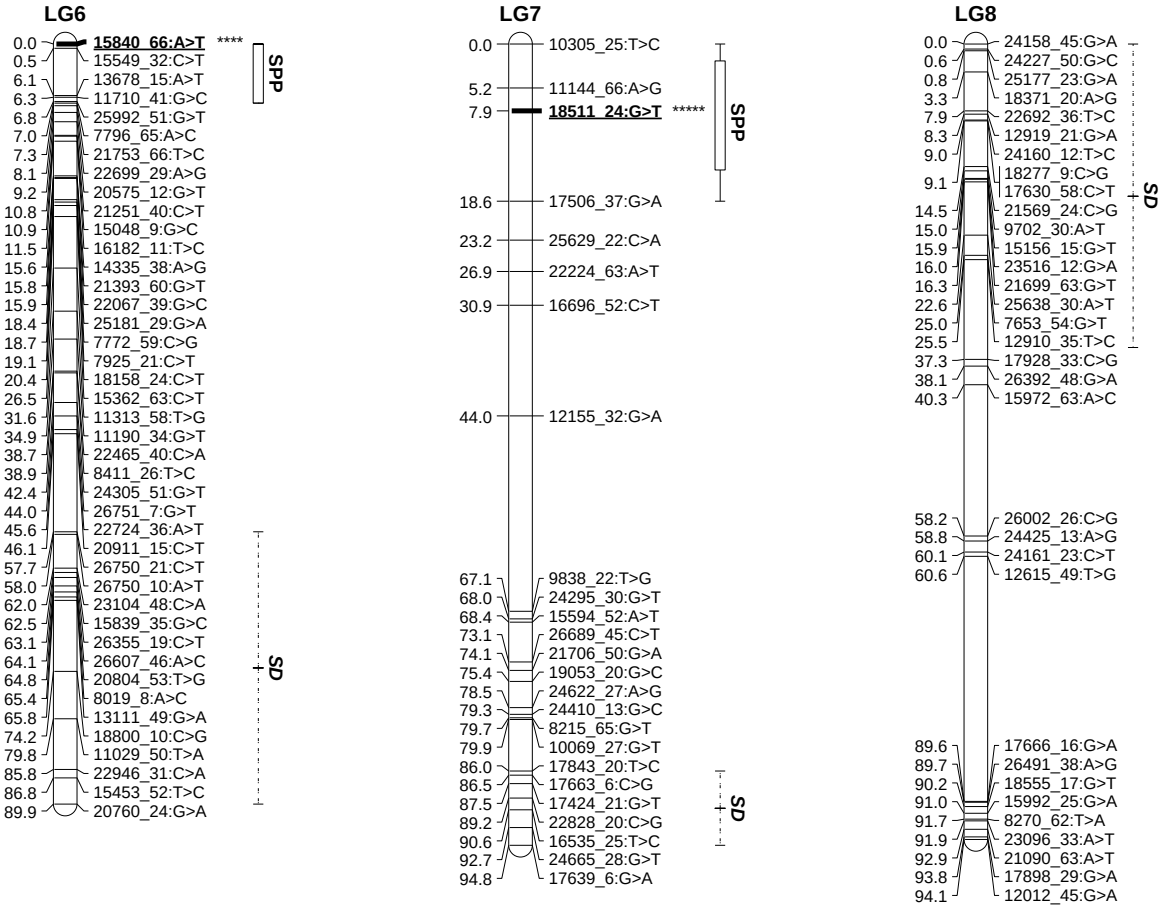
GW: genome-wide LOD threshold from permutation tests; Locus: marker in the indicated position; LG: linkage group; K*: Kruskal-Wallis test statistics; 3* to 7*: p= .01 to .0001; %VE: percentage of phenotypic variance explained; Ad. and Do.: additive and dominance effect; 2-LOD: support interval; SD: significantly distorted markers.

					KW		IM					MQM							
Trait	**GW**	**Locus**	**LG**	**Position**	**K***	**Sign. Level**	**LOD**	**%VE**	**Ad.**	**Do.**	**2-LOD (cM)**	**Locus**	**LG**	**Position**	**LOD**	**%VE**	**Ad.**	**Do.**	**SD**
InL	3.8	24841_65:G>T	5	68.66	12.25	4*	2.59	13.0	-0.5	1.8		24841_65:G>T	5	68.66	6.16	22.7	1.33	2.74	
		18922_15:A>C	3	135.76	11.08	4*	2.10	10.6	-1.1	-0.7		7641_27:G>A	5	47.26	2.63	12.2	-1.27	-0.32	
												12485_19:T>C	3	124.97	3.09	10.5	-1.04	-0.97	
NoN	3.9	19348_18:A>G	4	18.77	11.36	4*	2.57	12.9	-1.6	-0.8		19348_18:A>G	4	18.77	4.15	17.5	-0.198	-0.74	
												13499_33:C>T	5	26.22	4.05	17	1.58	-1.5	
StL	3.9	24841_65:G>T	5	68.66	16.40	6*	4.38	21.6	-19.6	54.9	18.9	24841_65:G>T	5	68.66	5.6	23.7	-20.13	58	
												19348_18:A>G	4	18.77	3.43	13.6	-33.91	-15.83	
NoB	3.8	15602_68:T>A	2	145.86	14.16	5*	3.47	16.9	1.0	-0.2	34.4	15602_68:T>A	2	145.86	3.47	16.9	1.04	-0.15	x
NoBTOT	3.9	15602_68:T>A	2	145.86	18.17	6*	4.43	21.1	1.8	-0.4	22.5	15602_68:T>A	2	145.86	4.92	19.2	1.45	-0.74	x
		22986_20:T>C	1	144.62	15.08	6*	3.66	17.8	-1.5	-0.2	40.9	10035_10:T>A	1	142.23	3.98	15.1	-1.29	-0.7	
LoB	3.8	13849_61:A>C	3	101.38	14.38	5*	3.60	17.5	25.8	-30.6	27.2	13849_61:A>C	3	101.38	4.85	20.3	29.35	-32.61	
												15972_63:A>C	8	53.78	3.8	13.5	-18.64	26.22	
PPP	3.9	7847_38:T>C	3	22.04	16.79	6*	2.99	15.5	4.1	5.1	8.6	7847_38:T>C	3	22.04	2.99	15.5	4.11	5.06	x
SPP	3.8	18511_24:G>T	7	7.92	14.91	5*	3.42	17.5	1.9	-0.3	12.9	18511_24:G>T	7	7.92	3.42	17.5	1.88	-0.31	x
		15840_66:A>T	6	0.00	13.36	4*	3.15	16.2	1.31	0.63	7.0	15840_66:A>T	6	0.00	2.76	11.8	1.17	0.50	
PoL	3.8	24171_29:A>C	4	58.52	20.06	7*	4.57	22.6	-1.4	0.7	16.5	24171_29:A>C	4	58.52	4.57	22.6	-1.41	0.65	x
		11154_46:G>T	9	16.21	13.49	4*	2.98	15.4	-1.4	-1.3	18.4	11154_46:G>T	9	16.21	1.94	8.0	-1.04	-1.04	x
HSW	3.8	8814_57:C>T	4	66.75	10.46	3*	2.64	13.8	-2.5	1.2		8814_57:C>T	4	66.75	3.86	17.4	-3.52	1.68	x
												19178_25:C>T	4	98.11	3.22	14.3	2.19	3.14	x
DtF	3.9	13440_29:C>T	9	8.82	9.40	3*	1.67	8.6	2.3	0.8		13440_29:C>T	9	8.82	1.67	8.6	2.3	0.8	x
PMT	3.8	22543_38:C>T	9	106.00	11.81	4*	3.05	15.4	-0.7	-5.2	19.5	22543_38:C>T	9	106.00	3.05	15.4	-0.73	-5.24	
												21699_63:G>T	8	77.74	2.59	11.2	2.83	2.60	
SPAD	3.9	27087_27:T>C	5	12.43	35.74	7*	9.58	40.1	-6.8	0.8	12.2	27087_27:T>C	5	12.43	11.47	33.9	-6.27	0.91	
		19643_19:T>C	4	87.21	22.42	7*	6.65	30.0	-5.5	4.3	22.0	22002_34:T>C	4	87.13	7.52	19.8	-4.39	3.71	

Grey-highlighted LOD values were < GW-1 threshol


LG1
0.0 7405_37:A>G
0.8 22638_7:G>A
11.2 26843_56:C>G
12.8 10930_61:G>C
15.2 22783_35:G>A
19.9 26772_23:C>T
20.3 25152_60:A>T
21.8 19385_30:G>C
32.4 18469_9:G>A
33.3 26668_65:A>T
37.6 24210_28:A>G
51.9 26975_54:G>A
53.8 25343_20:T>A
58.3 14656_48:G>A
62.6 8526_12:C>G
70.6 23988_20:C>T
72.0 17891_25:T>C
81.7 21820_64:G>A
83.5 13470_64:C>T
86.9 15891_23:T>G
87.1 22694_61:G>A
93.7 25068_50:G>C
98.9 22426_52:A>C
103.5 13179_52:C>G
104.4 8783_20:C>T
106.5 23835_25:T>C
107.3 18882_10:T>C
108.6 12405_53:T>C
108.9 8591_7:A>C
109.6 12950_62:C>G
110.7 20358_48:A>G
111.5 11255_59:C>T
112.9 23899_9:C>G
114.0 21193_34:A>G
114.8 24300_12:T>C
114.9 24300_42:A>C
121.4 18669_27:C>T
123.6 22223_21:C>A
132.2 26329_10:T>A
132.3 20388_35:A>G
133.2 20508_27:A>T
133.7 12653_16:G>C
134.4 18279_34:C>T
135.3 23634_41:A>T
136.4 26014_32:A>G
137.0 20165_36:G>T
139.6 17913_20:A>G
140.7 8604_43:C>A
141.0 20379_67:G>C
142.2 10035_10:T>A
144.6 22986_20:T>C
146.0 25528_6:C>A
149.5 22509_8:G>A
150.4 21328_54:C>T
160.4 24864_21:C>T
167.5 20122_68:C>T
170.5 9815_10:C>G
171.3 26226_24:T>C
171.8 19992_14:C>A
172.3 22742_49:C>T
172.6 24021_26:A>C
******
SD
NoBTOT
LG2
0.0 22468_45:T>A
2.1 26500_41:T>C
4.2 16321_22:A>G
4.5 21324_46:T>C
7.4 25763_26:G>A
11.0 25702_43:A>G
15.8 7910_18:C>G
19.6 24093_19:C>T
26.1 17884_32:A>G
26.9 7688_58:C>T
33.2 13376_50:A>C
37.4 22709_12:C>G
45.8 17225_8:C>G
46.7 14405_26:T>G
50.7 12814_10:T>C
53.3 18614_13:C>G
54.0 14991_59:C>T
54.8 13429_5:G>A
56.4 23504_38:A>G
59.1 13560_30:G>T
64.6 22478_40:C>T
65.7 26777_47:A>C
69.7 11206_45:G>A
72.7 12863_40:C>T
72.9 23154_46:T>C
73.4 18923_12:G>C
87.2 25931_66:C>G
87.4 24370_8:A>G
88.7 15414_56:G>T
89.8 19686_31:G>A
90.1 22577_12:A>G
91.8 8787_62:C>T
95.9 22234_18:T>C
96.3 12665_68:G>T
96.6 12266_60:G>A
97.1 24198_51:T>C
97.4 13002_5:A>G
98.8 21418_41:T>G
99.1 9319_66:A>T
104.9 17452_22:C>A
108.9 23252_38:A>C
110.9 21961_63:T>G
111.6 22332_22:C>G
114.3 21834_5:G>A
114.8 17491_6:C>T
118.2 22093_14:T>A
118.7 14349_64:G>C
119.1 22708_68:C>A
120.0 15330_48:T>G
124.2 9312_5:T>A
129.7 22383_53:C>T
132.9 8749_67:G>T
139.1 26413_59:A>G
141.9 25933_39:G>A
145.9 15602_68:T>A
147.9 19202_12:A>T
150.9 9447_59:C>A
152.8 25606_20:A>G
160.4 22489_20:T>C
170.2 7939_48:G>A
170.8 12524_27:A>C
171.3 10970_18:C>A
173.2 19634_47:A>C
173.3 9452_59:C>A
175.9 12936_39:A>G
176.8 14474_44:G>A
******/******
SD
NoBTOT
NoB


LG3
0.0 24732_23:G>A
12.6 14631_48:A>C
14.7 24545_60:G>A
17.3 24406_22:C>G
18.4 8780_17:A>G
22.0 7847_38:T>C ******
25.9 8506_16:A>G
36.6 20984_68:G>A
50.0 17711_29:G>A
59.2 26307_31:C>G
60.9 15138_30:C>T
64.5 21926_8:C>G
68.3 20773_45:T>C
69.0 11686_11:A>G
81.3 23526_16:C>T
82.3 24594_59:T>G
83.6 21743_45:T>C
101.4 13849_61:A>C *****
102.0 14685_30:G>C
102.4 10456_15:A>G
109.3 26618_40:C>G
109.5 28539_54:G>A
111.3 8119_19:T>C
112.0 21034_27:G>C
114.8 23158_18:C>T
118.2 11911_66:A>T
118.4 8773_47:A>G
118.6 7487_54:C>T
120.1 21987_17:A>G
125.0 12485_19:T>C
129.3 8327_68:A>T
135.8 18922_15:A>C
136.9 11986_19:A>C
138.5 14511_47:G>A
138.8 18065_34:A>G
139.1 20916_55:G>A
139.8 25279_38:G>C
141.3 20255_41:A>T
153.9 27315_5:C>T
157.9 13082_55:C>G
159.1 17810_35:A>T
160.5 14181_16:T>G
PPP
SD
LoB
SD
LG4
0.0 8391_30:T>C
0.9 13209_11:G>A
1.9 13546_10:C>G
2.5 9748_29:G>A
3.0 14758_43:A>C
7.5 22544_18:G>A
18.8 19348_18:A>G
19.0 20510_42:G>A
12217_11:A>G
19.1 22781_60:G>A
19.3 25199_16:T>G
19.7 23918_6:C>G
19.8 23916_6:G>C
19.9 21447_5:A>C
20.1 17557_16:G>T
22.1 16144_55:A>G
22.5 22492_42:C>T
24.0 9616_37:G>A
24.6 17466_32:G>A
25903_58:A>C
25.1 24972_13:G>A
25.3 21754_67:T>C
28.3 19704_36:C>T
36.9 24910_50:T>A
55.2 12536_44:G>C
56.1 9419_60:T>G
56.6 20656_49:G>C
58.5 24171_29:A>C *******
61.0 13186_20:T>C
65.0 22252_49:A>G
65.4 24690_61:T>G
66.7 8814_57:C>T
73.6 7857_36:C>A
74.1 25005_30:G>T
83.0 22551_20:G>A
83.1 19732_29:G>A
83.4 20092_7:A>G
87.1 22002_34:T>C
87.2 19643_19:T>C *******
87.5 11489_5:T>C
87.6 15067_33:T>G
87.8 21678_64:C>A
88.2 20663_34:C>T
92.9 13184_65:C>A
98.1 19178_25:C>T
101.5 11743_37:G>T
102.1 10781_19:C>G
102.2 14640_28:G>A
102.5 7682_21:T>A
SD
PoL
SD
SPAD
LG5
0.0 14878_58:C>G
0.9 20546_12:T>C
1.3 21890_54:T>A
1.4 20841_16:A>G
2.9 20229_35:G>C
3.3 21086_19:A>T
3.7 19512_43:C>A
12.4 27087_27:T>C *******
12.7 24280_14:T>C
13.0 18243_43:A>C
18.5 19398_22:T>C
20.4 20258_35:C>G
23.9 27134_19:C>G
24.7 23520_14:A>G
26.2 13499_33:C>T
26.7 15579_38:C>G
26.8 14918_14:A>G
27.0 22736_56:C>A
27.8 23996_28:C>T
31.2 25260_14:A>C
35.7 24321_16:G>C
39.8 13352_39:A>G
47.3 7641_27:G>A
68.7 24841_65:G>T ******
69.8 18015_24:A>G
74.1 15624_16:G>A
77.5 17950_40:G>T
77.7 18409_13:A>T
91.7 21758_50:G>T
92.5 21506_60:A>G
93.6 15325_56:T>A
93.9 21278_9:C>G
94.0 8755_14:T>A
94.1 19256_18:A>C
8433_63:T>A
94.3 21906_51:T>C
94.4 24221_8:G>T
95.2 22060_22:G>A
101.9 17333_9:C>A
102.3 21932_54:C>T
103.2 16105_35:A>G
111.1 27062_23:G>A
111.5 26281_16:A>G
112.2 22116_16:G>T
114.7 14399_60:G>C
114.9 15824_19:G>A
115.3 23556_22:A>G
118.5 16300_42:T>C
120.1 8358_35:T>A
120.5 21716_14:T>A
125.5 22202_36:G>T
137.4 9859_25:A>G
160.8 7706_47:C>T
171.0 18402_9:C>T
SPAD
StL
SD


