Unione europea
Fondo sociale europeo

REPUBBLICA ITALIANA

REGIONE AUTÒNOMA DE SARDIGNA
REGIONE AUTONOMA DELLA SARDEGNA

FSE 2007-2013
POR SARDEGNA
obiettivo competitività regionale e occupazione

SCUOLA DI

Scienze e

A.D. MDLXII

UNIVERSITÀ DEGLI STUDI DI SASSARI
DOTTORATO DI RICERCA

Biotecnologie

dei Sistemi Agrari e Forestali

e delle Produzioni Alimentari

RVRALIS ET SILVESTRIS SAPIENTIA

Indirizzo Produttività delle piante coltivate

Ciclo XXVIII

Two case studies on plant genetic diversity: comparison of two historic collections of barley landraces and characterization of a common bean segregant population for domestication traits

dr.ssa Maria Leonarda Murgia

Direttore della Scuola	Prof. Antonello Cannas
Referente di Indirizzo	Prof.ssa Rosella Motzo
Docente Guida	Dr. Domenico Rau
Tutor	Prof.ssa Giovanna Attene

La presente tesi è stata prodotta durante la frequenza del corso di dottorato in "Scienze e Biotecnologie dei Sistemi Agrari e Forestali e delle Produzioni Alimentari" dell'Università degli Studi di Sassari, a.a. 2014/2015 - XXVIII ciclo, con il supporto, relativo solo agli ultimi cinque mesi di attività, di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività l.3.1 "Finanziamento di corsi di dottorato finalizzati alla formazione di capitale umano altamente specializzato, in particolare per i settori dell'ICT, delle nanotecnologie e delle biotecnologie, dell'energia e dello sviluppo sostenibile, dell'agroalimentare e dei materiali tradizionali".

Maria Leonarda Murgia gratefully acknowledges Sardinia Regional Government for the partial financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.)

CONTENTS

ABSTRACT

In this study we studied the change in the time of population of barley and in second chapter we analisys the traits of Syndrome Domestication in common bean.

Chapter 1

Comparison two collections of landraces Barley from N.I. Vavilov at '90s.

INTRODUCTION

The relevance of landraces

Plant genetic resources (species and varieties) are sources of genes directly usable or potentially exploitable for the genetic improvement of cultivated plants. Among these, the wild progenitors and local varieties (landraces) occupy a role of particular importance.

The wild relatives of cultivated plants often show high tolerance against abiotic stress or resistance to biotic agents (Hajjar and Hodgkin, 2007; Atwell et al., 2013). Examples are those of tolerance to high or low temperatures (Grossi et al., 1998; Lei et al., 2012), at high salt concentrations (Hand and Takeda, 1998; Nevo and Chen, 2010), water deficit (Nevo and Chen, 2010; Lakew et al., 2013; Hikmet et al., 2013; Monneveux et al., 2013) as well as tolerance to edaphic conditions not optimal or limiting due to abnormal pH and high concentrations of heavy metals (Baker et al., 1994; Watanabe and Osaki, 2002; Ganança et al., 2007). In the second case, we can include mainly by the resistance to fungal pathogens, bacterial and viral (Muehlbauer et al., 1994; Rao et al., 2003; Cooper et al., 2006; Jarosz and Davelos, 1995).

The landraces (local varieties) are populations that have evolved in subsistence agriculture in response to selection by man as well as the environment. They are the product of what has been defined as "evolutionarily sustainable production" (Brown et al., 2000). Traditional farmers have retained populations characterized by high genetic variability due to the existence of network interchange of seeds among farmers ("migration") and as a result of anthropic selection (for certain customs and traditions) and natural selection (Harlan 1975; Brush 2000). In general, the landraces may have adaptive characters for the agro-ecosystem in which they have evolved and are interesting sources of genes for the improvement of agricultural production both in quantity (especially in marginal environments) and quality. Often, the landraces appear as heterogeneous populations (multi-genotypic) and not as individual genotypes (Brown, 2000; Ceccarelli and Grando, 2000). This raises a question about the role of the variability observed within populations of landraces; whether it is only the result of demographic processes occurring over time or is rather important to determine local adaptation (Demissie and Bjornstad 1997; Jaradat and Shahid, 2006). Several

experiments conducted with landraces suggest that this variability could have a "buffer effect" against environmental changes; the variability of the population could be a tool to ensure that the population (at least with some of its individuals) is "in phase" with the environmental conditions. This would allow the adaptation in environments characterized by strong unpredictability of weather conditions (Ceccarelli and Grando, 2000). Local varieties can be successfully used for the identification and mapping of the genes that control the adaptive variation in cultivated species (Vigouroux et al., 2002; Mazzucato et al., 2008; Comadran et al., 2009). However, a deeper understanding of the role of genetic variability of landraces is an objective of primary importance not only to exploit its potential for genetic improvement, but also for the correct implementation of strategies of conservation of germplasm. In fact, the preservation of genetic diversity directly "on farm" or "in situ", has become an important topic (Brown, 2000).This depends on the belief that the intraspecific variability is related to the selection and adaptation local and allows the "answer" to new diseases or adversity, stress edaphic and climatic variations (Frankel et al., 1995; Brown, 2000). The study of the evolution of the populations of landraces can be made by analysing the change in allele frequencies in space (in populations grown in different locations with contrasting soil and climatic conditions), is studying the change in allele frequencies over time at the level of individual populations and / or meta-population (= system populations with extinction-recolonization connected by gene flow) (Lister et al., 2009; Leino et al., 2010, 2012).

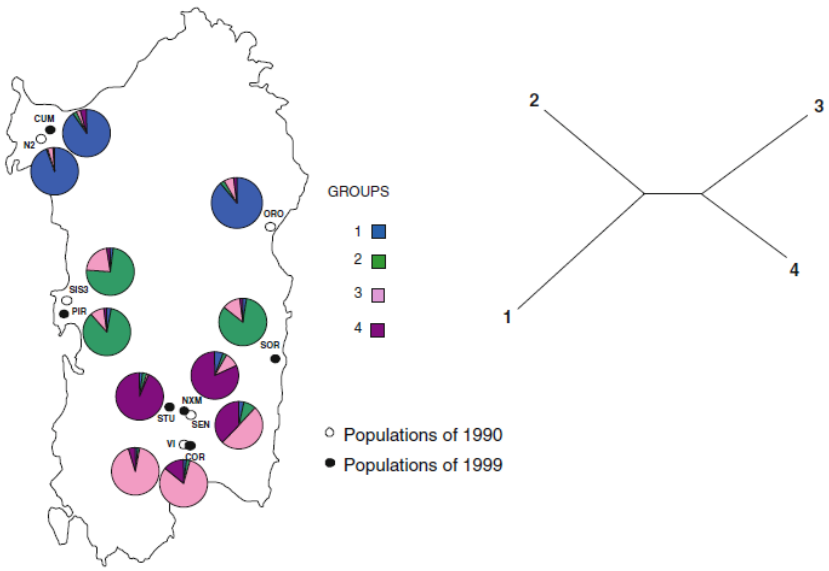
Barley (*Hordeum vulgare L.*) in Sardinia

Barley is a plant known to man since ancient ages; was already cultivated in the Middle East in the seventh millennium BC and from there it spread all over the world. In Italy occupies a cultivated area amounted to 360,000 hectares, with a production of 1.4 million tonnes. In areas where the climate is not suitable for the cultivation of wheat, barley was an important food for humans as a source of carbohydrates and protein, in many countries in the developing world it is still even. Instead, in most developed countries the grain of barley is the main destination (85-90%) in the animal feed livestock and secondarily (10-15%) of the malt industry. Cultivated since ancient times, in the past barley represented a significant food source in Sardinia. Today barley represents the most cultivated cereal after wheat occupying an area of 10,418 ha in 2011 (www.sardegna.coldiretti.it) In many areas there is a widespread cultivation of a local landrace polistic known as "S'orgiu sardu" (Attene et al., 1996)