LG6
0.0 15840_66:A>T ****
0.5 15549_32:C>T
6.1 13678_15:A>T
6.3 11710_41:G>C
6.8 25992_51:G>T
7.0 7796_65:A>C
7.3 21753_66:T>C
8.1 22699_29:A>G
9.2 20575_12:G>T
10.8 21251_40:C>T
10.9 15048_9:G>C
11.5 16182_11:T>C
15.6 14335_38:A>G
15.8 21393_60:G>T
15.9 22067_39:G>C
18.4 25181_29:G>A
18.7 7772_59:C>G
19.1 7925_21:C>T
20.4 18158_24:C>T
26.5 15362_63:C>T
31.6 11313_58:T>G
34.9 11190_34:G>T
38.7 22465_40:C>A
38.9 8411_26:T>C
42.4 24305_51:G>T
44.0 26751_7:G>T
45.6 22724_36:A>T
46.1 20911_15:C>T
57.7 26750_21:C>T
58.0 26750_10:A>T
62.0 23104_48:C>A
62.5 15839_35:G>C
63.1 26355_19:C>T
64.1 26607_46:A>C
64.8 20804_53:T>G
65.4 8019_8:A>C
65.8 13111_49:G>A
74.2 18800_10:C>G
79.8 11029_50:T>A
85.8 22946_31:C>A
86.8 15453_52:T>C
89.9 20760_24:G>A
SPP
SD
LG7
0.0 10305_25:T>C
5.2 11144_66:A>G
7.9 18511_24:G>T *****
18.6 17506_37:G>A
23.2 25629_22:C>A
26.9 22224_63:A>T
30.9 16696_52:C>T
44.0 12155_32:G>A
67.1 9838_22:T>G
68.0 24295_30:G>T
68.4 15594_52:A>T
73.1 26689_45:C>T
74.1 21706_50:G>A
75.4 19053_20:G>C
78.5 24622_27:A>G
79.3 24410_13:G>C
79.7 8215_65:G>T
79.9 10069_27:G>T
86.0 17843_20:T>C
86.5 17663_6:C>G
87.5 17424_21:G>T
89.2 22828_20:C>G
90.6 16535_25:T>C
92.7 24665_28:G>T
94.8 17639_6:G>A
SPP
SD
LG8
0.0 24158_45:G>A
0.6 24227_50:G>C
0.8 25177_23:G>A
3.3 18371_20:A>G
7.9 22692_36:T>C
8.3 12919_21:G>A
9.0 24160_12:T>C
9.1 18277_9:C>G
17630_58:C>T
14.5 21569_24:C>G
15.0 9702_30:A>T
15.9 15156_15:G>T
16.0 23516_12:G>A
16.3 21699_63:G>T
22.6 25638_30:A>T
25.0 7653_54:G>T
25.5 12910_35:T>C
37.3 17928_33:C>G
38.1 26392_48:G>A
40.3 15972_63:A>C
58.2 26002_26:C>G
58.8 24425_13:A>G
60.1 24161_23:C>T
60.6 12615_49:T>G
89.6 17666_16:G>A
89.7 26491_38:A>G
90.2 18555_17:G>T
91.0 15992_25:G>A
91.7 8270_62:T>A
91.9 23096_33:A>T
92.9 21090_63:A>T
93.8 17898_29:G>A
94.1 12012_45:G>A
SD

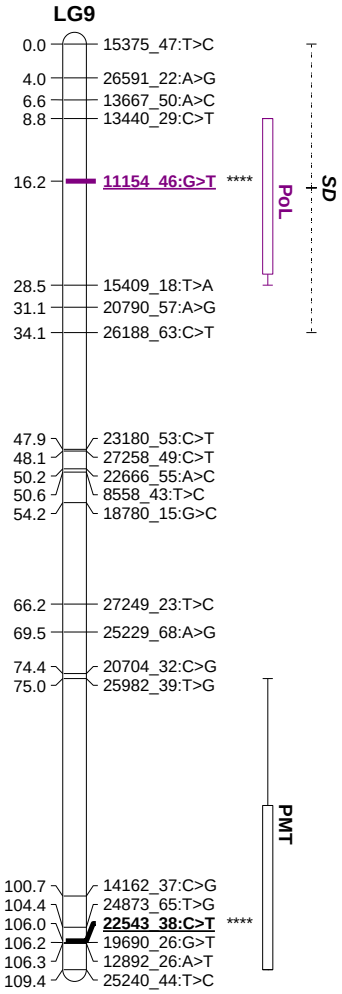


Figure 3.1: QTL mapping based on the complete map IM analysis. Asterisks (*) represent the significance level (*p*) from KW test (****= .005 to *******= .0001) for the marker linked to the putative or significant QTL. Rectangular bars represent the 2-LOD support interval, while lines show the flanking markers adjacent to the interval. Colour-filled rectangular bars represent significant QTLs, while empty bars represent putative QTLs. Segregation distortion (SD) regions are reported with dashed line on each LG.

3.4 Results using framework map

Similar results for morphological traits, *pod maturing time*, and *chlorophyll concentration* were obtained by carrying out the analysis with the framework map, i.e. without including any significantly distorted marker. QTLs previously showed for InL and NoN were even found above the GW-1 threshold, simply because of changes in GW LOD values from permutation tests. An exception occurred for NoBTOT, where the distorted marker on LG2 was not present, and no significant QTL was found anymore on such LG.

Differently, DtF and all yield-related traits revealed their QTLs to be linked to distorted markers during the first analysis (see Table 3.1). It is probably for this reason that when the framework map was used, IM analysis failed to find any QTL on the same or other LGs for such traits. The only exception was the putative QTL on LG6 for *seeds per pod*, which was found to be non-distorted in both analyses.

4 Discussion

4.1 Morphological traits

QTL analysis is a crucial step in order to unravel how many genomic regions control a phenotypic trait, what is their effect, and how they act. In this analysis, we have looked into 13 traits related to shoot morphology, yield, and development of winged bean, initially using a map with a density of about 3cM intermarker distance and 30% of distorted markers (i.e. the "complete" map described above). Although the high GW threshold and limited population size could have limited the discovery of QTL with minor effects, IM analyses have shown 8 putative QTL candidates and 5 significant QTLs across 9 traits, all explaining more than 10% of the phenotypic variation. MQM results have been treated carefully during the interpretation of the results, especially for those traits that do not showed normal phenotypic data distribution (see Table 2.1). The presence of a few major QTL per trait, in particular for morphological ones, is a common pattern found in most of the plant QTL studies reviewed (Ross-Ibarra 2005; Wang and Li 2008; Teichmann and Muhr 2015). Improved discovery, resolution, and effective estimation of QTLs would be achieved here by investigating a higher number of individuals, and by working on a population after a few generations of self-pollination (e.g. Recombinant Inbred Lines), allowing more recombination events to take place, and by replicating phenotyping across different locations and seasons. However, the information gathered from this study could allow the selection of individuals with some desired characteristics.

For *internode length* and *number of nodes*, marker-trait association analysis suggested the presence of QTLs on LG5 and 3, and LG4 respectively. IM analysis, though, showed no significant or putative QTL candidates on such linkage groups, whereas MQM results confirmed the QTL found in the non-parametric analysis (24841_65 and 18922_15 for InL; 19348_18 for NoN). In support of these finding, IM results for *stem length* found the same QTL on LG5 suggested for *internode length* (24841_65), while MQM analysis revealed as significant both this and the QTL previously shown for *number of nodes* (18922_15). Both InL and NoN phenotypic data were shown to be highly correlated to *stem length* (see Chapter 3), therefore a consensus across QTL analysis might be expected for these two fundamental morphological traits (InL and NoN) and the composite one (StL), such as the one shown here. Regarding the effect of these QTL, both showed relatively

high additive and dominance effects in the F_2, in particular with the 24841_65 allele coming from the parent M3 reducing StL and having a positive dominant effect over FP15-derived one.

The analysis of number of branches yielded better results when their total number along the main stem was taken into account. This could be ascribed to enhanced phenotypic differences for such trait among F_2 individuals: i.e. genotypes which had more branches in NoB, had even more branches in the NoBTOT dataset, compared to those which had lower branching habit. This is evidenced by the increased mean and variance between NoB and NoBTOT datasets (Chapter 3, appendix). Interval mapping, in both cases, showed a major QTL on LG2 linked to 15602_68, and a putative QTL was revealed on LG1. The presence of multiple genes with additive effect, rather than a single one controlling this trait, has been suggested before (Eagleton 1983) and could also explain the phenotypic distribution in the present F_2 (see histograms in Chapter 3, main section and appendix), although the two QTLs appeared to have opposite additive effects.

For *length of branch*, the significant QTLs revealed by MQM analysis have opposite effects in terms of additive and dominance effects and in both cases with relatively high values. Hence, in order to obtain individuals with longer branches it might be necessary to select individuals homozygous for the parental M3 allele for marker 13849_61, and homozygous for the parental FP15 allele for 15972_63. Although the segregating QTL found on LG3 seems close to the putative QTL found for InL, no significant correlation was found between these two traits (r= -0.05; p= 0.644).

The difference in QTL regions between NoB and LoB might be explained by different mechanisms underlying these two traits. While NoB might be related to the formation and initial release of secondary meristems (Mason et al. 2014; Teichmann and Muhr 2015), LoB could be the product of the internal control over the promotion of meristem growth itself (i.e. the consistent control from the main shoot apical meristem over secondary meristems). Differences in maintaining apical dominance over secondary branches were recorded in the male parent (M3) itself before flowering where, in some cases, axillary branches departing from lower nodes outgrow the main stem. In most of the F_2 individuals, branches remained below main shoot length, but a behaviour similar to M3 was observed in 11% of the population. The fact that different QTLs were found for

NoB (or NoBTOT) and for LoB, would encourage the idea of selecting for these two traits independently, given also the small and non-significant correlation found earlier between NoB and LoB phenotypic data (r_s= 0.20; p= 0.058, see Chapter 3).

4.2 Yield-related traits

Finding a way to achieve yield stability is one of the major targets in winged bean, in order to integrate this crop steadily into agricultural systems. The putative and segregating QTLs linked to yield components reported here could be used to support breeding efforts in this direction. For *pods per plant*, the finding of putative QTL on the same LG as LoB might be consistent with the significant positive correlation between these two traits (see Chapter 3), although the two QTLs were found about 80cM apart.

Seeds per pod and *dry pod length* reported QTLs on different LGs, perhaps suggesting that despite their correlation, QTLs detected for SPP might be more related to fertility, while PoL ones are related to the development of pod structure that accommodates developing seeds. Measured pod length, indeed, showed in some cases long structures in spite of a clear reduction in SPP: together with the current findings, it would suggest that the two traits could then be controlled by different pathways. Perhaps an interesting aspect, to consider for future studies is also the finding of the positive additive effect coming from the paternal alleles (M3), although this parent showed significantly less *seeds per pod* compared to the maternal parent FP15 (see Chapter 3).

Regarding HSW, unfortunately IM did not identify putative QTLs. However, trait phenotypic distribution allowed us to consider the MQM results. The 2 QTLs revealed by this latter are located on LG4, with 8814_57 located 8cM from a QTL linked to PoL: colocalization of QTLs underlying these traits have been reported before in mung bean (*Vigna radiata*) (Isemura et al. 2012) and faba bean (Ávila et al. 2017). However, in contrast with the mentioned studies, no correlation was found between these two traits in the present winged bean population.

4.3 Developmental and physiological traits

Analysis of developmental traits resulted only in a putative QTL for *pod maturing time*, which could be further explored. In particular, it would be worth exploring whether early pod maturity is due to mechanisms involving the whole plant, or the pod contribution to seed growth itself.

For *chlorophyll concentration*, the two significant QTLs could explain about 70% of phenotypic variance and were located on the same LGs as *stem length* (at a minimum distance of 50cM), although no phenotypic correlation was found between SPAD and these traits (see Chapter 3 appendix). It is not clear why *chlorophyll* accumulates later in newly formed leaves and the impact this has on plant photosynthetic rates. Whilst the plants eventually recover and do not show differences, using the markers found here would allow to select for individuals with extreme phenotypes and better understand the impact of this trait on final growth and development.

For all putative and significant QTLs from IM, the 2-LOD support interval was calculated and presented in Figure 3.1. In some cases, the presence of an abrupt decline in LOD score resulted in overlapping between the confidence interval and the flanking marker: this could probably be ascribed to the effect of population size, rather than marker misplacement, given that LOD score did not increased afterwards (Van Ooijen 2009).

Finally, with regards to the effect of SD on QTL detection, this has been suggested to depend on how closely linked are the distorted region and the QTL, and on population size. Zhang and colleagues (2010) proposed that SD will not impact on QTL detection, and the estimation of their position and effect; however, additive and dominance effects estimation have been suggested to be affected (Wen et al. 2013). The approach implemented here has included analysis using two maps, with and without SD markers, allowing us to verify directly the effect of SD on QTL analysis. The results using the framework map support the QTLs found in the first analysis that were not linked to a segregation distortion marker, as might be expected. For the remaining traits, lack of any QTL in non-distorted linkage space is as expected.

5 Conclusions

The QTL analysis reported in the present study has provided an initial insight into the genetic architecture of some important agnomical traits in winged bean. Using the markers linked to QTLs found here, could support the initial selection of individuals with a desired plant architecture, leading to shorter individuals with less dense canopies, and could allow for an increased planting density. Ultimately, this could pave the way for commercial cultivation of this important underutilised tropical legume. Yield-components especially, will require further studies to understand the role of the reported genomic regions. The information presented here could support future studies aimed to isolate candidate genes involved in relevant traits, for the improvement of winged bean.

References

Arumugan S, Perera ALT (1989) A study of the genetic architecture of some yield parameters of winged bean using model-fitting techniques. Trop Agric Res 1:

Ávila CM, Ruiz-Rodríguez MD, Cruz-Izquierdo S, et al (2017) Identification of plant architecture and yield-related QTL in Vicia faba L. Mol Breed 37:88. doi: 10.1007/s11032-017-0688-7

Cai G, Yang QQ, Chen H, et al (2016) Genetic dissection of plant architecture and yield-related traits in Brassica napus. Sci Rep 6:21625. doi: 10.1038/srep21625

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–71

Doust A (2007) Architectural Evolution and its Implications for Domestication in Grasses. Ann Bot 100:941–950. doi: 10.1093/aob/mcm040

Eagleton GE (1983) Evaluation of genetic resources in the winged bean (*Psophocarpus tetragonolobus* (L.) DC.) and their utilisation in the development of cultivars for higher latitudes. University of Western Australia

Harding J, Lugo Lopez. M. A., Perez Escolar R (1978) Promiscuous root nodulation of winged beans on an oxisol in Puerto Rico (*Psophocarpus tetragonolobus*). Trop. Agric.

Hildebrand DF, Harper JE, Hymowitz T (1981) Effects of nitrate level on nitrogen metabolism in winged bean and soya bean. Ann Bot 48:307–313. doi: 10.1093/oxfordjournals.aob.a086128

Iruthayathas EE, Vlassak K, Laeremans R (1985) Inheritance of nodulation and N2 fixation in winged beans. J Hered 76:237–242. doi: 10.1093/oxfordjournals.jhered.a110084

Isemura T, Kaga A, Tabata S, et al (2012) Construction of a Genetic Linkage Map and Genetic Analysis of Domestication Related Traits in Mungbean (Vigna radiata). PLoS One 7:e41304. doi: 10.1371/journal.pone.0041304

Kadam SS, Salunkhe DK, Luh BS (1984) Winged bean in human nutrition. C R C Crit Rev Food Sci Nutr 21:1–40. doi: 10.1080/10408398409527395

Koinange EMK, Singh SP, Gepts P (1996) Genetic Control of the Domestication Syndrome in Common Bean. Crop Sci 36:1037. doi: 10.2135/cropsci1996.0011183X003600040037x

Kortt AA, Caldwell JB (1984) Characteristics of the proteins of the tubers of winged bean (*Psophocarpus tetragonolobus* (L.) DC). J Sci Food Agric 35:304–313. doi: 10.1002/jsfa.2740350310

Liu B, Fujita T, Yan ZH, et al (2007) QTL mapping of domestication-related traits in soybean (*Glycine max*). Ann Bot 100:1027–1038. doi: 10.1093/aob/mcm149

Mallikarjuna BP, Samineni S, Thudi M, et al (2017) Molecular Mapping of Flowering Time Major Genes and QTLs in Chickpea (Cicer arietinum L.). Front Plant Sci 8:1–10. doi: 10.3389/fpls.2017.01140

Mason MG, Ross JJ, Babst BA, et al (2014) Sugar demand, not auxin, is the initial regulator of apical dominance. Proc Natl Acad Sci 111:6092–6097. doi: 10.1073/pnas.1322045111

Mayes S, Massawe FJ, Alderson PG, et al (2012) The potential for underutilized crops to improve security of food production. J Exp Bot 63:1075–1079. doi: 10.1093/jxb/err396

Prakash D, Misra PN, Misra PS (1987) Amino acid profile of winged bean (*Psophocarpus tetragonolobus* (L.) DC): a rich source of vegetable protein. Plant Foods Hum Nutr 37:261–264. doi: 10.1007/BF01091791

Ross-Ibarra J (2005) Quantitative trait loci and the study of plant domestication. Genetica 123:197–204. doi:

10.1007/s10709-004-2744-6
Schiavinato MA, Válio IFM (1996) Influence of staking on development of Winged Bean plants. RBrasFisiolVeg 8:99–103
Teichmann T, Muhr M (2015) Shaping plant architecture. Front Plant Sci 6:1–18. doi: 10.3389/fpls.2015.00233
Van Ooijen JW (2009) MapQTL ® 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species.
Varshney RK, Close TJ, Singh NK, et al (2009) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:202–210. doi: 10.1016/j.pbi.2008.12.004
Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. J Hered. doi: 10.1093/jhered/93.1.77
Wang J, Wang Z, Du X, et al (2017) A high-density genetic map and QTL analysis of agronomic traits in foxtail millet [Setaria italica (L.) P. Beauv.] using RAD-seq. PLoS One 12:e0179717. doi: 10.1371/journal.pone.0179717
Wang Y, Li J (2008) Molecular Basis of Plant Architecture. Annu Rev Plant Biol 59:253–279. doi: 10.1146/annurev.arplant.59.032607.092902
Weeden NF (2007) Genetic changes accompanying the domestication of Pisum sativum: Is there a common genetic basis to the "domestication syndrome" for legumes? Ann Bot 100:1017–1025. doi: 10.1093/aob/mcm122
Wen J, Can VT, Zhang YM (2013) Multi-QTL mapping for quantitative traits using distorted markers. Mol Breed 31:395–404. doi: 10.1007/s11032-012-9797-5
Zhang L, Wang S, Li H, et al (2010) Effects of missing marker and segregation distortion on QTL mapping in F_2 populations. Theor Appl Genet 121:1071–1082. doi: 10.1007/s00122-010-1372-z

Chapter 6 - Genetic diversity analysis in winged bean (*Psophocarpus tetragonolobus* (L.) DC.) based on DArTseq data

Abstract

Knowing how much genetic diversity is present among germplasm, allows us to design better strategies for conservation and improvement programmes. In winged bean, most of the held germplasm still has gaps in this context, increasing the danger of breeding programmes having a narrow genetic base to work with, and not planning comprehensive conservation projects to ensure that. Here we analysed 91 accessions from public and private collections of winged bean, using a total of 5981 Single Nucleotide Polymorphism (SNP) markers generated through Diversity Array Technology sequencing (DArTseq) technology. Genetic distance analyses showed potential agreement between genetic and geographic structure. Observed heterozygosity (7.3% average) support the idea of an inbreeding species, although within populations there still are relatively high levels of diversity. Germplasm from Indonesia, Malaysia, and Philippines, and Thailand showed higher differentiation among the analysed populations, while South American and African accessions are likely to be derived from South East Asian and Papua New Guinea stocks.

1 Introduction

Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) is a diploid species ($2n=2x=18$) belonging to the Fabaceae family, and the only domesticated species of genus *Psophocarpus* (Harder 1992; Harder and Smartt 1992; Vatanparast et al. 2016). It is cultivated for its green pod, tuberous roots, and mature seeds, with preferences between countries (NAS 1975; Eagleton 1999; Lepcha et al. 2017). The origin of the winged bean is a long-lasting matter of controversy: this has indeed been split between Africa, South Asia, South East Asia, and the South Pacific (Hymowitz and Boyd 1977; Verdcourt and Halliday 1978; Smartt 1980; Harder and Smartt 1992; Lepcha et al. 2017; Yang et al. 2018). Nonetheless, winged bean cultivation has not been reported in Africa, whereas it remains a crop mainly in tropical countries like India, Sri Lanka, Myanmar, Thailand, Malaysia, Indonesia, and Papua New Guinea (PNG). The latter has also been proposed as a centre of diversity for winged bean, based on phenotypic observation from Khan (1976), although Indonesia and Thailand were considered rich in variability (Pickersgill 1980).

To date, only a few attempts have been made to assess genetic diversity and population structure across winged bean germplasm. Using phenotypic markers (e.g. stem, or pod wing colour) on landraces sampled in highlands of PNG first revealed winged bean as an inbreeding species. Little gene-flow was observed, even between proximate locations and the little seen was probably due to occasional pollinators-assisted outcrossing (Erskine 1980; Erskine and Khan 1980; Pickersgill 1980). Subsequent studies with qualitative and quantitative traits (e.g. petals colour, days to flowering, seed weight, pod length, tubers yield) on a much broader range germplasm led by Thompson and Haryono (1980) suggested the presence of two major groups of germplasm for winged bean: Indonesian, and from Papua New Guinea. The authors went on suggesting that Nigerian material could have been derived from PNG, while Indonesian germplasm appeared to be phenotypically distinct from PNG, with late-maturing, and higher vegetative vigour and yield. Other studies seemed to support these initial observations, and included Thailand germplasm as another distinct group (Sastrapradja and Aminah Lubis 1975; Thompson and Haryono 1980; Pickersgill 1980; Anonymous 1981; Chandel et al. 1984). With the advent of DNA-based markers, it has become possible to evaluate genetic diversity in a more objective way, since genomic sequences are not biased by growing conditions, or by subjective recording of a trait. In winged bean, such studies were initially carried out using Inter Simple Sequence Repeats (ISSR) and Random Amplification of Polymorphic DNA (RAPD) markers (Mohanty et al. 2013; Chen et al. 2015). The first investigation involved a group of 24 accessions from at least 3 countries (India, Nigeria, Thailand, and additional accessions from unspecified countries), screened through 13 RAPD and 7 ISSR markers. Among these 2 types of markers, ISSR showed higher average PIC (0.281, against 0.170 from RAPD), diversity index (0.299, against 0.261), and polymorphism (94.5%, against 68.4%), although a Mantel test revealed significant correlation between distance matrices obtained from the two marker sets, and between each of them and the matrix obtained with the cumulative dataset. Later, 3 of these ISSR, and an additional 2 (average polymorphism of 65.7% for the 5 markers in total) were used to screen a different set of 45 accessions from PNG, Indonesia, Thailand, Nigeria, Columbia, Sri Lanka, Costa Rica, and China (Chen et al. 2015). Both studies led to UPGMA cluster analysis that revealed no consensus between genetic distance and geographical origin of the analysed accessions. A few years later, Simple Sequence Repeats (SSR) markers were developed for winged bean (Chapman 2015; Vatanparast et al. 2016; Wong et al. 2017). A set of 10 SSR markers (average

PIC of 0.546) were then implemented to investigate diversity across 53 accessions representing a broad germplasm with American, African, Asian, and South Pacific accessions. PCoA revealed no particular clustering according to geographical origins of the accessions. This was subsequently confirmed by the 4 clusters resulting from STRUCTURE analysis, 3 of which included accessions from all the major geographical areas (Yang et al. 2018). Overall, the three studies used different accessions and with different implemented markers, but reached the same conclusions by finding no agreement between observed clusters according to molecular data, and theoretical groups based on the declared geographical origin of analysed accessions. Yang and colleagues, given also the presence of admixture, suggested crossbreeding phenomena probably related to trading and sharing of seeds and ancestral admixed populations from which current accessions are derived, leading to contradictory passport data for some accessions.

In order to help resolve the disagreement between geographical and genetic clusters, higher density markers distributed along the entire genome could be used to determine the degree of relatedness among accessions from different countries, by looking at a larger portion of genomic regions. In this study we have carried out Genotyping-by-Sequencing (GbS) of accessions representing a broad geographical range, with 4 continents and 14 countries in total, coming from both publicly-available and private collections of winged bean. The aim was to evaluate genetic diversity of winged bean from different geographical areas, and sources, using Single Nucleotide Polymorphism markers (SNP) derived from DArTseq genotyping (Diversity Arrays Technology). This could eventually support the development and implementation of effective conservation, selection and utilisation strategies and in particular for desirable materials to be included in winged bean improvement programmes.

2 Material and Methods

2.1 Plant material

A total of 91 accessions were genotyped, representing 4 continents and 14 countries according to the sources of the germplasm. These were the World Vegetable Centre (WVC, previously known as Asian Vegetable Research and Development Center, WVC, Taiwan), the International Institute of Tropical Agriculture (IITA, Nigeria), the Malaysian Agricultural Research and Development Institute (MARDI, Malaysia), East-West Seed (EWS, Thailand), and a few more samples made available by private donors in Malaysia (see Table a.1

in appendix for full list of accessions). Accessions from WVC, according to the information available, have been collected in different regions of each country, and with different backgrounds (landrace or traditional cultivar, from backyards or village markets). Less information was available for IITA germplasm (http://my.iita.org/accession2/), in spite of the high volume of past research using these accessions.

Table 2.1: Summary of countries divided by continent represented in this study, and number of accessions from each of them.

Continent	Country (Code)	N. of Accessions	Sources
Africa	Liberia (LBR)	2	IITA
Africa	Nigeria (NGA)	4	EWS, IITA
America	Colombia (COL)	5	EWS
Asia (East)	China (CHN)	1	EWS
Asia (South East)	Indonesia (IDN)	10	WVC, EWS, IITA
Asia (South East)	Malaysia (MYS)	8	WVC, Donors, EWS, MARDI
Asia (South East)	Myanmar (MMR)	2	Donor
Asia (South East)	Philippines (PHL)	23	WVC, EWS
Asia (South East)	Singapore (SGP)	1	Donor
Asia (South East)	Thailand (THA)	22	WVC, EWS
Asia (South)	Bangladesh (BGD)	2	EWS, IITA
Asia (South)	Sri Lanka (LKA)	6	EWS
Oceania (South Pacific)	Fiji (FJI)	1	WVC
Oceania (South Pacific)	Papua New Guinea (PNG)	4	EWS, IITA

Six accessions included here were in common with the previous study from Yang and colleagues (2018): TPt-5,12,17,19,31,53 from IITA (see detailed table of accessions in the appendix below for countries of origin).

2.2 DNA extraction and genotyping by DArTseq

Genomic DNA was extracted from single plants of each accession using QIAGEN 96 Plant kit following manufacturer instruction. DNA integrity and concentration were verified with 1% agarose gel run in TAE buffer 80V (5µL of sample loaded) and NanoDrop 1000 (Thermo Scientific, NH, USA). Quality of DNA was checked by *Hin*dIII-HF restriction enzyme reaction with 100ng DNA in 1X CutSmart® buffer (New England BioLabs, Ipswich, MA) (37°C, 1h), followed by gel electrophoresis of the digested DNA. A total of 2µg of genomic DNA of each sample were sent to Diversity Array Technology Pty. Ltd. (Canberra, Australia) for DArTseq genotyping.

2.3 Marker selection and analysis

All markers with missing data were removed from the initial DArTseq dataset, leaving 5891 SNP markers across the 91 accessions, with a mean average polymorphism information content (PIC) of 0.25 (range 0 to

0.5). PIC was calculated for each SNP according to the formula PIC $= 1 - (f_a^2 + f_b^2)$, where f_a and f_b are the frequencies of the alleles *a* and *b*. Frequency and distance-based analyses were carried out using GenAlEx v6.5 (Peakall and Smouse 2012), for generating pairwise individual-by-individual genetic distance (GD) matrix, Nei's D coefficient genetic distance matrix, Principal Coordinate Analysis (PCoA), heterozygosity and F-statistics by populations, and Analysis of Molecular Variance (AMOVA, Codominant-Allelic). A similarity matrix was also created with the Dice's (Nei and Li's) coefficient in Genstat 18th Ed. (VSN International 2015) for pairwise individual-by-individual similarity. This matrix was then transformed for each value into a dissimilarity matrix by simple arithmetic subtraction (1- value) to carry out PCoA in GenAlEx and compare results obtained from the Euclidean GD matrix, as well as to test the correlation between the two matrices through a Mantel test (999 permutations) for cross-validation of results. MEGA 7.0 (Kumar et al. 2016) was used for cluster analysis and tree construction upon GD matrix using Neighbour-Joining method (NJ).

2.4 STRUCTURE analysis

An analysis using STRUCTURE software v2.3.4 (Pritchard et al. 2010) was carried out using the following parameters: length of burning period 200000, number of MCM reps after burning 500000, K 1 to 10 with 5 replicates each, using an admixture model (pre-configured settings), and allele frequencies correlated among populations. In order to reduce the computational load, only the SNPs with an PIC value above the median were considered: this included 2957 SNPs with an average PIC of 0.38 (range 0.16 to 0.50). Results were subsequently analysed using STRUCTURE HARVESTER (Earl and VonHoldt 2012), implementing the Evanno method (Evanno et al. 2005).

3 Results

3.1 Principal Coordinate Analysis based on pairwise individual-by-individual

PCoA was able to explain a cumulative 31.35% of the variation between the two coordinates. Based on the GD matrix, some of the groups agreed with the country of origin, in particular for material originally from the Philippines and Thailand. Accessions from other South-East Asian countries (e.g. Malaysia and Indonesia) appeared to be closer and to form another group, including African and remaining material, which could not be differentiated (Figure 3.1).