This is used both for the production of fresh forage through the direct grazing (sheep) and for the production of grain. Local varieties of barley were grown in different environments at different latitudes, soil types and with different cultural practices and this may have allowed the development of locally adapted populations of barley. Section of Agronomy, Crop Science and Genetics, Department of Agriculture, University of Sassari, in the summer of 1990 (Attene et al., 1996) has performed an initial collection of 27 populations of barley in different growing areas of Sardinia. They were taken into account farms that for at least 30 years had used seeds of own production. From each field were taken at random 100 ears, whose kernels were multiplied in single ear, at the Experimental Farm of the Department of Agriculture. The analysis of 240 lines extracted from 12 of the 27 local populations collected in Sardinia in 1990 showed the existence of high genetic variability for many characters such as: date of ear emergence, plant height, kernel weight, grain yield, biomass total and harvest index (Papa, 1993).The collection of populations of Sardinian barley was repeated in 1999, visiting the same areas sampled in 1990 (Rau, 2002). The comparison of the collections of 1990 and 1999 using microsatellite molecular markers (SSR, Simple Sequence Repeats) and based on retrotransposons (SSAP; Simple Sequence Amplified Polymorphism) not showed neither changes in the levels or in the structure of genetic diversity, indicating that farmers implement a system of conservation "in situ" capable of preserving the genetic identity of the populations, even in the absence of specific conservation projects for these plant materials (Rodriguez et al., 2012; Bellucci et al., 2013). The analysis of the spatial structure of the genetic variability showed a level of genetic divergence of populations moderate-low, measured by a fixation index Wright (Fst; 1931) 0.18 (Rodriguez et al., 2012). This also clears up that 18% of molecular variance (the allele frequencies) is explained by the differences between populations and 82% of the variance is observed within the population. This finding is in line with that obtained with other types of molecular marker such as isozymes and RAPD (Papa et al., 1998) and SSR (Bellucci et al., 2013) with Fst values of 0.16, 0.11 and 0.14, respectively. The high proportion of variability within this population is also confirmed by the fact that, with a relatively small number of molecular marker (134 SSAP or SSR 12; Rodriguez et al., 2012; Bellucci et al., 2013) all individuals analyzed (over 350) are a unique genetic profile, two individuals are still genetically distinguishable (Rodriguez et al., 2012; Bellucci et al., 2013). Moreover, it was highlighted a well-defined genetic structure of populations, with four main genetic groups and a clear structure according to the geographical site of origin (Figure 1.1, Rodriguez et al., 2012). In particular, the genetic

variability is mainly oriented along the north-south axis of the island and there is a structure known as hierarchical island (or "archipelago"), in which peoples geographically close (North, the center, the center-south and south) tend to be more similar than that to other populations. This complex may indicate the existence of exchange of seed among farmers in network mostly local, and also suggests the existence of diversifying selection (wonders of local adaptation; Rodriguez et al., 2012)

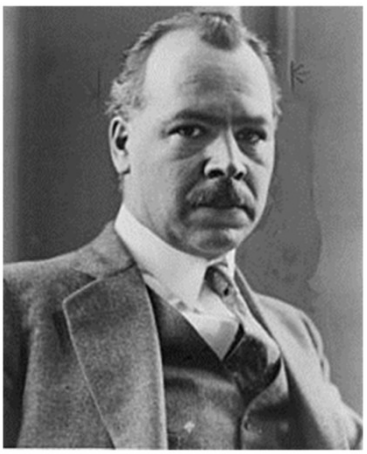
Figure 1.1 – Population structure of Sardinian landraces (From Rodriguez et al. 2012)



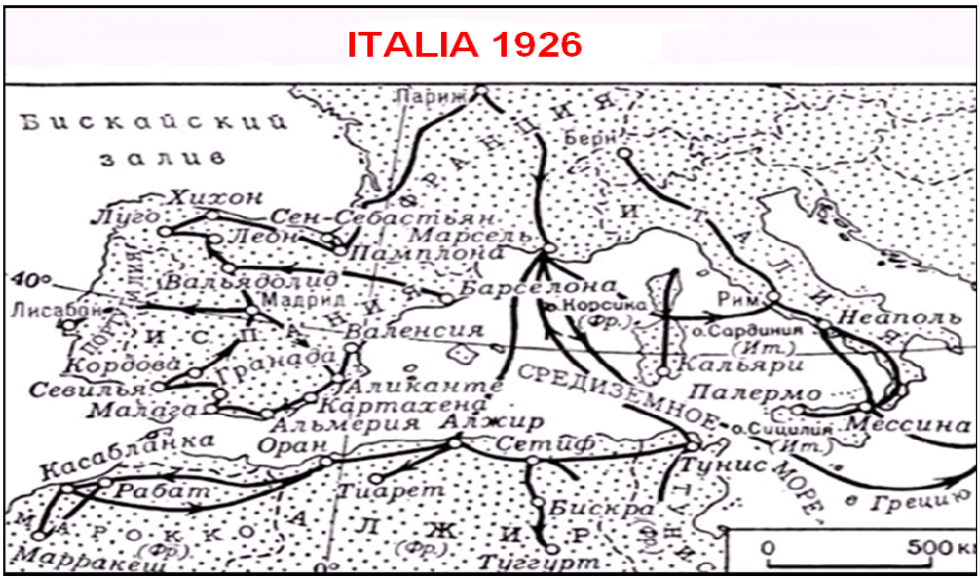
The comparison between the degree of divergence of populations for neutral molecular markers (Fst) and quantitative traits (Qst) about it can be particularly informative. In fact, according to the theoretical models (Merilä and Crnokrak, 2001) is expected that in a survey neutral Fst = Qst, in a survey in which balancing selection Qst <Fst and in a survey in which divergent selection Qst> Fst. For the Sardinian population was calculated Qst equal 0.40 (Rau, unpublished; Papa, unpublished), the degree of divergence medium for quantitative traits was 2-3 times higher than the Fst, supporting the hypothesis of selection for local adaptation in different environments. Finally, the analysis of linkage disequilibrium showed rather low levels (13% of pairs of locus, with P <0.01) and a decay with the distance of the map. The analysis of multilocus LD showed that genetic drift and founder effect have played an important role in shaping the genetic diversity present in this landrace, although the

epistatic selection may also have played a role in this system. Overall, the results suggest that these populations could be used profitably for gene mapping by association mapping (Rodriguez et al., 2012).