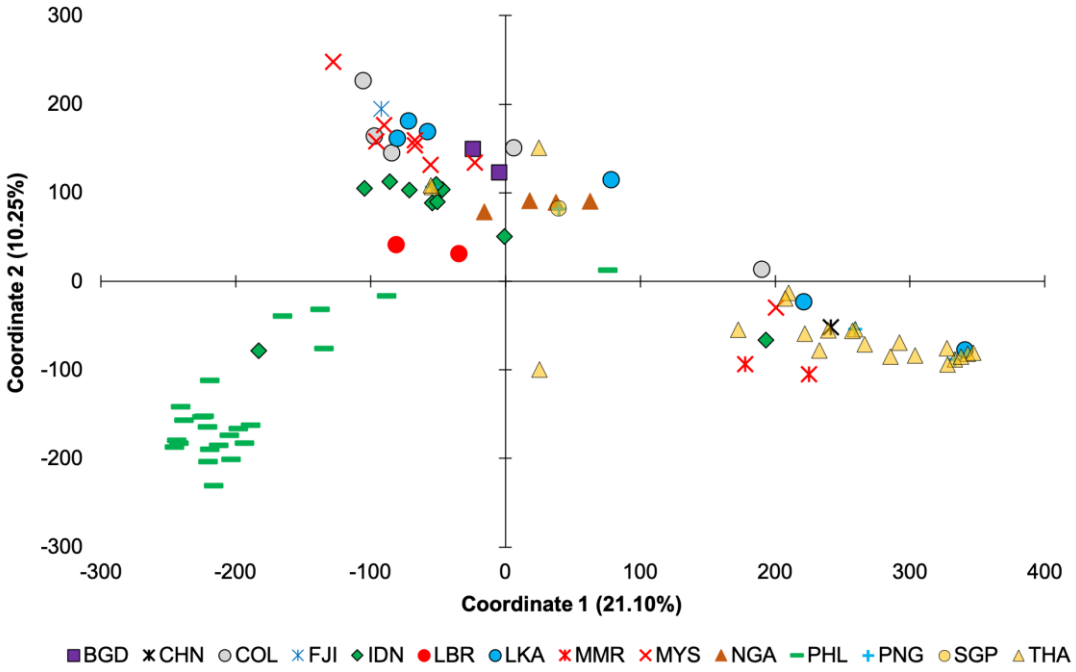


Figure 3.1: Principal Coordinate Analysis (PCoA) from pairwise individual-by-individual genetic distance (GD) analysis of accessions from 15 countries. 31.35% of cumulative variation was explained by the 2 axes. Legend reports the symbols for each country of origin (see country codes in Table 2.1, material and methods)

Pairwise genetic distance (GD) from GenAlEx and Dice's coefficient (Nei and Li) similarity matrix (transformed into dissimilarity matrix) from Genstat showed consistent results, and the Mantel test resulted in a significant correlation between these two matrices (R_{xy}= .828, p<.001 with 999 permutations). PCoA results upon Dice (Nei and Li's) coefficient-derived dissimilarity matrix are reported in the appendix.

3.2 Cluster analysis

Neighbour-joining tree from the pairwise GD matrix showed 6 potential clusters. Cluster I was formed by Malaysian and Indonesian accessions, with material coming from different sources (WVC, EWS, and MARDI), and the two countries generated two distinct subgroups. A certain degree of consistency was obserevd between cluster and country of origin was observed in clusters I, with accessions from Malaysia and Indonesia; in III, predominantly with accessions from Philippines; and the sub-cluster in cluster II, grouping Thai germplasm (Figure 3.2).

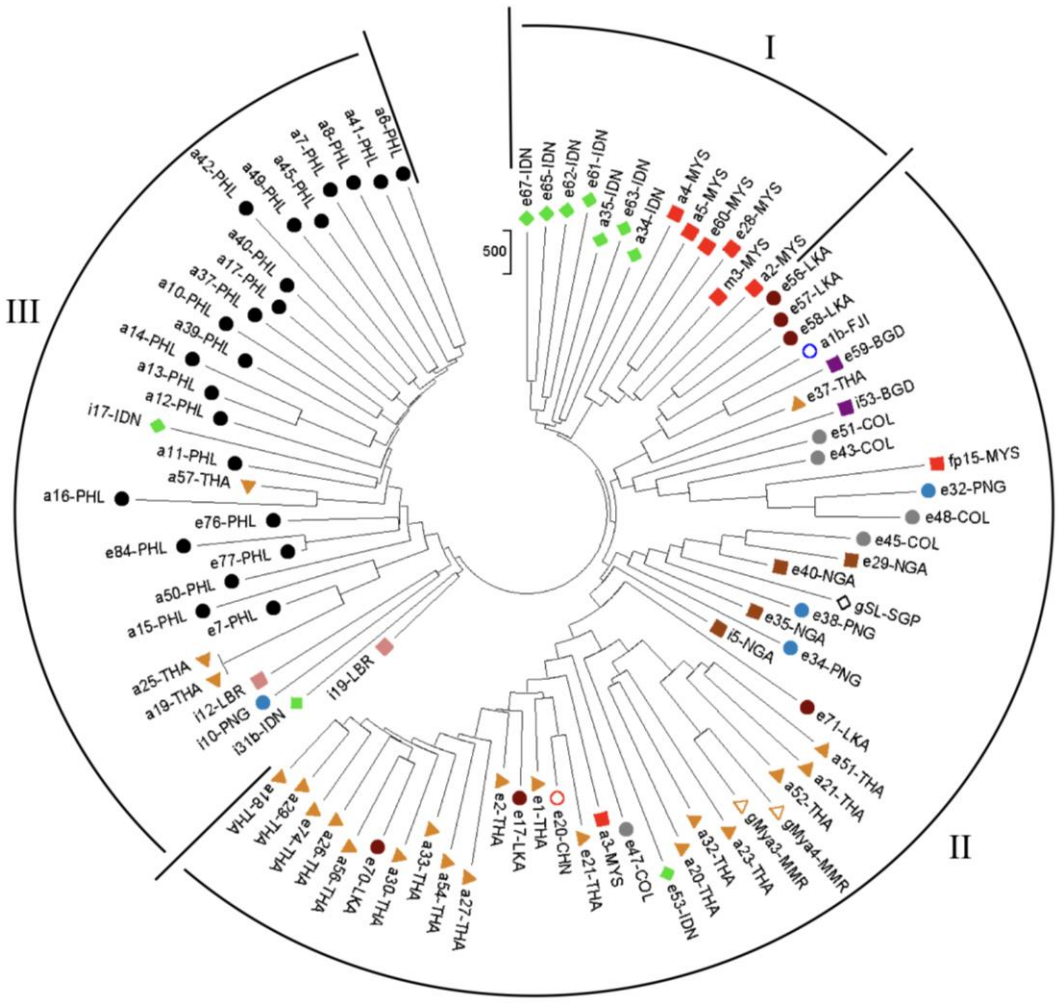


Figure 3.2: Neighbour-Joining (NJ) method tree generated using the GD matrix in MEGA 7. Branch lengths are to the same scale as genetic distances.

African material was split, with NGA accessions in cluster II and LBR accessions in cluster III, and they did not appear to be particularly related to each other. Both African germplasm sources, however, show relatedness to accessions from Papua New Guinea or Indonesia. Finally, cluster II has the highest number of countries of origins: 47 accessions were divided among 11 countries, with Thailand, Sri Lanka and Colombia being the most represented. Trees generated through the UPGMA method using the same GD matrix showed similar results, while NJ-trees based upon Nei's D coefficient distance matrix led to the same results (both reported in the appendix).

3.3 STRUCTURE results

Results from STRUCTURE HARVESTER (Figure 3.3) suggested a maximum of 7 clusters that could explain the underlying population structure of the included germplasm. However, looking directly into of the values from the Q-matrix revealed that above from 6 and above each accession was equally partitioned across clusters (i.e. each had a 1/K values for each cluster, with K being the number of clusters considered). This would indicate no real structure with such number of clusters.

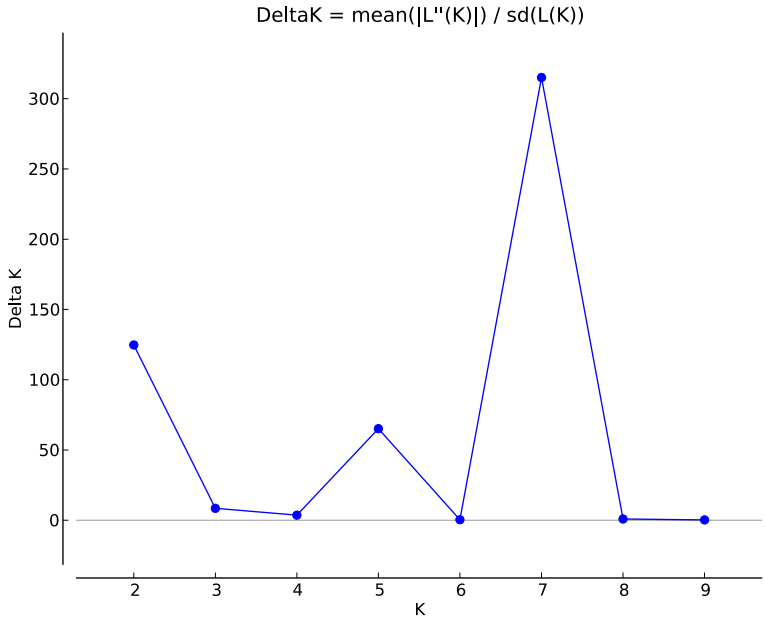


Figure 3.3: Delta-K graph, a parameter developed by Evanno et al. (2005) to predict the real number of clusters underlying the analyses population.

3.4 AMOVA and F-statistics

To reveal how genetically distinct the material is, an AMOVA based on allelic distance has been carried out. Eighty-seven accessions were allocated to groups according to their country of origin, or according to countries' geographical proximity and PCoA results: i.e. Thailand and Myanmar were merged, as well as Liberia and Nigeria. Populations, number of accessions, and descriptive indexes are presented in Table 3.1.

Table 3.1: Results from frequency analysis in GenAlEx. Mean Heterozygosity (H_o, observed; H_e, expected), Fixation Index (F), and percentage of polymorphic loci (P) calculated across 5891 SNPs in each population.

Pop	Accessions	Ho	He	F	P
COL	5	0.078±0.002	0.211±0.003	0.578±0.006	56.87%
IDN	10	0.068±0.001	0.219±0.002	0.658±0.005	65.83%
LBR+NGA	6	0.153±0.002	0.193±0.003	0.160±0.005	57.27%

LKA	6	0.050±0.001	0.189±0.003	0.674±0.005	50.94%
MYS	8	0.036±0.001	0.195±0.002	0.774±0.005	59.07%
PHL	23	0.089±0.001	0.218±0.002	0.538±0.004	79.41%
PNG	4	0.057±0.002	0.199±0.003	0.659±0.006	51.18%
THA+MMR	24	0.054±0.001	0.191±0.002	0.605±0.005	69.55%

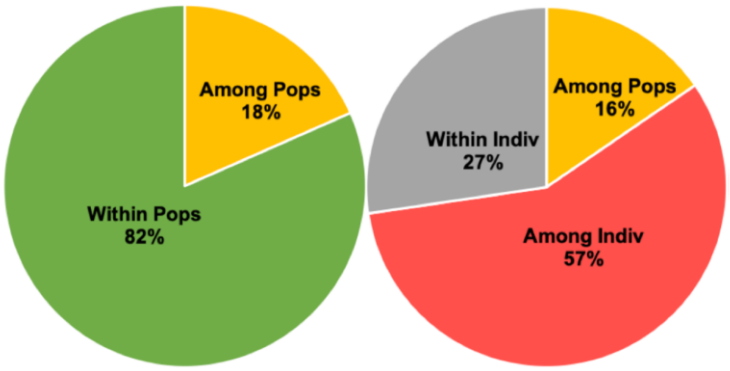
Among these populations, Malaysian material showed the lowest observed heterozygosity (H_o), resulting in the highest inbreeding coefficient (F= 0.774±0.005), while African material showed an opposite trend. THA+MMR accessions, the largest population, reported a relatively low level of observed heterozygosity, while having the second highest level of polymorphism; Sri Lankan material appeared relatively inbred as well (0.674±0.005), with relatively low level of heterozygosity, and low levels of polymorphism. For all populations, observed heterozygosity was lower than expected, with an average of 7.3%.

Table 3.2: Genetic differentiation between populations (pairwise F_{st} values test, probability based on 999 permutations).

	COL	**IDN**	**LBR+NGA**	**LKA**	**MYS**	**PHL**	**PNG**	**THA+MMR**
COL	-							
IDN	0.107**	-						
LBR+NGA	0.085*	0.109**	-					
LKA	0.143**	0.117**	0.156**	-				
MYS	0.134**	0.116**	0.160**	0.120**	-			
PHL	0.191**	0.138**	0.176**	0.208**	0.196**	-		
PNG	0.034	0.096**	0.060*	0.120*	0.135**	0.178**	-	
THA+MMR	0.201**	0.184**	0.195**	0.129**	0.225	0.254**	0.202**	-

**and *: significant difference at .001 and .05 respectively.

From the AMOVA analysis, the overall fixation index (F_{st}) value of 0.184 (F'_{st}= 0.234) indicated significant (p<0.001) and moderately strong genetic differentiation between the analysed populations. Pairwise population F_{st} values comparison, indeed, revealed moderate significant (p< 0.01) differentiation between all populations, except between PNG and COL germplasm (Table 3.2). A high degree of differentiation was shown between THA+MMR and PHL, with both these two populations differing from most of the others, while the least significant differentiation occurred between PNG and African populations.



	Variation Within Pop
COL	620.54
IDN	644.195
LBR+NGA	568.5347
LKA	555.625
MYS	575.3984
PHL	641.3554
PNG	587.1406
THA+MMR	562.0104
Average	594.3499

Figure 3.4: AMOVA results for codominant allelic distance matrix: suppressed (left) and non-suppressed (right) within individual analysis. The table reports variance within each population, calculated on sum of squared values from AMOVA (SSWP/N; SSWP sum of squares population, N number of accessions within the population).

Looking into the partitioning of genetic diversity between and within populations, the analysis showed that most of the variation was within population, rather than between them (see Figure 3.3 left). Repeating the analysis without suppressing within-individual variation, revealed intra-population and inter-individual variations as the main source of genetic variation, although between populations still showed 16% of variation. F-statistics in AMOVA also reported high and significant (p<0.001) inbreeding values within individuals, relative to each population and to the total (F_{is} and F_{it} values, 0.677 and 0.727, respectively). Finally, higher genetic distance variance within population was found in Indonesian and the Philippines germplasm, while Sri Lankan and THA+MMR accessions reported lower values (Figure 3.3, table on the right).

4 Discussion

Conservation and crop improvement programmes can be aided by a genetic diversity assessment of the held germplasm. In this study we used multiple approaches to investigate the diversity among accessions coming originally from 15 countries, including material from gene banks that could be requested by anyone working on winged bean. The PCoA based on a genetic distance matrix reveals possible agreement between geographical and genetic groups. This is true, in particular, for the Philippines, Thai, Malaysian, and Indonesian germplasm. In agreement with PCoA findings, the 3 clusters identified by NJ-tree, grouped most of the Philippines and Thai material according to the respective country of origin, and fairly distinct from each other. Malaysian and Indonesian accessions were also grouped according their country of origin, but these two groups were related enough to form a single cluster. Interpretation of both PCoA and NJ-tree results needs to

consider the more likely historical distribution of winged bean, which sees cultivation predominantly in Asia and the Pacific, and potential mislabelling due to seed trading and sharing during past international trials (Yang et al., 2018; Chapter 1). For these reasons, the observed closeness of accessions from Africa and South America to material from South East Asia and South Pacific could indicate that these latter are the true regions of origin for material found in the seed collections from the first two continents, a possibility once also suggested based on phenotypic observations of Nigerian accessions (Thompson and Haryono 1980). Indeed, African accessions were not differentiated into a clear separate cluster, but both LBR and NGA groups have some degree of relatedness to PNG or IDN germplasm. This is also in agreement with a previous STRUCTURE analysis that examined, among others, some of the accessions included here (TPt5, 12, 19; IITA) through SSR markers (Yang et al. 2018).