Nikolaj I. Vavilov and Vavilov Insitute for Plant Genetic Resources (VIR)



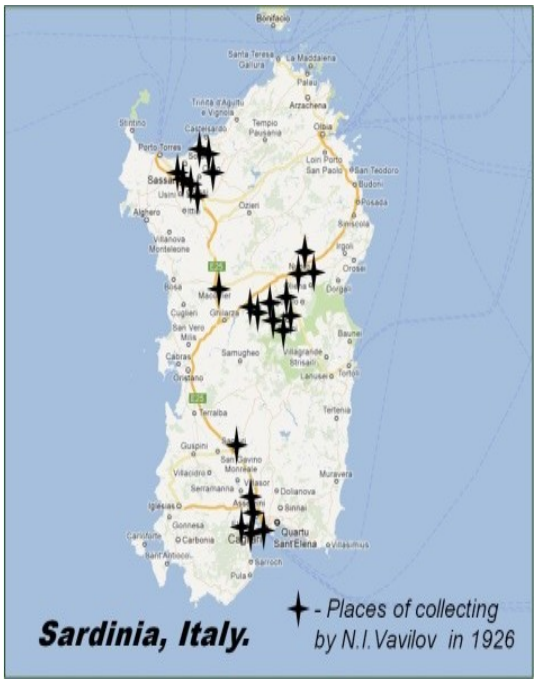
The Russian Nikolai Ivanovich Vavilov (1887-1943) was a pioneer in studies on biodiversity vegetable. In 1920 he was called by Lenin to direct the then Soviet Institute of Plant Industry Agricultural now become NI Vavilov Research Institute of Plant Industry, St. Petersburg, Russia (VIR). He directed this institute from 1920 to 1940. Vavilov considered the central role of genetic resources in breeding. The experience accumulated during a multitude of travel observation and study of the variability of the species and breeds of agricultural plants or their "kin" wild in different areas of the world led him to theorize the existence of the centers of origin of the plants grown (where species domesticated have originated and their evolutionary history).The important insight of Vavilov consisted in assuming that the observed differences were to be related to the length of the evolutionary history of the species in a given location. According to the reasoning of Vavilov, the place with greater variability is the one that also has a link with the oldest species in question that is the place where the species has had more time to diversify and from which possibly has subsequently expanded to colonize other regions. Although more recent studies have not always confirmed for the various species of centers of origin identified by Vavilov, he was responsible for the brilliant idea and incredibly far-sighted to center its scientific work on the identification of the centers of origin and domestication of agricultural plants an industry that is allowing you to get and, promises to do to get in the next few years, enormous advances in the genetic improvement of crop plants (Olsen and Wendel, 2012, 2013). In the period 1926-1927 Vavilov and collaborators begin an expedition in the countries of the Mediterranean dock (including, Syria, Palestine, Transjordan, Greece, Cyprus, Crete, Italy, France, Spain, Portugal, Morocco, Algeria, Tunisia and Egypt) and in Abyssinia (Djibouti, Addis Ababa, the Nile valley, lake Tsana), Eritrea (Massawa) and Yemen (Hodeida, Jidda, Hedjas). From



Cyprus, as evidenced by texts received, Vavilov headed for Italy who had previously visited many times. He visited Sicily, along the road from Palermo to Catania and performed also an exploration of Sardinia.

Vavilov was convinced that the exploration of Italy and its islands were very important to fully understand the Mediterranean culture (Loskutov, 1993).On the information contained in the archives of the VIR, Vavilov was not conducted exploration in Sardinia personally, but under his direction and supervision, by a close collaborator the Dr. Gaissinsky.The collection of these materials would have taken place during a trip from South to North Island with a through even in inland areas. During this trip is been visited several sites and collected accessions of many species of agricultural interest.

Figure 1.2– Sites visited in Sardinia (based on archive data of (VIR).



Although it know the places where Gaissinsky stopped and he collected plant material (Figure 1.2), to date, from the information available it was not possible to deduce the exact site of origin of each accession although it is fair to assume that these are from sites distributed throughout the region since as noted by Gaissinsky: "Barley. It is cultivated Throughout Sardinia, less in Sassari. Samples are abundant". However, it is important to note that the sampling sites of Gaissinsky, concern mostly the same areas in which were sampled the populations of landraces of Hordeum vulgare in 1990 and 1999 .Figure….

The existence of landraces of barley collected in the 1926 is a great opportunity : it might be possible to study the time variation of allele frequencies comparing the population of 1926 and those collected in 1990 and 1999, a distance of about 70 years. This would allow to understand what was the role demography and natural selection in " shaping" the genetic diversity of landraces,also allowing to better understand the genetic basis of adaptation of barley.

OBJECTIVES

In this chapter are presented the results of a pilot study where the aim of wigh was to compared population of landrcaes collected in different ages (1926 and 1990-1999) in Sardinia.

To observe that the population of landraces of 1926, 1990 and 1999 share the same gene pool would suggest a close relationship between populations of different ages. This would allow to account populations of 1926 as a progenitors of Sardinian landraces stressing the antiquity of the same and opening the possibility in the next future to conduct analysis of allele frequency variation in time. This could allow a better understanding of the genetic bases of plant adaptation to the agro-ecosystem.

Materials and Methods

Phenotypic Analysis

The first phenotypic analysis was performed on eight populations collected in 1926. A the end of February 2013, 40 seeds were sown for each of the eight population. After about 15 days of growth, stage of third leaf by the plants, second leaf has been used for the extraction DNA. After a further period of recovery of the seedlings plants were transplanted in the field. For each population was counted the number of number of plants that did flower. Then, the following characters were measured: date of flowering, plants height (height of the culm in cm), spike length (cm), awns length (cm), awns length/spike length ratio, number of seeds per plant, weight of 1000 seeds (g) and grain yield per plant (g). The flowering date was determined for each population as the percentage of plants come to earing in one survey on 10 June 2013.

Molecular Analysis

For molecular analysis we used 12 microsatellites markers already successful in describing the populations structure of Sardinian barley landraces (Bellucci et al., 2013). Primer sequences, DNA extraction and PCR conditions were as described in Bellucci et al. (2014).

Analysis was performed on 240 individuals collected in 1926 together with 12 individuals representing a core collection of the Sardinian barley landraces. This set comprised five individual belonging to the "Northern gene pool" of the he Sardinian landraces, five belonging to the southern and two individual that were admixed between the two gene pool (Bellucci et al. 2014).

Statistical analysis

The differences between populations were assayed using analysis of variance (ANOVA) considering the individuals of the population as "replicates". To compare population means, the Tukey-Kramer multiple comparison test was used. The relationships between populations

were studied by principal component analysis (PCA) using the means of the populations for all the nine phenotypic traits. Analyses were performed with the software JMP ver. 7.00 (SAS Institute, 2007). Population genetic analyses were conducted using Arlequin vers. 3.5.1 and Structure software.

Results

The analysis of variance showed that in 2013 the differences between populations were highly significant ($P<0.0001$) for five traits and with lesser extent ($P<0.01$) for plant height and spike length. Population differences did not reach significance for number of kernel per plant ($P=0.07$) and yield per plant ($P=0.10$) (Table 1.1).

The Table 1.2 shows averages for all the nine characters and for all populations. In particular, it must be noted the high degree of variation among population in the number of plants that did not flower. The Sardinian barley landraces are mainly of intermediate type (Papa et al, 1999), thus variation for this trait might be particularly informative to discriminate between potential "old relatives" of the recent SBLs.

Table1.1 – Comparison (by one-way ANOVA) among the eight populations collected in 1926 for the nine phenotypic traits recorded in 2013. d.f. = degree of freedom; M.S. = Mean Square.

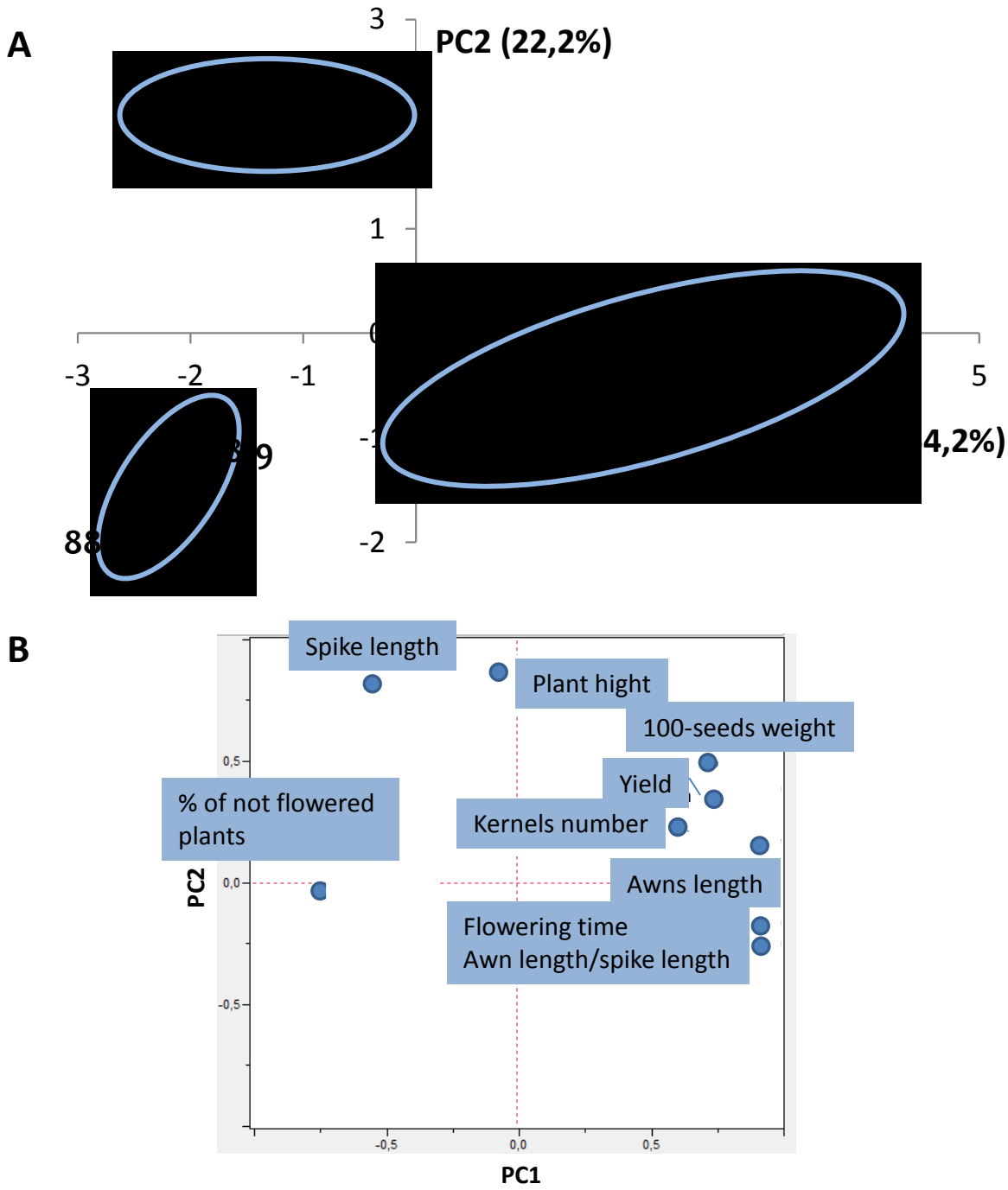
Trait	**d.f.**	**M.S.**	***F***	***P***		**d.f.**	**M.S.**	***F***	***P***
% of flowered plants					**Awns lenght/spike lengh**				
Population	7	1.6	8.62	2.33E-09	Population	7	5.1	7.27	2.22E-07
Error	225	0.2			Error	134	0.7		
Flowering date (days)					**N. of kernels**				
Population	7	1.7	11.8	1.04E-11	Population	7	4224.7	1.92	0.0711
Error	136	0.1			Error	132	2199.8		
Plan hight					**Yield per plant**				
Population	7	311.3	2.79	0.0095	Population	7	6.5	1.74	0.1046
Error	136	111.5			Error	132	3.7		
Spike lenght					**1000-seeds weigth**				
Population	7	8.5	3.33	0.0027	Population	7	530.9	4.60	0.0001
Error	134	2.6			Error	132	115.4		
Awn lenght									
Population	7	42.7	6.16	2.97E-06					
Error	134	6.9							

Table 1.2 – Means of the eight populations for each of the nine phenotypic trait recorded in 2013. For each trait, means that do not share the same letter are statistically different at the Tukey-Kramer HSD test for multiple comparisons.

Population	% of not flowered plants	Flowering time (% at the 10 of june)	Culm length (cm)	Spike Lenght (cm)	Awn lenght (cm)	Awn lenght / spike lenght	N. of kernel per plant	Yield per plant (g)	1000-seeds weight (g)
8819	33.0 BD	0.0 D	52.6 AB	5.7 AB	8.3 C	1.5 B	23.1	0.7	30.3 C
8825	41.2A-D	10.0 BD	58.6 AB	6.9 A	9.9 A-C	1.5 B	50.1	1.9	37.7 A-C
8815	15.2 CD	82.1 A	54.0 AB	4.8 B	12.1 A	2.8 A	47.9	2.1	43.4 A
8822	3.0 D	34.4 BC	50.3 AB	5.8 AB	10.9 AB	2.1 B	41.8	1.7	41.3 AB
8832	36.4A-C	33.3 BD	46.6 B	5.6 AB	9.0 BC	1.7 B	70.5	2.4	32.8 BC
8813	61.5 AB	0.0 D	59.8 A	7.2 A	8.4 BC	1.2 B	33.7	1.4	36.9 A-C
8827	64.5 A	0.0 D	46.7 B	5.5 AB	7.8 C	1.5 B	21.9	0.8	31.6 BC
8836	60.0 AB	50.0 AB	50.6 AB	5.5 AB	10.0 A-C	2.0 B	43.1	1.4	31.3 BC
Minimo	3.00	0.00	46.62	4.79	7.80	1.22	21.90	0.71	30.31
Massimo	64.50	82.14	59.80	7.15	12.11	2.83	70.45	2.42	43.37

The principal component analysis applied to average population suggest that accessions collected in the 1926 can be grouped into three main groups (Figure 1.3). The Figure 1.3 shows also the correlation the original variables and the first two principal component.

Figure 1.3 – A) Relationships among the populations collected in 1926 based on nine phenotypic traits; sowing: February 2013. Loadings of the traits on the principal components.



In particular the first principal component (PC1) captures more than 50% of the phenotypic variance, and separated the populations 8832, 8822 and 8815 and to a lesser extent the population 8836, from the remaining four population 8813, 8825, 8819 and 8827. In particular the populations with high values for the PC1 showed a low percentage of not -flowered plants (i.e. they did not require vernalization and have low photoperiodic sensitivity), tend to be early. Moreover, these plants showed long awns and high awns length/spike length ratio. These populations tended to have higher yield, that was achieved with more seeds and with a high weight of 1000-seeds. The second principal component (PC2) explained more than 20% of the total variance and it separated populations mainly based on plant height spike length. The PC2 also reflected, although to a lesser extent, some productive characteristics such as weight of 1000-seeds and yield per plant. The biggest difference for PC2 is that between 8827 versus 8825 and 8813; the populations 8815, 8819, 8822, 8832 and 8836 occupy an intermediate position for PC2 (Figure 1.3)

Microsatellite genotyping

The analysis conducted with SSR markers showed that the populations collected in 1926 possess diversity levels quite variable (Figure). One population (8819) was monomorphic while the population 8832, the most diverse, showed 25 haplotypes out of 30 individuals analyzed with an expected heterozygosity of 0.626. A similar level of diversity was observed for the population 8822.

Figure 1.4 – Descriptive diversity statistics for the populations collected in 1926 and the core collection of Sardinian barley landraces of 1990 and 1999 .