The results from STRUCTURE and STRUCTURE HARVESTER were interpreted according to the results from PCoA and NJ-tree. A maximum of five potentially meaningful clusters were identified, although it is likely that the results obtained from STRUCTURE HARVESTER could simply indicate a breaking-down of major clusters into sub-clusters, resulting in a higher K value. This might be the case with material from Indonesia, Malaysia, and/or the distinction between Thai germplasm from the rest of the accessions grouped into cluster II showed in the NJ-tree. Such value would then represent the maximum number of cluster to take into consideration. From a biological point of view, this would suggest once again that 2 to 3 is the real number of clusters that capture most of the underlying population structure, in agreement with the results observed in the PCoA analysis and given the more likely origin of accessions collected from countries where winged bean has probably been introduced by recent germplasm exchanges (see Chapter 1, Tanzi et al. 2019). The data interpretation has also to consider the limitations that come from using germplasm representing an inbreeding species, with potential relatedness among individuals, and with a structure likely to be far from a small number of discrete populations, therefore hampering the identification of an accurate cluster value through this software (Falush et al. 2003; Pritchard et al. 2010).

Observed heterozygosity, inbreeding coefficient, and percentage of polymorphic loci could give some insight into the population history. Although the analysed sample size could influence these indices, these also reflect the amount of allelic diversity in the population, outcrossing rates, admixture, and selective pressure. The

results, and in particular the inbreeding coefficient (F) and the percentage of polymorphic loci, appear to agree with what is expected in an inbreeding species. The consistently lower than expected heterozygosity could be related to a high degree of self-pollination, while higher polymorphism could point towards a high degree of genetic variability between individuals within a population. Also, the average observed heterozygosity of 7.3% recalls the 7.6% reported using phenotypic markers (Erskine 1980). Inbreeding and high variation could eventually lead to genetic differentiation, which appeared indeed to be significant, from AMOVA results. Relatively high intra-population and inter-individual genetic variability was confirmed in the AMOVA and F-Statistics analysis. In particular, genetic differentiation is higher between germplasm from Thailand-Myanmar, the Philippines, Malaysia, and Sri Lanka (F_{st} values test, Table 3.2). The high amount of genetic variability partitioned within populations could be due to grouping according to countries, which could include very diverse accessions in self-pollinating species. On the other hand, the seed trading/sharing, and the mislabelling mentioned earlier, could have led to overinflated within-population variability due to inclusion of accessions that actually did not originate in a particular country-group, but to another. Accessions from Papua New Guinea did show a relatively moderate amount of variability (see Table 3.1 and Figure 3.3), but further germplasm should be included in future studies to have a better picture of whether the country can be considered a centre of diversity and to particularly investigate the relationship with African material. The Sri Lankan population showed potentially the lowest variability (as indicated by observed heterozygosity, percentage of polymorphic loci, and genetic variability in AMOVA), and yet accessions split into 2 clusters in NJ-tree. A possible explanation, as for Colombian and African accessions, could be that the current accessions are derived from different initial populations, possibly imported at different times in the past. As in the case of PNG, further investigation with additional germplasm from each of these countries will likely clarify their genetic distance. Furthermore, the results discussed here about genetic diversity from AMOVA are based on grouping accessions according to their geographical origin, partially supported by the results from individual-by-individual pairwise comparison showed in PCoA and NJ-tree.

5 Conclusions

Unlike previous studies, groups observed in PCoA and NJ-tree analysis agreed to some extent with geographical structure of the analysed accessions. Genetic differentiation of groups of accessions from WVC

and IITA, along with private collections, has been reported. The genetic differentiation between germplasm from Thailand, Philippines, Indonesia, and Malaysia could support more effective selection of accessions within each of these population, to efficiently harness their genetic diversity in improvement programmes.

References

Anonymous (1981) The Winged Bean Flyer (3 No.2). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Chandel KPS, Pant KC, Arora RK (1984) Winged Bean in India. NBPGR Sci. Monogr. No. 8. National Bureau of Plant Genetics Resources, New Delhi, India

Chapman MA (2015) Transcriptome Sequencing and Marker Development for Four Underutilized Legumes. Appl Plant Sci 3:1400111. doi: 10.3732/apps.1400111

Chen D, Yi X, Yang H, et al (2015) Genetic diversity evaluation of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) using inter-simple sequence repeat (ISSR). Genet Resour Crop Evol 62:823–828. doi: 10.1007/s10722-015-0261-3

Eagleton G (1999) Winged Bean in Myanmar, Revisted. Econ Bot 53:342–352

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. doi: 10.1007/s12686-011-9548-7

Erskine W (1980) Measurements of the cross-pollination of winged bean in Papua New Guinea. SABRAO J. 12:11–14

Erskine W, Khan TN (1980) Variation within and between land races of winged bean (Psophocarpus tetragonolobus (L.) DC.). F Crop Res 3:359–364. doi: 10.1016/0378-4290(80)90041-6

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587

Harder DK (1992) Chromosome Counts in *Psophocarpus*. Kew Bull 47:529. doi: 10.2307/4110581

Harder DK, Smartt J (1992) Further evidence on the origin of the cultivated winged bean, *Psophocarpus tetragonolobus* (L.) DC. (Fabaceae): Chromosome numbers and the presence of a host-specific fungus. Econ Bot 46:187–191. doi: 10.1007/BF02930637

Hymowitz T, Boyd J (1977) Origin, ethnobotany and agricultural potential of the winged bean-*Psophocarpus tetragonolobus*. Econ Bot 31:180–188. doi: 10.1007/BF02866589

Khan TN (1976) Papua New Guinea: A centre of genetic diversity in winged bean (*Psophocarpus tetragonologus* (L.) D c.). Euphytica 25:693–705. doi: 10.1007/BF00041608

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874. doi: 10.1093/molbev/msw054

Lepcha P, Egan AN, Doyle JJ, Sathyanarayana N (2017) A Review on Current Status and Future Prospects of Winged Bean (*Psophocarpus tetragonolobus*) in Tropical Agriculture. Plant Foods Hum Nutr 72:225–235. doi: 10.1007/s11130-017-0627-0

Mohanty CS, Verma S, Singh V, et al (2013) Characterization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) based on molecular , chemical and physiological parameters. Am J Mol Biol 3:187–197. doi: 10.4236/ajmb.2013.34025

NAS (1975) The Winged Bean: A High-Protein Crop for the Humid Tropics. National Academy of Sciences, Washington, D.C.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28:2537–9. doi: 10.1093/bioinformatics/bts460

Pickersgill B (1980) Cytology of two species of winged bean, *Psophocarpus tetragonolobus* (L.) DC. and *P. scandens* (Endl.) Verdc. (Leguminosae). Bot J Linn Soc 80:279–291. doi: 10.1111/j.1095-8339.1980.tb01665.x

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: Version 2.3. 1–38

Sastrapradja S, Aminah Lubis SH (1975) *Psophocarpus tetragonolobus* as minor garden vegetable in Java. Bogor, Indonesia

Smartt J (1980) Some observations on the origin and evolution of the winged bean (*Psophocarpus tetragonolobus*). Euphytica 29:121–123. doi: 10.1007/BF00037256

Tanzi AS, Eagleton GE, Ho WK, et al (2019) Winged bean (Psophocarpus tetragonolobus (L.) DC.) for food and nutritional security: synthesis of past research and future direction. Planta 1–21. doi: 10.1007/s00425-019-03141-2

Thompson AE, Haryono SK (1980) Winged bean: unexploited tropical food crop. HortScience 15:233–237
Vatanparast M, Shetty P, Chopra R, et al (2016) Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Sci Rep 6:1–14. doi: 10.1038/srep29070
Verdcourt B, Halliday P (1978) A Revision of *Psophocarpus* (*Leguminosae-Papilionoideae-Phaseoleae*). Kew Bull 33:191. doi: 10.2307/4109575
VSN International (2015) Genstat *for Windows* 18th Edition.
Wong Q, Tanzi AS, Ho W, et al (2017) Development of Gene-Based SSR Markers in Winged Bean (Psophocarpus tetragonolobus (L.) DC.) for Diversity Assessment. Genes (Basel) 8:100. doi: 10.3390/genes8030100
Yang S, Grall A, Chapman MA (2018) Origin and diversification of winged bean (*Psophocarpus tetragonolobus* (L.) DC.), a multipurpose underutilized legume. Am J Bot 105:888–897. doi: 10.1002/ajb2.1093

Appendix – Chapter 6

a.1 List of accessions included into genetic diversity analysis

Table a.1: List of accessions included in this study, ordered by origin according to the source providing the seeds. Where available, the source original label has been included for each accession (e.g. for WVC material, it is reported the vegetable introduction number), along with region from where the accession was originally collected.

Continent	Country (Code)	Region	Study label	Source	Source label
Africa	Liberia (LBR)	-	i12	IITA (i)	TPt12
Africa	Liberia (LBR)	-	i19	IITA (i)	TPt19
Africa	Nigeria (NGA)	-	e29	EWS (e)	GB000029-1
Africa	Nigeria (NGA)	-	e35	EWS (e)	GB000035-1
Africa	Nigeria (NGA)	-	e40	EWS (e)	GB000040-1
Africa	Nigeria (NGA)	-	i5	IITA (i)	TPt5
America	Colombia (COL)	-	e43	EWS (e)	GB000043-1
America	Colombia (COL)	-	e45	EWS (e)	GB000045-1
America	Colombia (COL)	-	e47	EWS (e)	GB000047-1
America	Colombia (COL)	-	e48	EWS (e)	GB000048-1
America	Colombia (COL)	-	e51	EWS (e)	GB000051-1
Asia (East)	China (CHN)	-	e20	EWS (e)	GB000020-1
Asia (South-East)	Indonesia (IDN)	-	a34	WVC (a)	VI043344
Asia (South-East)	Indonesia (IDN)	-	a35	WVC (a)	VI043348
Asia (South-East)	Indonesia (IDN)	-	e53	EWS (e)	GB000053-1
Asia (South-East)	Indonesia (IDN)	-	e61	EWS (e)	GB000061-1
Asia (South-East)	Indonesia (IDN)	-	e62	EWS (e)	GB000062-1
Asia (South-East)	Indonesia (IDN)	-	e63	EWS (e)	GB000063-1
Asia (South-East)	Indonesia (IDN)	-	e65	EWS (e)	GB000065-1
Asia (South-East)	Indonesia (IDN)	-	e67	EWS (e)	GB000067-1
Asia (South-East)	Indonesia (IDN)	-	i17	IITA (i)	TPt17
Asia (South-East)	Indonesia (IDN)	-	i31b	IITA (i)	TPt31
Asia (South-East)	Malaysia (MYS)	-	fp15	Donor	NA
Asia (South-East)	Malaysia (MYS)	-	e28	EWS (e)	GB000028-1
Asia (South-East)	Malaysia (MYS)	-	e60	EWS (e)	GB000060-1
Asia (South-East)	Malaysia (MYS)	-	m3	MARDI (m)	NA
Asia (South-East)	Malaysia (MYS)	Sabah	a5	WVC (a)	VI034627
Asia (South-East)	Malaysia (MYS)	Sarawak	a2	WVC (a)	VI034544
Asia (South-East)	Malaysia (MYS)	Sarawak	a3	WVC (a)	VI034581
Asia (South-East)	Malaysia (MYS)	Sarawak	a4	WVC (a)	VI034582
Asia (South-East)	Myanmar (MMR)	-	gMya3	Donor	NA
Asia (South-East)	Myanmar (MMR)	-	gMya4	Donor	NA
Asia (South-East)	Philippines (PHL)	-	a49	WVC (a)	VI034446
Asia (South-East)	Philippines (PHL)	-	a50	WVC (a)	VI034466
Asia (South-East)	Philippines (PHL)	-	e7	EWS (e)	GB00007-1
Asia (South-East)	Philippines (PHL)	-	e76	EWS (e)	GB000076-1
Asia (South-East)	Philippines (PHL)	-	e77	EWS (e)	GB000077-1
Asia (South-East)	Philippines (PHL)	-	e84	EWS (e)	GB000084-1
Asia (South-East)	Philippines (PHL)	Bangasinan	a10	WVC (a)	VI034794
Asia (South-East)	Philippines (PHL)	Bangasinan	a37	WVC (a)	VI034797
Asia (South-East)	Philippines (PHL)	Bulacan	a13	WVC (a)	VI036282
Asia (South-East)	Philippines (PHL)	Ilocos Sur	a39	WVC (a)	VI034768
Asia (South-East)	Philippines (PHL)	Ilocos Sur	a40	WVC (a)	VI041583
Asia (South-East)	Philippines (PHL)	Ilocos Sur	a7	WVC (a)	VI034706
Asia (South-East)	Philippines (PHL)	Iloilo	a16	WVC (a)	VI041685
Asia (South-East)	Philippines (PHL)	La Union	a17	WVC (a)	VI043602
Asia (South-East)	Philippines (PHL)	La Union	a41	WVC (a)	VI034650
Asia (South-East)	Philippines (PHL)	La Union	a42	WVC (a)	VI034661
Asia (South-East)	Philippines (PHL)	La Union	a6	WVC (a)	VI034651
Asia (South-East)	Philippines (PHL)	Pangasinan	a45	WVC (a)	VI034750
Asia (South-East)	Philippines (PHL)	Pangasinan	a8	WVC (a)	VI034752
Asia (South-East)	Philippines (PHL)	Rizal	a14	WVC (a)	VI036285-A
Asia (South-East)	Philippines (PHL)	Rizal	a15	WVC (a)	VI036285-B

Asia (South-East)	Philippines (PHL)	Tarlac	a11	WVC (a)	VI034811
Asia (South-East)	Philippines (PHL)	Tarlac	a12	WVC (a)	VI034826
Asia (South-East)	Singapore (SGP)	-	gSL	Donor	NA
Asia (South-East)	Thailand (THA)	-	e1	EWS (e)	GB00001-1
Asia (South-East)	Thailand (THA)	-	e2	EWS (e)	GB00002-1
Asia (South-East)	Thailand (THA)	-	e21	EWS (e)	GB000021-1
Asia (South-East)	Thailand (THA)	-	e37	EWS (e)	GB000037-1
Asia (South-East)	Thailand (THA)	-	e74	EWS (e)	GB000074-1
Asia (South-East)	Thailand (THA)	Chanthaburi	a33	WVC (a)	VI041038
Asia (South-East)	Thailand (THA)	Chiang Mai	a51	WVC (a)	VI038095
Asia (South-East)	Thailand (THA)	Chiang Rai	a52	WVC (a)	VI038155
Asia (South-East)	Thailand (THA)	Kamphaeng Phet	a19	WVC (a)	VI038089
Asia (South-East)	Thailand (THA)	Kamphaeng Phet	a23	WVC (a)	VI039751
Asia (South-East)	Thailand (THA)	Kanchanburi	a54	WVC (a)	VI040854
Asia (South-East)	Thailand (THA)	Loei	a25	WVC (a)	VI040052
Asia (South-East)	Thailand (THA)	Lopburi	a26	WVC (a)	VI040591
Asia (South-East)	Thailand (THA)	Nakhon Sawan	a18	WVC (a)	VI038088
Asia (South-East)	Thailand (THA)	Phayao	a21	WVC (a)	VI038093
Asia (South-East)	Thailand (THA)	Phetchaburi	a29	WVC (a)	VI040769
Asia (South-East)	Thailand (THA)	Phetchaburi	a30	WVC (a)	VI040776
Asia (South-East)	Thailand (THA)	Phitsanulok	a20	WVC (a)	VI038091
Asia (South-East)	Thailand (THA)	Prachin Buri	a32	WVC (a)	VI041019
Asia (South-East)	Thailand (THA)	Prachin Buri	a56	WVC (a)	VI041008
Asia (South-East)	Thailand (THA)	Ratchaburi	a27	WVC (a)	VI040732
Asia (South-East)	Thailand (THA)	Uthai Thani	a57	WVC (a)	VI040662
Asia (South)	Bangladesh (BGD)	-	e59	EWS (e)	GB000059-1
Asia (South)	Bangladesh (BGD)	-	i53	IITA (i)	TPt53
Asia (South)	Sri Lanka (LKA)	-	e17	EWS (e)	GB000017-1
Asia (South)	Sri Lanka (LKA)	-	e56	EWS (e)	GB000056-1
Asia (South)	Sri Lanka (LKA)	-	e57	EWS (e)	GB000057-1
Asia (South)	Sri Lanka (LKA)	-	e58	EWS (e)	GB000058-1
Asia (South)	Sri Lanka (LKA)	-	e70	EWS (e)	GB000070-1
Asia (South)	Sri Lanka (LKA)	-	e71	EWS (e)	GB000071-1
Oceania-Pacific (South)	Fiji (FJI)	-	a1b	WVC (a)	VI034496
Oceania-Pacific (South)	Papua New Guinea (PNG)	-	e32	EWS (e)	GB000032-1
Oceania-Pacific (South)	Papua New Guinea (PNG)	-	e34	EWS (e)	GB000034-1
Oceania-Pacific (South)	Papua New Guinea (PNG)	-	e38	EWS (e)	GB000038-1
Oceania-Pacific (South)	Papua New Guinea (PNG)	-	i10	IITA (i)	TPt10

a.2 Additional PCoA and NJ-tree graphs using alternative coefficients.

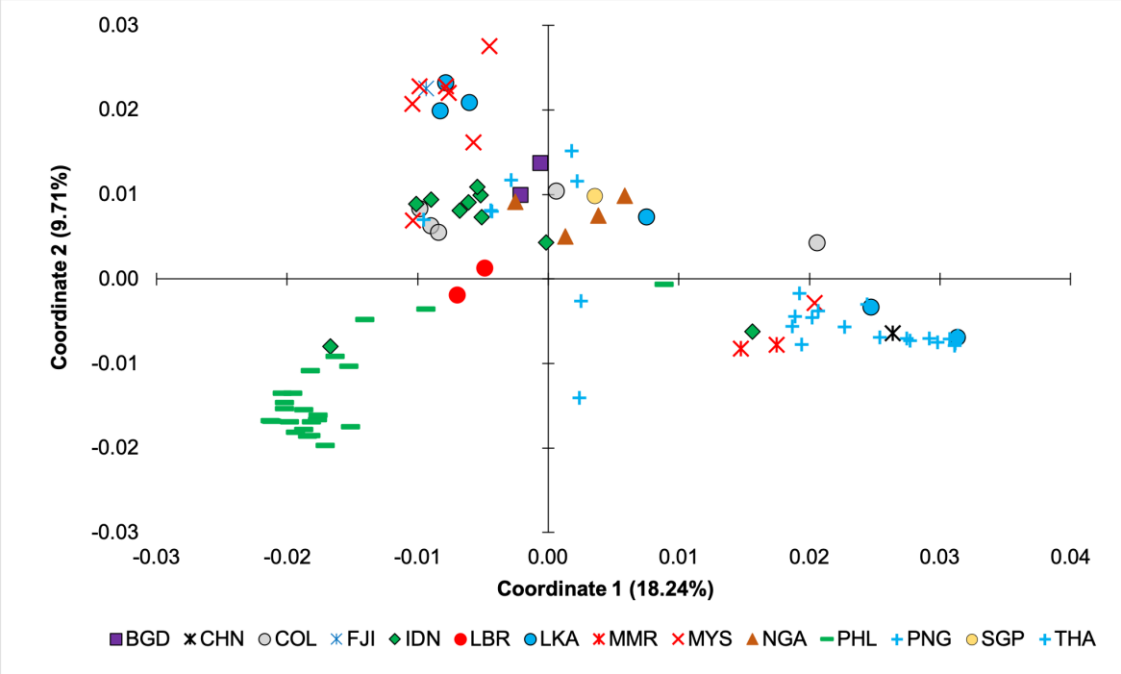


Figure a.1: PCoA from Nei and Li's (Dice) coefficient similarity matrix from Genstat. Values were transformed (1-value) to carry out the analysis in GenAlEx.

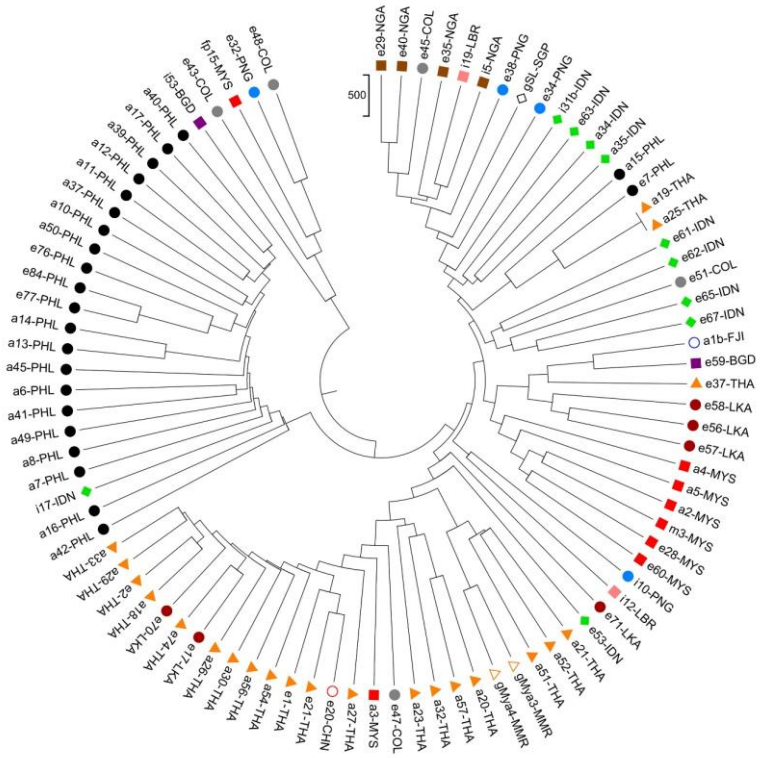


Figure a.2: UPGMA tree from GD matrix.

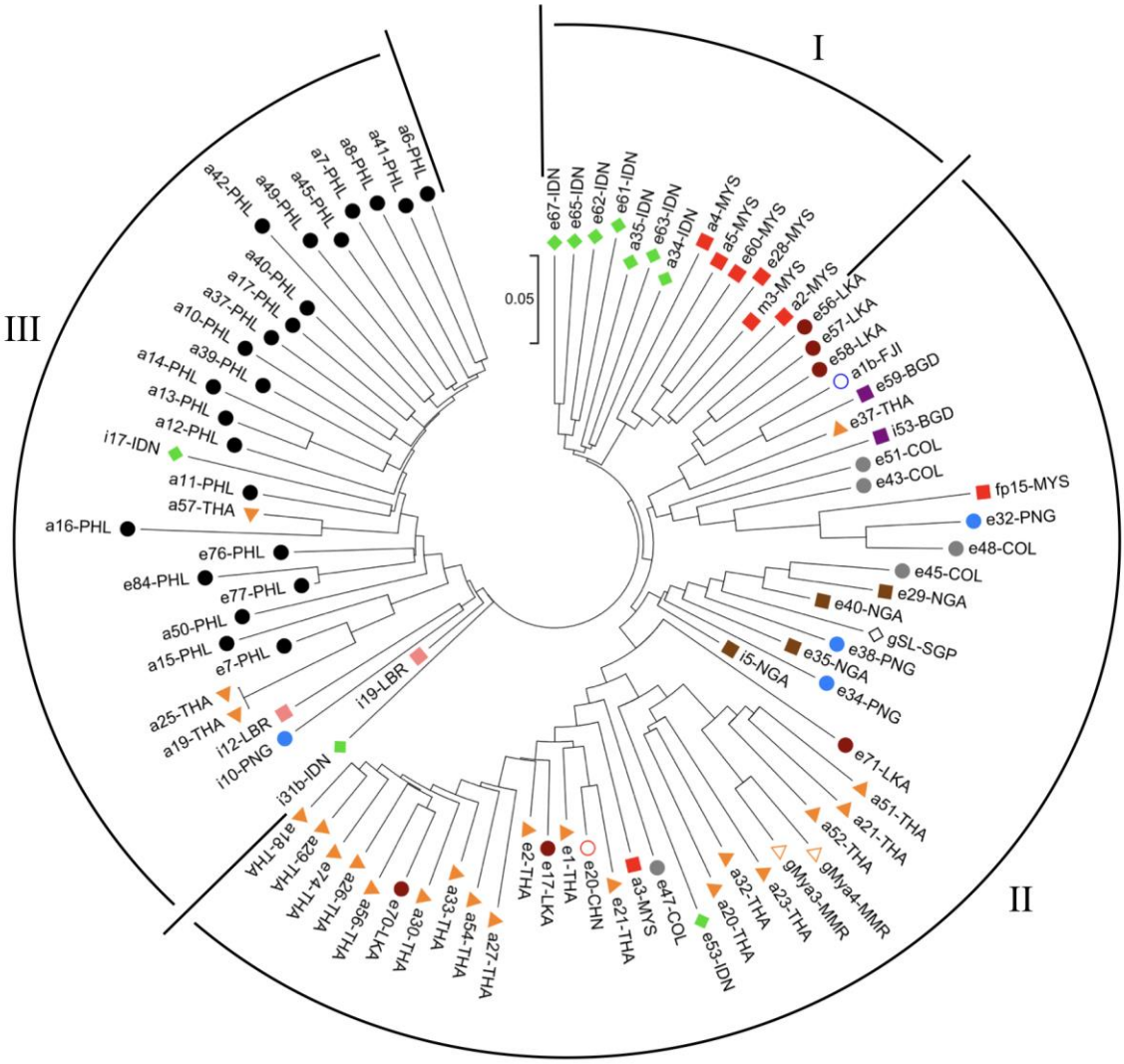


Figure a.3: NJ-tree from Nei's D coefficient genetic distance.

a.3 Additional graph of PCoA based on GD matrix, with labelled parental genotypes used in crosses.

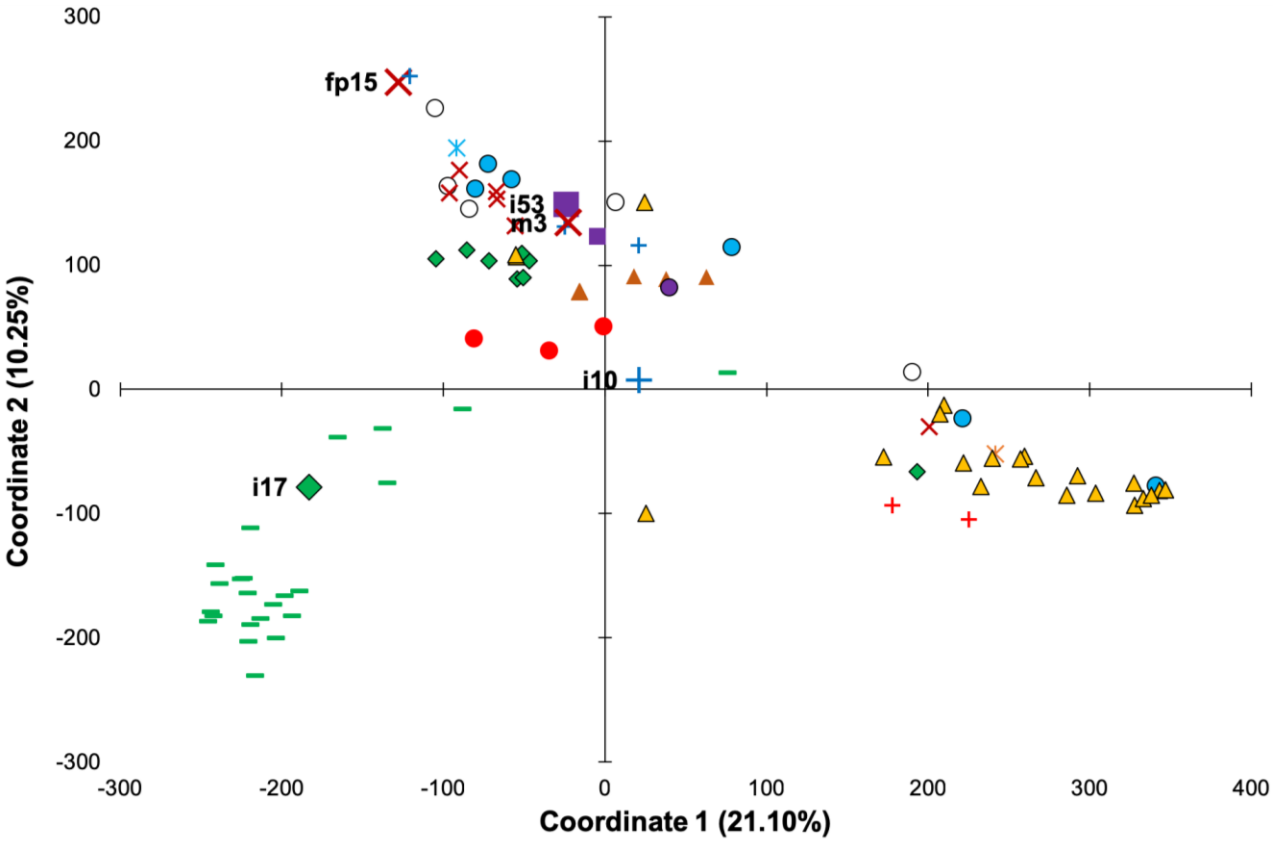


Figure a. 4: PCoA based on GD matrix (as in Figure 3.1 in the results section above), with labels for the parental genotypes used during crosses described in Chapter 3.

1 Background

In the context of tropical areas, winged bean has great potential to contribute towards food diversification, benefitting agricultural systems and supporting efforts towards food security. However, despite past research, its cultivation remains predominantly limited to a small scale. One of the factors that could hamper a wider integration of this crops into agriculture could be its growth habit, and the lack of research into how to modify it. In this regard, past research has overlooked the process of designing different ideotypes that could match the ways and the purposes for which this crop has been traditionally grown. Winged bean, in fact, remains a horticultural crop first, cultivated for its green pods and tuberous roots, in large part through the help of vertical supporting structures. Pods productivity has always been seen as the final aim to increase seed yield, without looking into the morphological traits underlying it and how much each of these contributed to the final number of pods per plant. A fresh approach focused on the linkages between morphology, development, and yield components, supported by the development of molecular technologies such as molecular markers and next-generation-sequencing, can lead to a better understanding of how winged bean grows. Eventually, this would focus towards the selection of individuals with desired plant architecture that could fit into specific growing conditions and for specific purposes.

In the present study traits related to plant architecture, development, and productivity of winged bean were analysed. The final aim has been to identify traits that could be part of potential ideotypes, primarily for improving pod yield when growing this crop on vertical supporting structures. This passed through the development of molecular tools, phenotyping of an F_2 population, construction of a genetic linkage map, and QTL analysis. Additionally, an initial genetic diversity analysis was carried out on a set of winged bean accessions from private and public sources, to identify the underlying population structure and identify genetic material of interest to include in future studies and improvement programmes.

2 Development of molecular markers in winged bean

Two types of molecular markers have been utilised in this project: Simple Sequence Repeats (SSR) developed using transcript data, and Single Nucleotide Polymorphism (SNP) developed through genotyping-by-sequencing technology.

The second chapter reported on the development and validation of a set of genic-SSR markers. Although the inclusion of more genotypes would have given a better estimate of the information that they can provide (e.g. PIC value), this has been partially done in a recent study using 53 accessions (Yang et al. 2018), the validated set has allowed us to have an initial idea of the relatedness between our accessions, and the residual heterozygosity. This has been crucial for the designing of the crossing scheme, as it directed us towards the use of single individuals rather than multiple individuals, and of the selection method used to advance F_1 to F_2 generation (single plant). The same markers also enabled us to validate the crosses facilitating the selection of the most promising among the four parental combinations, and the establishment of the F_2 population.