Population	Sample size	Haplotypes	Haplotype/Sampe size	HE
8819	30	1	0.03	0
8825	30	16	0.53	0.505
8815	30	17	0.57	0.527
8822	30	24	0.80	0.613
8832	30	25	0.83	0.626

8813	30	20	0.67	0.535
8827	30	15	0.50	0.443
8836	30	20	0.67	0.482
Core	12	12	1.00	0.592

The lack of genetic diversity for the population 8819 could be explained, given the strong autogamy of barley (selfing >99% Briggs, 1975) , assuming that for this this population the conservation units was likely the single spike.

The analysis conducted with the software STRUCTURE using a240 individuals collected in 1926 and the core collection SBLs and the method of Evanno et al., (2005) determined K = 2 as the most likely number of genetic groups within the collection of individuals. This solution is represented in (Figure 1.5). Overall, the coefficient of membership of the individuals ins quite high indicating well-defined population structure. Indeed the 161 individuals (~67%) were attributed to one of the two groups (Cl.1 or Cl.2) with $q_i > 0.80$. Among the two groups identified by STRUCTURE the one also comprise the SBL of 1990 and 1999 (Cl.1) was characterized by a higher genetic diversity as compared to C1.2 (0.650 versus 0.300) as also by a much lower F_K estimate (0.135 *versus* 0.694). The parameter F_K represents the estimated drift from the inferred common ancestor of all populations, thus similar to F_{ST} but specific for each cluster is expected to be proportional to the divergence from a common ancestral population. A low F_K value indicates little drift away from the ancestral state.
From this picture it was evident that three populations 8822, 8832 and 8815 were more similar to SBL than the other populations collected in 1926. This was corroborated by the inspection of the pairwise F_{ST} matrix between populations (Table 1.3). Moreover, the F_{ST} between the core collection and these tree populations, is very close to the average F_{ST} among SBL populations estimated by using different marker system (Papa et al., 1998; Rodriguez et al., 2012; Bellucci et al., 2013). Other populations (8819, 8827, 8813 and 8836) were not seen in close relationship with the recent Sardinian Landraces being the F_{ST} around 0.300.

Table 1.3 – pairwise distance matrix (F_{ST}) between samples. **Bold:** the three lowest distance between the core collection and the populations collected in 1926. Underlined: F_{ST} between the three populations collected in 1926

	8819	8825	8815	8822	8832	8813	8827	8836	Core
8819	0.000								
8825	0.431	0.000							
8815	0.638	0.329	0.000						
8822	0.543	0.166	0.203	0.000					
8832	0.584	0.267	0.156	0.110	0.000				
8813	0.471	0.181	0.316	0.240	0.263	0.000			
8827	0.593	0.276	0.407	0.297	0.367	0.277	0.000		
8836	0.592	0.310	0.371	0.203	0.275	0.323	0.104	0.000	
Core	0.609	0.200	**0.165**	**0.096**	**0.148**	0.287	0.349	0.277	0.000

Moreover, at K=3, the individuals of the SBL core collection divided according to their provenance (Northern gene pool, Southern gene pool and admixed; Figure 1.5 and 1.6). At the same time, among the five populations of 1926 more similar to the SBL, there was the separation of 8813, 8825 and 8822 from 8815 and 8832. Thus, the North/South structure that in characterized Sardinian barley landraces in the ’90 (Papa et al., 1998; Rodriguez et al. 2012; Bellucci et al. 2014) might have a long history and might had persist for decades until nowadays.

Figura 1.5 – Genetic structure of the eight populations collected in 1926 (240 individuals, 30 per population). For comparison, a core collection (12 individuals) of the Sardinian barley landraces (SBL) collected in 1990 and in 1999.was also included in the analysis. The core collection included five individuals belonging to the "Northern gene pool" of SBL, five to the "Southern gene pool" and two admixed (see Rodriguez et al. 2012; Bellucci et al., 2013).

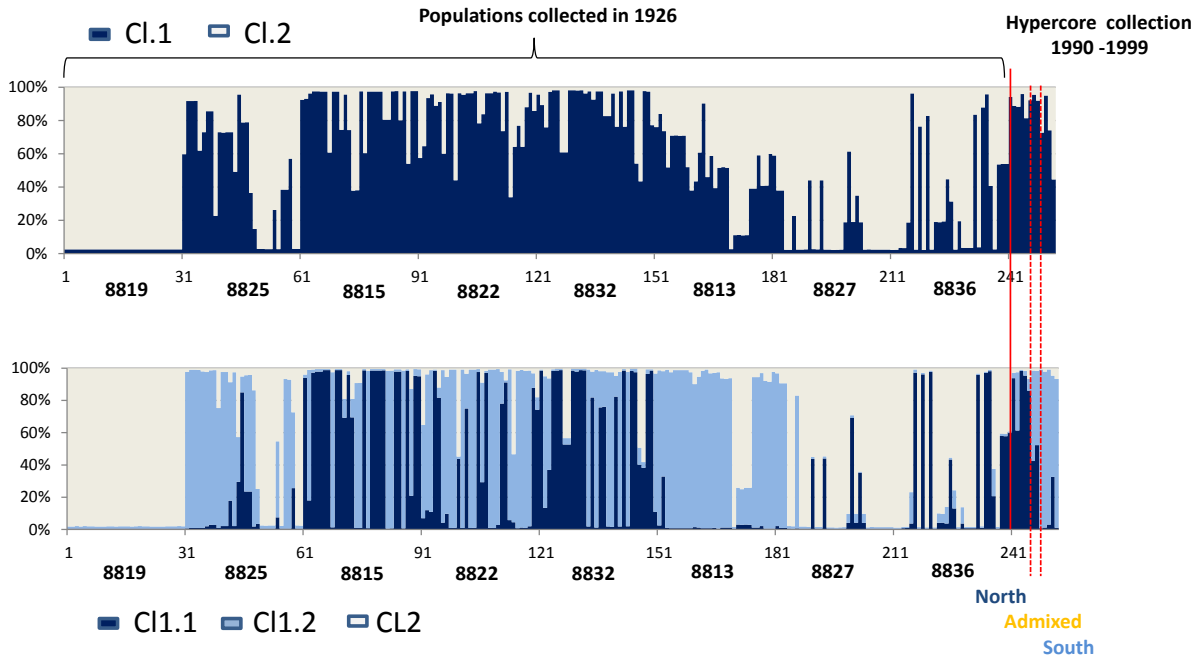
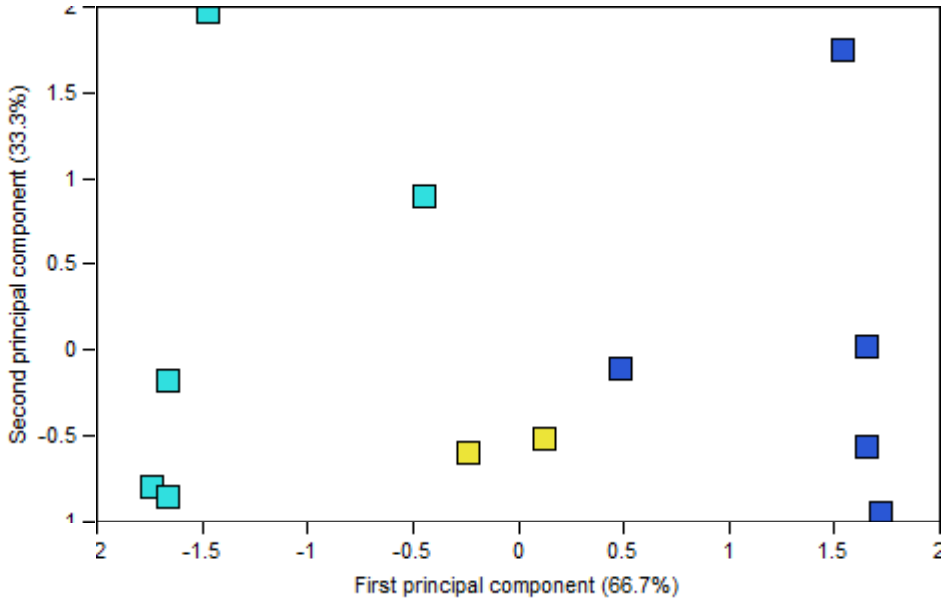