Regarding the SNPs, these have been developed through Diversity Array Technology sequencing (DArTseq) after the establishment of the F_2 population. DArTseq yielded 2467 SNPs between the parents from which the analysed F_2 was derived (M3 and FP15): of these, 1937 had no missing values across the segregating population, hence carrying a call rate value of 1. These SNPs were used to construct a dense genetic linkage map. No missing genotypic information also provides the best base for downstream QTL analysis. In addition, DArTseq data was generated for a large set of markers to carry out a genetic diversity analysis across 91 winged bean accessions.

3 Phenotypic analysis of the M3 × FP15 derived F_2 population, and related QTLs

One of the project aims was the analysis of plant architecture, development, and yield-related traits, through selection of parental genotypes for generation of hybrids, and assessment in field conditions of F_1 and selected F_2 populations. This could lead to an understanding of the genetic basis of the recorded traits, by combining phenotypic and genetic data generated from the same population.

Five parents were chosen based on an earlier characterisation (Wong, Q.N., unpublished) carried out before this project started and modified by seeds availability for the selected accessions. The local Malaysian accession retrieved from Malaysian Agriculture Research and Development Institute (MARDI) was chosen as a common male parent to be crossed with four different females. This is effectively progeny testing, and it was planned for the future improvement of locally adapted material. The crossing technique was based on the paper from Erskine and Bala (1976): this has led us to obtain pods in 53% of the crosses performed in controlled environment, a value comparable to the 57.9% in shade-house conditions reported by Erskine and Bala. Hybrids were then validated through SSR markers, at the beginning of the F_1 field assessment for the selection of the parental combinations to advance towards the F_2. Such selection, based on phenotypic analysis of traits like *stem length*, *internode length*, *leaves per plant*, and *pod length*, was partially affected by the low number of replicates used, and the loss of individuals in the field due to bacterial infection (Chapter 3, appendix). Nonetheless, the phenotypic data was sufficient to direct us towards the choice of the parental combination that showed the greatest contrast for these traits of interest: M3 × FP15.

Although the selection of a single individual, from which the F_2 was established, has limited the assessment of trait inheritance, it has allowed us to select for limited residual heterozygosity, in particular from M3, that could have affected phenotypic and genotypic analysis by introducing more genetic variation.

3.1 Plant architecture

From the analysed F_2 population, one of the main findings has been the stronger correlation of *stem length* (StL) with *internode length* (InL), rather than with the *number of nodes* (NoN). For internode elongation, major plant hormones appear involved: some, like auxin, gibberellins (GA), and strigolactone (SL) promoting elongation (Lester et al. 1997; Sasaki et al. 2002; de Saint Germain et al. 2013); whereas others, like abscisic acid and ethylene might have an inhibitory effect (Ross and Reid 1986). Altering winged bean *stem length* through *internode length* could therefore be a viable path to obtain individuals with shorter phenotypes. This could allow the use of lesser and shorter physical supports required to contain the plant growth during the entire cycle, rather than having the stem outgrowing the structure and start intertwining with neighbouring individuals. The QTL found in common between *internode length* and *stem length* (24841_65) might be

exploited to select individuals with opposite alleles, to carry out further assessment and elucidate which other genomic regions are involved.

Given the lack of correlation between *internode length* and *branch number per plant*, there could be potential to breed for shorter individuals without affecting lateral growth. PCA results appear to support this idea, as the two traits were on unrelated coordinates. Furthermore, studies in pea (*Pisum sativum*) *rms-1*, *rms-4*, and *Psbrc1* mutants have also showed the effect of strigolactone (SL) on internode length; in these SL-deficient or SL-insensitive individuals, might be independent from the SL-dependent bud outgrowth (de Saint Germain et al. 2013), again separating genetic mechanisms involved in internode elongation and branching behaviour. This could lead to either something akin to grape, green pea, tomato cultivation, or a system of complete automation without supports.

Branching, here analysed through the number of branches within the first 10 nodes (basal portion of the stem), is itself a mechanism controlled by multiple endogenous factors, and constant feedback received by environment and plant development (Leyser 2009; Teichmann and Muhr 2015). Across the F_2 population, *branch number per plant* (NoB) showed transgressive segregation, with additive mechanisms likely to be involved. This latter was postulated by Eagleton (1983), who interestingly reported also minimal environmental effect for this trait, in his replicated studies. In support of a mechanism involving more than one gene, the QTL analysis reported in chapter 5 showed 2 significant QTLs on the linkage groups 1 and 2 when *total branch number per plant* (NoBTOT) was analysed. Such QTLs explained a relatively large percentage of the observed variation (38.9%VE combined), but seemed to have opposite additive effects. For the selection of individuals with limited branch number, this will require the isolation of individuals carrying the allelic combinations from both parental genotypes, pushing towards reduced bud outgrowth. At the same time, it would be interesting to investigate possible combinations for the two genes. In an investigation on pea, *rms6* mutants showed increased branching from lower nodes, suggesting the presence of different mechanisms involved in basal and upper node branching (Rameau et al. 2002). The presence of such a mechanism in winged bean would perhaps allow us to prevent the formation of branches at upper nodes, which might not contribute much to final yield (see chapter 3, appendix) but causes management problems resulting from an increased level of intertwining with neighbouring individuals. The presence of individuals, among the parent M3 and the F_2, with branches departing from basal nodes overgrowing the main stem, perhaps points in this direction.

3.2 Plant architecture and yield

Although branch number has been suggested to affect pod number (Anonymous 1981) and total pod dry matter (Eagleton 1983), previous research has not looked specifically into the possible correlation between branch and pod number per plant, or the distribution of branches along the main stem.

Branch number per plant and *pods per plant* (PPP) showed a positive correlation, with medium effect size (r_s= .38). However, the mere increase in *branch number per plant* might not necessarily lead to a proportional increase in yield, whereas it could result in higher vegetative biomass. This might be due to the presence, at the top of each lateral branch, of nodes that have not developed mature inflorescences, therefore limiting the reproductive load in shorter branches, and result only in vegetative biomass. These individuals have shown that there is potential to increase pod productivity, without necessarily selecting for a higher NoB, which would instead lead to higher vegetative biomass, and denser canopies difficult to manage on the field. One trait to select for would be a higher *length of branch* (LoB): i.e. the selection towards individuals with fewer but longer lateral branches. This could decrease the impact of unreproductive nodes by reducing the number of younger growing stems (i.e. developing at later growing stages), reducing the amount of vegetative biomass, and hence resulting in higher harvest index. In support of this idea is the significant positive correlation found between *length of branch* and *pods per plant* (r_s= .44), even higher than the previous with *branch number per plant*. It is also possible that investigation into fertilization rates might provide more evidence for improving the rate of pod setting per inflorescence, and achieve better pod productivity while keeping lower vegetative biomass growth. For *length of branch* and *pods per plant*, the QTL analysis revealed two major putative QTLs co-localised on the same linkage group (LG3), as might be expected from traits showing significant correlation (Julier et al. 2007) such as these two. Further validation studies for these QTLs could be critical for the design of winged bean ideotypes to grow in different conditions, with different support systems, and with modulating the branching habit of the plant as key target. Increasing planting density is likely to benefit from individuals that have lower numbers of branches, although a careful assessment of competition between individuals and final canopy density has to be done, in order to calibrate the number of plants per unit area. Quite different could be the cultivation of winged bean on field borders, or in small scale as an additional cash crop. In these cases, with a focus on single individuals rather than entire field areas, it could be desirable to grow individuals with more branches, aiming at higher number of pods per single plant instead of per unit area.

Proposed ideotypes of winged bean are reported in Figure 1 below.

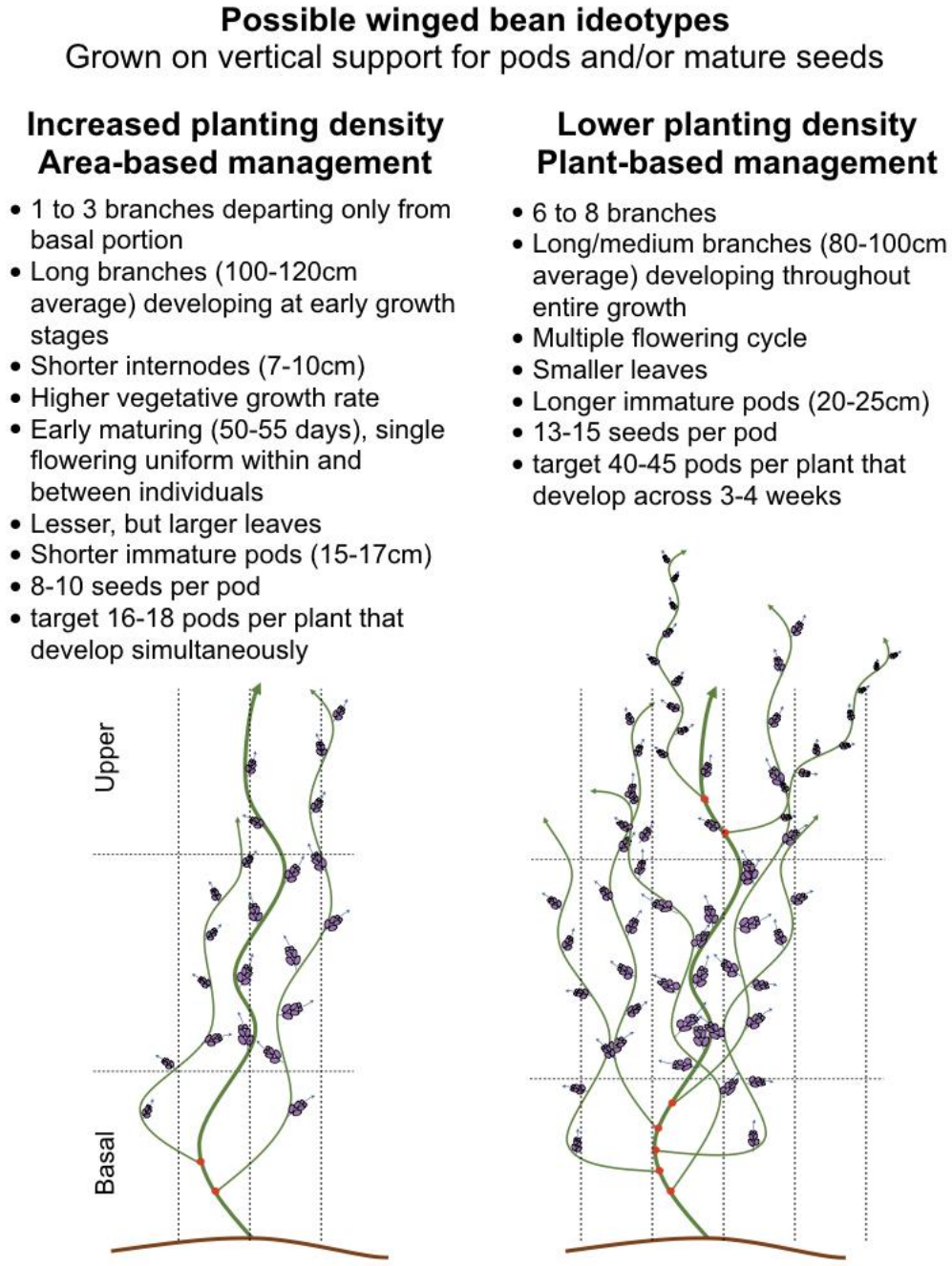


Figure 7.1: Proposed ideotypes for winged bean, for two different cropping systems.

Regarding other yield traits, *Dry pod length* (PoL) was significantly correlated with *seeds per pod* (SPP), which might encourage the selection for longer pods for both pod and seed yield. However, a couple of considerations need to be made: the first is in regard to the length of pods and the plant structure upon which they would grow. In fact, longer pods would probably require taller plants to allow adequate spacing between them.

Second, there could be two distinct mechanisms linking *seeds per pod*, and pod structure growth, as suggested by the presence of pod within the average length, but with far fewer seeds than in the average pod. In agreement with this second hypothesis, major significant and putative QTLs found for PoL and SPP did not co-localised on common linkage groups. Therefore, improvements for winged bean as a pulse crop should probably be looking into SPP individually, and investigate traits such as number of ovules per ovary and proportion of fertilised ovules. In this context, despite the two parents significantly differing for SPP, the analysis across this F_2 is likely to have been affected by the segregation distortion found during genetic linkage mapping, limiting the interpretation of the results.

3.3 Developmental traits and chlorophyll content

Concerning developmental traits, *days to first open flower* (DtF) appeared to be polygenic, with the F_2 population average closer to the later-flowering parent M3, but showing also transgressive segregation for higher and lower values. Early flowering individuals were shown to have also shorter *pod maturing time* (PMT), a trait for which a putative QTL was found on LG9. However, QTL analysis struggled to find loci linked to DtF, perhaps due to population size and the narrow difference between parental values (about 9 days, according to their mean values)

A significant correlation was found between SPAD measurements (*chlorophyll concentration*) and visual scoring of leaf greenness. Individuals from M3 parent and some F_2 individuals, showed pale green or yellow colour in their top leaves, suggesting that these leaves had significantly lower *chlorophyll concentrations*. A similar phenotype was observed in soybean crossings between wild-type and induced mutants carrying a single gene mutation (*cd1*) conferring pale-green leaves (Zhang et al. 2011): in contrast with that study, a 3:1 phenotypic ratio was not observed in the present F_2 population. Also, despite the lower chlorophyll content, growth and development of these individuals did not appear to be affected for any of the analysed trait during F_2 phenotyping.

Analysing a larger population is likely to improve the inferences made here, especially for traits related to yield. However, the correlations observed in this study could allow identification of critical traits to alter plant

architecture for the first time, and the design of ideotypes for different cultivation conditions. Additional traits, in particular fertilisation rate, have also been identified for the first time as being important and could be included into further studies aimed to improve yield, for both immature pod and seed productivity.

4 Segregation distortion

In chapter 4 is presented the construction of a genetic linkage map, based upon the segregation data from 221 F_2 individuals genotyped using DArTseq. Before this, Principal Coordinate Analysis (PCoA) and genetic distance analysis using the markers included for genetic linkage mapping revealed the F_2 as a group to be genetically closer to the male parent M3. Such an observation was later supported by the analysis of segregation patterns, which revealed about 73% of the markers to show significant deviation from the expected zygotic ratio (1:2:1=*aa*:*ab*:*bb*) (FDR confirming a total of 53%). Such high levels of segregation distortion might be expected in wide crosses, performed between genetically distant individuals, with underlying mechanisms acting at gametogenesis or at zygotic level, involving nuclear and/or cytological effects (Falconer and Mackay 1989; Lu et al. 2002; Castro et al. 2011; Li et al. 2011; Reflinur et al. 2014). In order to identify the possible causes within the present population, a few considerations have to be made:

1. In these F_1 and F_2, a low *seeds per pods* average was observed, while this phenomenon was not recorded across the other parental combinations. In particular, pods appeared to have a number of small undeveloped seeds (about 1mm diameter or less). This would suggest a mechanism of selection acting after fertilisation, perhaps when seed development progresses from maternal to zygotic control (Weber et al. 2005; Baroux and Grossniklaus 2015), hence at zygotic level. A possible cause could be incomplete genomic set within the zygote, due to abnormal recombination between inverted genomic regions.

2. While usually this amount of segregation distortion is observed in wide crosses, the genetic diversity analysis does not place FP15 and M3 relatively far apart. This male parent was actually, genetically, closer to FP15 than to other female parents such as TPt17 (see Chapter 3 appendix, and Chapter 6). Yet, combinations between M3 and those other genotypes showed a *seeds per pod* average value even higher than the better parent. Segregation distortion could perhaps suggest a cytological effect in the M3 × FP15 offspring, but given that almost the entire allelic distortion was in favour of the male alleles, this hypothesis seems less likely.