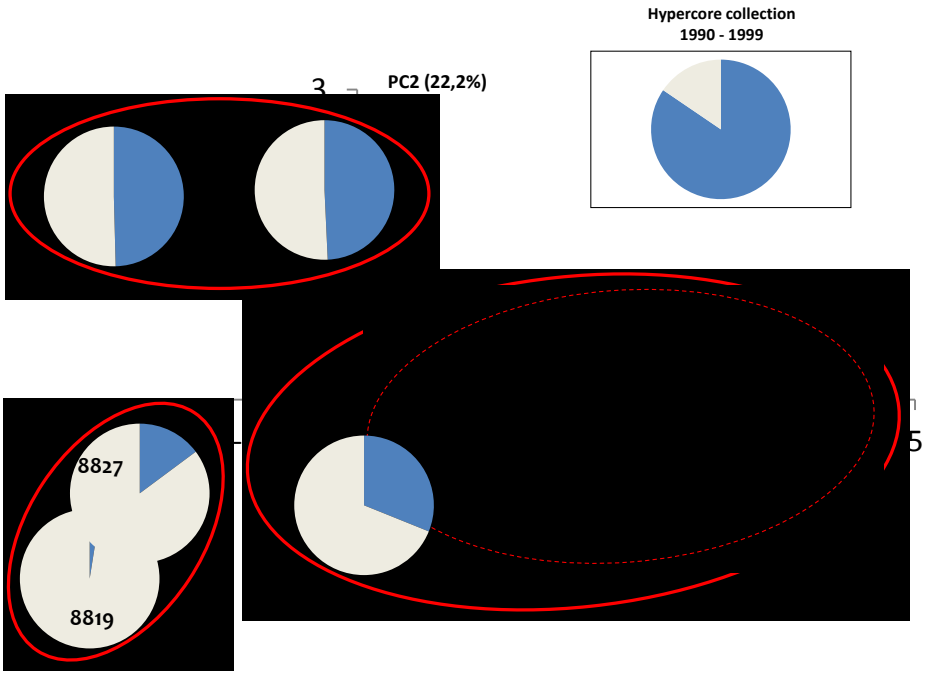
Figure 1.6 – Relationships between the individuals of the core collection of SBLs. The scatterplot represents the result of the Principal Component Analysis (PCA) using the q_i coefficient of membership calculated by STRUCTURE. The different color represent the Color keys: Light blue : SBL belonging to the Southern gene pool; individuals belonging to the "Northern gene pool" are in blue, those belonging to the Southern gene pool in light blue and the admixed in yellow.



Comparison between the phenotypic and molecular characterization of the populations collected in 1926.

It was detected a correspondence between phenotypic and molecular structure. Indeed, the phenotypic group composed by the populations 8836, 8832, 8822 and 8815 also tend to be genetically different for the other groups of populations collected in 1926 ([8827,8819] and [8825 e 8813]) and with a gene pool more similar to that of the SBL. In particular three populations, 8832, 8822 and 8815appeared related to the SBL collected in 1990 and 1999.

Figure 1.7– The figure synthesizes the relationship between the eight population collected in 1926 for the year 2013. Populations are separated based on the first two principal component (PC1 and PC2) calculated on using the mean of nine phenotypic traits recorded in 2013, Each population is represented by a pie which reported the average contribution (q_i) of each genetic group identified by STRUCTURE to the gene pool off the population.



Overall these results stress the biological significance for the population structure identified within the collection of 1926 by microsatellite analysis and led us to hypothesizes that some the populations collected in 1926 can be regarded with interest, as the likely are "progenitors" (and in particular three of them) of the SBL collected about 70 year after in 1990 and 1999. opening the possibility in the next future to conduct analysis of allele frequency variation in time This could allow a better understanding of the genetic bases of plant adaptation to the agro-ecosystem.

Bibliography

Attene G., Ceccarelli S., and Papa R. (1996). The barley (Hordeum vulgare L.) of Sardinia, Italy. Genetic Resources and Crop Evolution, 43(5), 385-393.

Atwell B.J., Han W., Scafaro A.P. (2013). Could abiotic stress tolerance in wild relatives of rice be used to improve *Oryza sativa*?. Plant Science.

Baker A.J.M., Reeves R.D., Hajar A.S.M. (1994). Heavy metal accumulation and tolerance in British populations of the metallophyte Thlaspi caerulescens J. & C. Presl (Brassicaceae). New Phytologist, 127(1), 61-68.

Bellucci E., Bitocchi E., Rau D., Nanni L., Ferradini N., Giardini A., Rodriguez M., Attene G., Papa R. Population structure of barley landrace populations and gene-flow with modern varieties. Accettato per la pubblicazione su PLOSone.

Comadran J., Thomas W. T. B., Van Eeuwijk, F. A., Ceccarelli, S., Grando, S., Stanca, A. M., Pecchioni N., Akar T., Al-Yassin A, Benbelkacem A., Oubbou H., Bort J., Romagosa I., Hackett C.A., Russell J. R. (2009). Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theoretical and applied genetics, 119(1), 175-187.

Cooper I., Jones R.A. (2006). Wild Plants and Viruses: Under-Investigated Ecosystems. Advances in virus research, 67, 1-47.

Demissie A., Bjørnstad A. (1997). Geographical, altitude and agro-ecological differentiation of isozyme and hordein genotypes of landrace barley from Ethiopia: implications to germplasm conservation. Genetic Resources and Crop Evolution 44: 43–55.

Earl Dent A. and von Holdt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources vol. 4 (2) pp. 359-361 doi: 10.1007/s12686-011-9548-7

Evanno G, Reganut E, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620.

Excoffier, L. and H.E. L. Lischer (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics, 164:1567–1587.

Ganança J.F.T., Abreu I., Sousa N.F., Paz R.F., Caldeira P., Dos Santos T.M.M., Costa G., Slaski J.J., Pinheiro de Carvalho M.A.A. (2007). Soil conditions and evolution of aluminium resistance among cultivated and wild plant species on the Island of Madeira. Plant Soil and Environment, 53(6), 239.

Grossi M., Giorni E., Rizza F., Stanca A.M., Cattivelli L. (1998). Wild and cultivated barleys show differences in the expression pattern of a cold-regulated gene family under different light and temperature conditions. Plant molecular biology, 38(6), 1061-1069.

Hagenblad J., Zie J., Leino, M. W. (2012). Exploring the population genetics of genebank and historical landrace varieties. Genetic Resources and Crop Evolution, 59(6), 1185-1199.

Hajjar R., T. Hodgkin (2007). The use of wild relatives in crop improvement: a surveyof developments over the last 20 years, Euphytica 156, 1–13.

Hikmet B., Melda K., Kuaybe Yucebilgili K. (2013). Drought Tolerance in Modern and Wild Wheat. The Scientific World Journal, 2013.

Jaradat A.A., Shahid M. (2006). Population and multilocus isozyme structures in a barley landrace. Plant Genetic Resources, 4(2), 108-116.

Jarosz A.M., Davelos A.L. (1995). Tansley Review No. 81. Effects of disease in wild plant populations and the evolution of pathogen aggressiveness. New Phytologist, 371-387.

Jordi Comadran, W. T. B. Thomas, F. Á. van Eeuwijk, S. Ceccarelli, S. Grando, A. M. Stanca, N. Pecchioni, T. Akar, A. Al-Yassin, A. Benbelkacem, H. Ouabbou, J. Bort, I. Romagosa, C. A. Hackett, J. R. Russell hide

Lakew B., Henry R.J., Eglinton J., Baum M., Ceccarelli S., Grando S. (2013). SSR analysis of introgression of drought tolerance from the genome of *Hordeum spontaneum* into cultivated barley (*Hordeum vulgare* ssp *vulgare*). Euphytica, 1-13.

Lei D., Tan L., Liu F., Chen L., Sun C. (2012). Identification of heat-sensitive QTL derived from common wild rice (*Oryza rufipogon* Griff.). Plant Science, 201–202, 121–127.

Leino M. W., Boström E., Hagenblad J. (2012). Twentieth-century changes in the genetic composition of Swedish field pea metapopulations. Heredity, 110(4), 338-346.

Leino MW, Hagenblad J (2010). 19th century seeds reveal the population genetics of landrace barley (Hordeum vulgare). Mol Biol Evol 27: 964–973.

Leino MW, Hagenblad J, Edqvist J, Karlsson Strese E-M (2009). DNA preservation and utility of a historic seed collection. Seed Sci Res 19: 125–135.

Lister DL, Thaw S, Bower MA, Jones H, Charles MP, Jones G et al. (2009). Latitudinal variation in a photoperiod response gene in European barley: insight into the dynamics of agricultural spread from 'historic' specimens. J Arch Sci 36: 1092–1098.

Loskutov I (1993). Vavilov and his Institute. The history of the world collection of plant genetic resources in Russia. IPGRI publication.

Mano Y., Takeda K. (1998). Genetic resources of salt tolerance in wild Hordeum species. Euphytica, 103(1), 137-141.

Mazzucato A., Papa R., Bitocchi E., Mosconi P., Nanni L., Negri V., Picarella M.E., Siligato F., Soressi G.P., Tiranti B., Veronesi F. (2008). Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (Solanum lycopersicum L.) *landraces*. Theoretical and Applied Genetics, 116(5), 657-669.

Merilä J., Crnokrak P. (2001). Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology, 14(6), 892-903.

Monneveux P., Ramírez D.A., Pino M.T. (2013). Drought tolerance in potato (*S. tuberosum* L.): Can we learn from drought tolerance research in cereals?, Plant Science, 205–206, 76-86.

Muehlbauer F.J., Kaiser W.J., Simon C.J. (1993). Potential for wild species in cool season food legume breeding. Euphytica, 73(1-2), 109-114.

Nevo E., Chen G. (2010). Drought and salt tolerances in wild relatives for wheat and barley improvement. Plant, Cell & Environment, 33(4), 670-685.

Olsen K. M., and Wendel, J. F. (2013). A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. Annual review of plant biology, 64, 47-70.

Olsen K. M., and Wendel, J. F. (2013). Crop plants as models for understanding plant adaptation and diversification. Frontiers in plant science, 4.

Papa R. Diversità and adattamento in germoplasma sardo di orzo (*Hordeum vulgare* L.). Dottorato di Ricerca (PhD) Thesis, Università degli Studi di Sassari, Sassari, Italy.

Papa R., Attene G., Barcaccia G., Ohgata A., Konishi, T. (1998). Genetic diversity in landrace populations of Hordeum vulgare L. from Sardinia, Italy, as revealed by RAPDs, isozymes and morphophenological traits. Plant Breeding, 117(6), 523-530.

Rao N.K., Reddy L.J., Bramel P.J. (2003). Potential of wild species for genetic enhancement of some semi-arid food crops. Genetic Resources and Crop Evolution, 50(7), 707-721.

Rau D. (2002). Effetto dei fattori agro-climatici sulla struttura della diversità genetica del sistema pianta-patogeno Hordeum vulgare-Pyrenophora teres. Dottorato di Ricerca (PhD) Thesis, Università degli Studi di Sassari, Sassari, Italy.

Rodriguez, M., Rau, D., O'Sullivan, D., Brown, A. H., Papa, R., & Attene, G. (2012). Genetic structure and linkage disequilibrium in landrace populations of barley in Sardinia. Theoretical and Applied Genetics, 125(1), 171-184.

Vigouroux Y., McMullen M., Hittinger C. T., Houchins K., Schulz L., Kresovich S., Y. Matsuoka, Doebley J. (2002). Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proceedings of the National Academy of Sciences, 99(15), 9650-9655.

Watanabe T., Osaki M. (2002). Mechanisms of adaptation to high aluminum condition in native plant species growing in acid soils: a review. Communications in Soil Science and Plant Analysis, 33(7-8), 1247-1260.

Wright, S. I., & Gaut, B. S. (2005). Molecular population genetics and the search for adaptive evolution in plants. Molecular biology and evolution, 22(3), 506-519.

CHAPTER II

Characterization of a collection of introgression lines of common bean (*Phaseolus vulgaris* L.) for the study of the genetic bases of the domestication syndrome at pod level and a close up on the pod shattering trait

Introduction

The domestication of plants

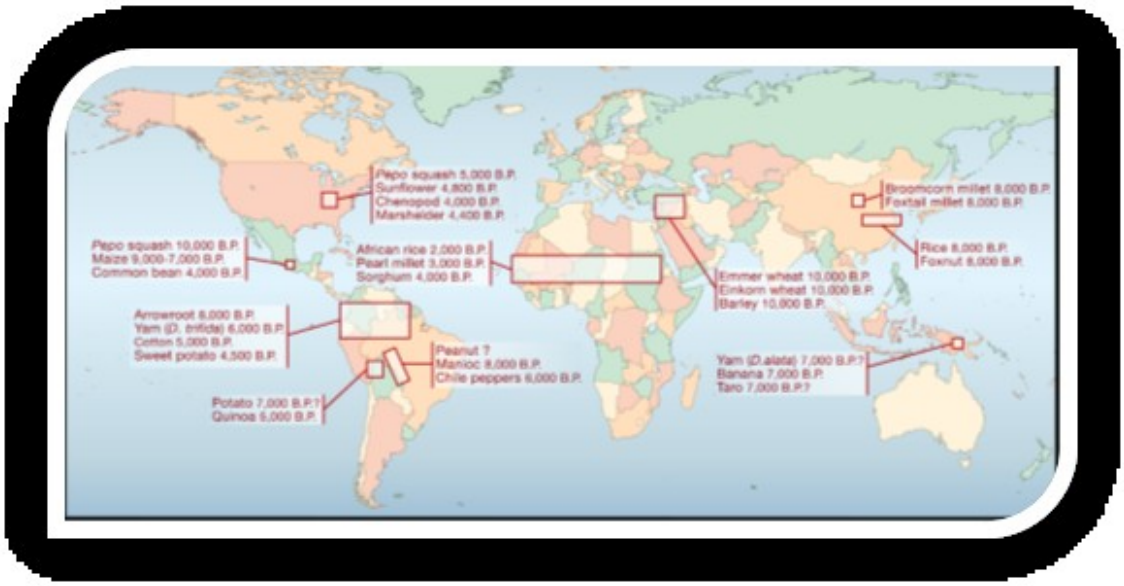
The transition from a nomadic life style dedicated to hunting and gathering of food to a settled civilization based on agriculture has marked a turning point in human history defined as the "Neolithic Revolution" (Smith, 2006). Several approaches have been followed to study this transition such as, among others, archaeological botanical, and genetic. The first prehistoric communities began to cultivate wild plants about 10,000 years ago in various regions of the world (Gepts 2004; Cohen 2009) giving birth to their domestication. Te domestication of plants is the evolutionary process by which genetic modification of wild species has led to new forms of plants, modified to satisfy human needs (Doebley et al., 2006). Considerable efforts have been made to investigate the processes involved in domestication. These studies in factallow to understand the genetic basis of crop adaptation to the agro-ecosystem having an impact on evolutionary biology studies, conservation biology and plant breeding (Lenser and Theißen, 2013; Olsen and Wendel 2013a,b; Dong and Wang 2015).