3. Segregation distortion regions (SDR), due to loci causing segregation distortion (SDL), were across all the linkage groups rather than localised on a few. This would exclude the presence of a few responsible SDLs, but rather the presence of a fundamental mechanisms acting on the whole genome, or multiple mechanisms acting within the same cross.

4. Allelic distortion was widespread among zygotic distorted markers (95%), usually with increase in homozygous *aa*. This would suggest that marker distortion could be due to selection among gametes.

5. All germinated seeds progressed through to flowering stage and produced seeds, indicating that selection did not act at zygotic level after seed germination. SDLs are hence not involved in post-germination zygotic fitness, perhaps pointing towards an underlying meiotic factor.

Distortion in favour of male alleles and the incidence of allelic distortion would suggest the presence of selection at gametic level. However, the presence of aborted seeds in pods of F_1 and F_2 individuals, might indicate a post-fertilization selection. Identifying the cause of segregation distortion would then require further studies, complicated by the possible presence of multiple mechanisms, nuclear and cytological, that can act within the same cross and with multiple mechanisms each, as has been shown in rice (Reflinur et al. 2014; Bravo Núñez et al. 2018) and *Medicago sativa* (Li et al. 2011). A first step however could be the analysis of reciprocal crosses and backcrosses, in order to narrow the number of possible causes underlying the observed segregation distortion. As alternative, analysing a different segregating population obtained from a cross with the same male M3 parent would give an indication of whether this carries SDLs that generate the observed segregation distortion.

For the present study, segregation distortion had to be considered in order to obtain a genetic linkage map. Distorted loci can, in fact, alter the correct marker ordering and distance calculations (Li et al. 2010; Zhou et al. 2015). For such reasons, we adopted a two-step approach: first we excluded all distorted makers, in order to obtain a framework map with correct order and distance of non-distorted markers. Then, distorted markers were re-introduced, at each round checking that the undistorted framework markers remained in the same order. To date, there is no standard way to carry out construction of genetic linkage map with strongly distorted

markers, but the approach implemented here should allow us to avoid as much as possible errors in the map order (Xian-Liang et al. 2006; Li et al. 2010). This has allowed QTL analysis to be performed based on a complete map of 1171.6cM length, and 395 SNPs (3cM/SNP; 36% of distorted markers).

Regarding the effect of segregation distortion in QTL analysis (chapter 5), past literature has not shown a complete consensus regarding this subject. This is probably because the effect of a distorted marker in QTL analysis varies according to map density, population size, proximity of distorted marker to QTL, segregation distortion direction, missing information, and phenotypic variance explained by the QTL to be detected (Hackett and Broadfoot 2003; Xu 2008; Zhang et al. 2010; Li et al. 2011; Wen et al. 2013; Cui et al. 2015). What appears from the present investigation is that the power to detect QTL was possibly reduced by an increased Genome Wide LOD threshold and by the small population size, hampering the ability to detect QTLs with minor effects, or the correct estimation of the explained phenotypic variance. However, the combined use of non-parametric marker-trait analysis, Interval Mapping, a 2-LOD support interval, repeated analysis with the framework map, and the high LOD threshold itself, should support the validity of the QTLs detected in our analysis.

5 Genetic diversity analysis

A diversity analysis was presented in chapter 6, in order to assess the genetic diversity, and infer the population structure within germplasm from both private and public collections. The aim was to identify more efficiently genetic material to include in future winged bean improvement programmes, to harness as much diversity as possible.

One of the issues faced in this kind of analysis, as mentioned in chapter 1, is the amount of available and reliable information for winged bean accessions: mislabelling, relabelling, and loss of seed passport information during international seed trials have hampered the possibility to establish a coherent seed collection. In addition, the movement of germplasm ‘prehistorically’ is unknown. These could be part of the reason why past research has not reconciled geographical origin and observed groups, based on molecular techniques like Inter-Simple Sequence Repeat (ISSR) (Chen et al. 2015), combined ISSR and Random Amplified Polymorphic DNA (RAPD) (Mohanty et al. 2013), and Simple Sequence Repeat (SSR) (Yang et

al. 2018), although those investigations also did not use an extensive set of markers and accessions, while there are also concerns with methods such as ISSR and RAPD.

The present investigation utilised 5891 SNPs generated through DArTseq on 91 winged bean accessions, coming from 15 countries spread across 4 continents. In particular, for material from Philippines, Thailand, Indonesia, and Malaysia, it was possible to identify potential individuals belonging to these country groups that could be advanced for phenotypic characterisation and breeding programmes. Although further analyses are needed to clarify the population structure in our germplasm, PCoA and NJ-tree found a degree of consensus between geographical and genetic groups. This is in contrast with the previous studies mentioned earlier, potentially due to the use of numerous accessions representing single countries (in particular, Philippines and Thailand) and our larger set of markers. It should be considered that the nature of the markers themselves (SNPs rather than SSRs or RAPDs) could yield different results upon the same set of accessions, given the differences between mechanisms through which each marker evolves (single point mutation, rather than indels during DNA replication), the genomic regions they might represent (active or inactive) and the reliability of the marker system. In rice, for example, it has been shown how SNP were able to achieve better resolution in investigating population structure compared to SSRs, which nonetheless seemed more efficient (Singh et al. 2013). Having generated our SNPs through DArTseq method, which should target predominantly active genomic regions (Baloch et al. 2017), would allow the present analysis to look into variation across functional regions underlying phenotypic traits. These would give our selection process the advantage of identifying more efficiently the desired phenotypic variation to include into future improvement programmes, as well as a significantly deeper sampling of the genetic structure of the accessions

6 Summary of the progress achieved in this project

This project was aimed to generate molecular tools and investigate agronomic traits in order to support breeding improvement programmes in winged bean.

The main results could be summarised as follow:

1. Review of past efforts and identification of breeding targets in winged bean (Chapter 1);
2. Development of the first validated set of genic-SSR markers for genetic diversity, and breeding system analysis in winged bean (Chapter 2);

3. Generation of four F_1 hybrid populations, validated through molecular markers and characterised under field conditions (Chapter 3);
4. Investigation of basic traits determining plant architecture, and their correlation with yield components in one cross of winged bean. Identification of traits and ideotypes to improve pod yield in winged bean grown on a vertical trellis structure (Chapter 3);
5. Construction of the first winged bean genetic linkage map, based on SNPs discovered through DArTseq method across an F_2 mapping population (Chapter 4);
6. Identification of segregation distortion, and potential breeding barriers in winged bean (Chapter 4);
7. QTL mapping of agronomic traits, and identification of 5 segregating QTLs and 8 putative QTLs distributed across 9 traits, including main targets such as *branch number per plant*, *length of branch*, *dry pod length*, and *chlorophyll content* (Chapter 5);
8. Assessing the genetic diversity of a germplasm composed by 91 accessions, including publicly retrievable ones, and effective selection of material to be included in future winged bean improvement programmes (Chapter 6).

7 Future work

The results obtained here pave the way for further studies.

These could begin by taking advantage of the QTLs identified for traits such as *internode length*, *branch number per plant*, *length of branch*, and *pods per plant*, to select for contrasting traits and further characterise the obtained plant architectures. Phenotyping of these individuals could also include additional traits like *harvest index*. In addition to this, a further thorough analysis will be required to improve winged bean pod and seed productivity, by looking into the number of inflorescences produced and the number of pods harvested. It is also worth considering that in the present study *pods per plant* had a higher correlation with the number of branches within the first 10 nodes (NoB), than with the total number of branches along the entire stem (NoBTOT). This could suggest that branches departing at higher nodes might contribute less than branches departing at lower ones. However, further phenotyping and partitioning of *pods per plant* between branches departing at different nodes is needed to confirm such a hypothesis.

Another possible avenue of research could be the characterisation of photosynthetic rates and biomass growth in individuals with contrasting chlorophyll concentrations, selected either through visual scoring or the significant QTLs revealed in this study. Under different growing conditions it is possible that such a trait might impact on growth and yield, thus making it a necessary target to be included into breeding programmes.

Overall, the current findings need to be validated through replicated trails across different populations, locations, and seasons. This will allow an estimation of environmental effects, validation of current QTLs, with an idea of the marker potential for Marker-Assisted Selection (MAS), and, ideally, finding additional loci involved in polygenic traits like *internode length*, or *branch number per plant*. Fine mapping of the QTLs could also be achieved through larger population sizes, and by establishing mapping populations such as Recombinant Inbred Lines (RILs). Such populations will support a more efficient selection of desired phenotypes, and the development of designed ideotypes.

In relation to the genetic linkage map, this could be used to carry out an initial comparative analysis for the identification of synthenic regions in other well studies species, and the isolation of candidate genes for specific traits.

Further investigation could also look into the causes for the observed segregation distortion, by genotyping one of the available F_2 obtained through crosses with the common parent M3. In such analyses, levels of segregation distortion and its direction would allow us to understand which parent and what underlying mechanisms might have been involved in the M3 × FP15 cross.

Regarding targets that have not been part of this project, nodulation efficiency and nitrogen fixation activity deserve further attention, given the potential of winged bean in contributing to agricultural system resilience in low-input conditions.

Finally, if winged bean is to reach its full potential to contribute to human nutrition, the establishment of breeding programmes towards this aim will be critical. In this context, it will be necessary to screen germplasm for nutritional and antinutritional contents in pods, seeds, and tubers, while making use of the research carried out so far. Such efforts, as much as those spent in this project, will be crucial to fulfil the potential of winged bean, and contribute towards food security.

References

Anonymous (1981) The Winged Bean Flyer (3 No.2). International Documentation Centre for the Winged Bean Agricultural Information Bank of Asia, Laguna, Philippines

Baloch FS, Alsaleh A, Shahid MQ, et al (2017) A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. PLoS One 12:1–18. doi: 10.1371/journal.pone.0167821

Baroux C, Grossniklaus U (2015) The Maternal-to-Zygotic Transition in Flowering Plants. Evidence, Mechanisms, and Plasticity. Curr Top Dev Biol 113:351–371. doi: 10.1016/bs.ctdb.2015.06.005
Bravo Núñez MA, Nuckolls NL, Zanders SE (2018) Genetic Villains: Killer Meiotic Drivers. Trends Genet 34:. doi: 10.1016/j.tig.2018.02.003
Castro P, Rubio J, Cabrera A, et al (2011) A segregation distortion locus located on linkage group 4 of the chickpea genetic map. Euphytica 179:515–523. doi: 10.1007/s10681-011-0356-7
Chen D, Yi X, Yang H, et al (2015) Genetic diversity evaluation of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) using inter-simple sequence repeat (ISSR). Genet Resour Crop Evol 62:823–828. doi: 10.1007/s10722-015-0261-3
Cui Y, Zhang F, Xu J, et al (2015) Mapping quantitative trait loci in selected breeding populations: A segregation distortion approach. Heredity (Edinb) 115:538–546. doi: 10.1038/hdy.2015.56
de Saint Germain A, Ligerot Y, Dun EA, et al (2013) Strigolactones Stimulate Internode Elongation Independently of Gibberellins. Plant Physiol 163:1012–1025. doi: 10.1104/pp.113.220541
Eagleton GE (1983) Evaluation of genetic resources in the winged bean (*Psophocarpus tetragonolobus* (L.) DC.) and their utilisation in the development of cultivars for higher latitudes. University of Western Australia
Erskine W, Bala AA (1976) Crossing technique in winged bean. Trop. Grain Legum. Bullettine 6:32–35
Falconer DS, Mackay TF (1989) Introduction to Quantitative Genetics Third Edition
Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity (Edinb) 90:33–38. doi: 10.1038/sj.hdy.6800173
Julier B, Huguet T, Chardon F, et al (2007) Identification of quantitative trait loci influencing aerial morphogenesis in the model legume Medicago truncatula. Theor Appl Genet 114:1391–1406. doi: 10.1007/s00122-007-0525-1
Lester DR, Ross JJ, Davies PJ, Reid JB (1997) Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell 9:1435–43. doi: 10.1105/tpc.9.8.1435
Leyser O (2009) The control of shoot branching: An example of plant information processing. Plant, Cell Environ 32:694–703. doi: 10.1111/j.1365-3040.2009.01930.x
Li H, Kilian A, Zhou M, et al (2010) Construction of a high-density composite map and comparative mapping of segregation distortion regions in barley. Mol Genet Genomics 284:319–331. doi: 10.1007/s00438-010-0570-3
Li X, Wang X, Wei Y, Brummer EC (2011) Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. Theor Appl Genet 123:667–679. doi: 10.1007/s00122-011-1617-5
Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. TAG Theor Appl Genet 105:622–628. doi: 10.1007/s00122-002-0970-9
Mohanty CS, Verma S, Singh V, et al (2013) Characterization of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) based on molecular , chemical and physiological parameters. Am J Mol Biol 3:187–197. doi: 10.4236/ajmb.2013.34025
Rameau C, Murfet IC, Laucou V, et al (2002) Pea rms6 mutants exhibit increased basal branching. Physiol Plant 115:458–467. doi: 10.1034/j.1399-3054.2002.1150316.x
Reflinur, Kim B, Jang SM, et al (2014) Analysis of segregation distortion and its relationship to hybrid barriers in rice. Rice 7:3. doi: 10.1186/s12284-014-0003-8
Ross JJ, Reid JB (1986) Internode length in Pisum. The involvement of ethylene with the gibberellin-insensitive erectoides phenotype. Physiol Plant. doi: 10.1111/j.1399-3054.1986.tb05076.x
Sasaki A, Ashikari M, Ueguchi-Tanaka M, et al (2002) Green revolution: A mutant gibberellin-synthesis gene in rice. Nature 416:701–702. doi: 10.1038/416701a
Singh NK, Choudhury DR, Singh AK, et al (2013) Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. PLoS One 8:e84136. doi: 10.1371/journal.pone.0084136
Teichmann T, Muhr M (2015) Shaping plant architecture. Front Plant Sci 6:1–18. doi: 10.3389/fpls.2015.00233
Weber H, Borisjuk L, Wobus U (2005) Molecular Physiology of Legume Seed Development. Annu Rev Plant Biol 56:253–279. doi: 10.1146/annurev.arplant.56.032604.144201
Wen J, Can VT, Zhang YM (2013) Multi-QTL mapping for quantitative traits using distorted markers. Mol Breed 31:395–404. doi: 10.1007/s11032-012-9797-5
Xian-Liang S, Xue-Zhen S, Tian-Zhen Z (2006) Segregation distortion and its effect on genetic mapping in plants. Chinese J Agric Biotechnol 3:163–169. doi: 10.1079/CJB2006110
Xu S (2008) Quantitative Trait Locus Mapping Can Benefit From Segregation Distortion. Genetics 180:2201–2208. doi: 10.1534/genetics.108.090688
Yang S, Grall A, Chapman MA (2018) Origin and diversification of winged bean (*Psophocarpus tetragonolobus* (L.) DC.), a multipurpose underutilized legume. Am J Bot 105:888–897. doi: 10.1002/ajb2.1093
Zhang H, Zhang D, Han S, et al (2011) Identification and gene mapping of a soybean chlorophyll-deficient mutant. Plant Breed 130:133–138. doi: 10.1111/j.1439-0523.2010.01844.x
Zhang L, Wang S, Li H, et al (2010) Effects of missing marker and segregation distortion on QTL mapping in F_2 populations. Theor Appl Genet 121:1071–1082. doi: 10.1007/s00122-010-1372-z

Zhou W, Tang Z, Hou J, et al (2015) Genetic Map Construction and Detection of Genetic Loci Underlying Segregation Distortion in an Intraspecific Cross of Populus deltoides. PLoS One 10:e0126077. doi: 10.1371/journal.pone.0126077