Charles Darwin, the founder of the theory of evolution presented in the book "The Origin of Species by Means of Natural Selection" (1859), has devoted the first chapter to the changes occurred between wild and domesticated organisms during the process of domestication. One of the main observations made by Darwin was that morphological changes selected during domestication were such that the growth of many crops was completely dependent on human action. For many crops this meant losing entirely the ability to spread naturally, as for example in corn and cauliflower (Doebley et al., 2006). Although the “birth” of the domesticated forms on a millennial time scale has been relatively rapid (Purugganan and Fuller, 2011), the process was long and very complex and it was characterized by random events (mutation, genetic drift) as also by natural and human mediated selection, being this latter both conscious (Ladizinsky, 1987; Kerem et al., 2007; Abbo et al., 2009, 2011, 2012, 2014), or unconscious (Purugganan and Fuller, 2011; Kluyver et al., 2013).

Sites of Origin

Today there are about a dozen of recognized centres of domestication (Doelby et al., 2006) differing in geographical location, size, number and diversity of domesticated species. These were localized in South East Asia (The Fertile Crescent), Africa (Ethiopia and the Sahel), Mesoamerica, the Andes and South America, West Asia and South China (Hawkes 1983; Harlan et al., 1992; Smartt and Simmonds 1995) (Figure 2.1).

Figure 2.1. – Distribution of the recognized independent centers of crop domestication (Doelby et al., 2006).



Other areas, such as Africa, North America, India, New Guinea, East-North America and Europe were less relevant than previously thought (Kluyver et al., 2013). Different events occurred independently in different sites as parallel selection was exerted on similar traits across different species (Olsen and Wendel, 2013).

Domestication Syndrome

The phenotypic changes occurred during the domestication process were substantial, and many of these were in common among different domesticated plants. The set of common traits that characterize the species domesticated from their wild ancestors is known as domestication syndrome (Harlan et al., 1973; Hammer, 1984; Zohary, 1989; Fuller, 2007; Brown et al., 2009).

These traits include:

a) loss of the seed dispersal mechanism. The domesticated plants have lost the ability to disperse the seeds at maturity. This was achieved in different ways by the different species. For example in cereals, this was due to the loss of the "fragile rachis" trait which produces the disarticulation of the spike and the release of the seeds (Gepts, 2004);

b) increased number of seed and, fruit and seeds size size. The fruit and the seeds produced by domesticated plants are usually bigger than those of the wild progenitors. For example, wheat kernels are from 5 to 10 times bigger than those of the wild relatives (Gepts, 2004);

c) reduction of seed dormancy. The seeds of the domesticated plants are characterized by a reduced dormancy while wild ancestors usually have highly dormant seeds. The seed dormancy prevents premature germination and this can be particularly important under environmental conditions that may prevent the growth of seedlings. On the other hand, the lack of seed dormancy allow the growth of more uniform populations as all plants show similar developing stage (Gepts, 2004);

d) compact growth habit. Corn represents a fitting example as teosinte, its wild relative, shows a branching growth habit, which sharply contrasts with the uniculm architecture of the domesticated crop (Gepts, 2004). A compact growth habit allows more uniform growing conditions within the field, this reduces the competition among plants for resources such as water, soil nutrients and solar radiation and finally allows an increase in productivity and a non-scalar maturation and harvest (Evans 1993);

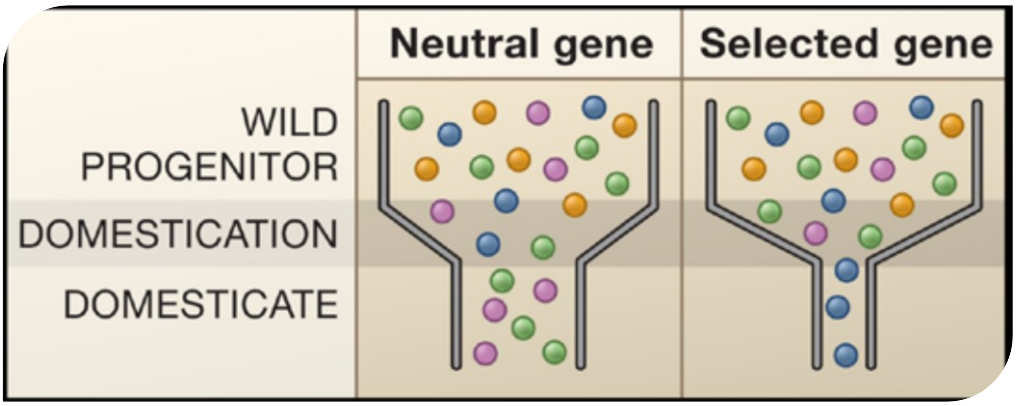
e) a decrease in the number (Kluyver et al., 2013) and angle of tillers (Jin et al., 2008): the wild species of *Oryza rufipogon* shows a prostrate growth habit with a wider tiller angle and short stature with many tillers. This plant architecture increases leaf shade

and therefore decreases photosynthetic efficiency, prohibiting dense plantings (Jin et al., 2008). Such undesirable plant architecture was targeted and continuously selected against by ancient humans, which gradually resulted in the more desirable plant architecture of domesticated rice (*O. sativa*). Domesticated rice shows relatively erect growth (a narrow tiller angle) and fewer tillers, which allows for effective high-yield cultivation. Therefore, plant-architecture selection was a pivotal event in rice domestication.

Effects of domestication

The greatest effect of domestication in crops has been the reduction of the genetic diversity compared to wild progenitors (Gepts, 2004; see Bitocchi et al., 2012 and 2013 for a comparison of bean with other species). Only a portion of the variability of the wild gene pool is today present in the gene pool of the domesticated species, especially in the improved varieties. However, the extent of this reduction varies depending on crop species. The initial agricultural practices led inevitably to variations in the levels of genetic diversity of the cultivated wild populations (Frankel, 1974). Moreover, the early farmers probably have selected a limited number of individuals for the cultivation; this lead to both demographic and selection bottleneck (Doelby et al., 1989; Glemin and Bataillon, 2009). The loss of diversity was a function of the size of the populations and the length of domestication (Eyre-Walker et al., 1998). In particular, the loss of diversity was not evenly distributed along the genome: it was much stronger for the selected loci (and for physically linked "neutral polymorphisms" due to hitchhiking) than for neutral loci that do not influenced the selected phenotypes (called genes neutral) (Wright et al., 2005; Doelby et al., 2006) (Figure 2.2).

Figure 2.2 – The effects of the domestication bottleneck (Doelby et al., 2006)



The common bean

Phaseolus vulgaris belongs to the genus Phaseolus that comprises 70 species; it is one of the five species of the genus, which have been domesticated (*P. dumosus, P. coccineus, P. acutifolius, P. lunatus, P.vulgaris*) (Delgado-Salinas et al., 2006). Both wild and domesticated forms of *P. vulgaris* are self-pollinating, diploid, with 2n = 2x = 22 chromosomes. The common bean (*Phaseolus vulgaris* L.) is the most important legume crop of the world for human consumption; it is a source of complex carbohydrates, protein, fiber and minerals (Broughton et al., 2003). The bean seeds contain about 20-25% protein, of these the highest percentage is constituted by the seed storage protein named phaseolin (Ma and Bliss, 1978).

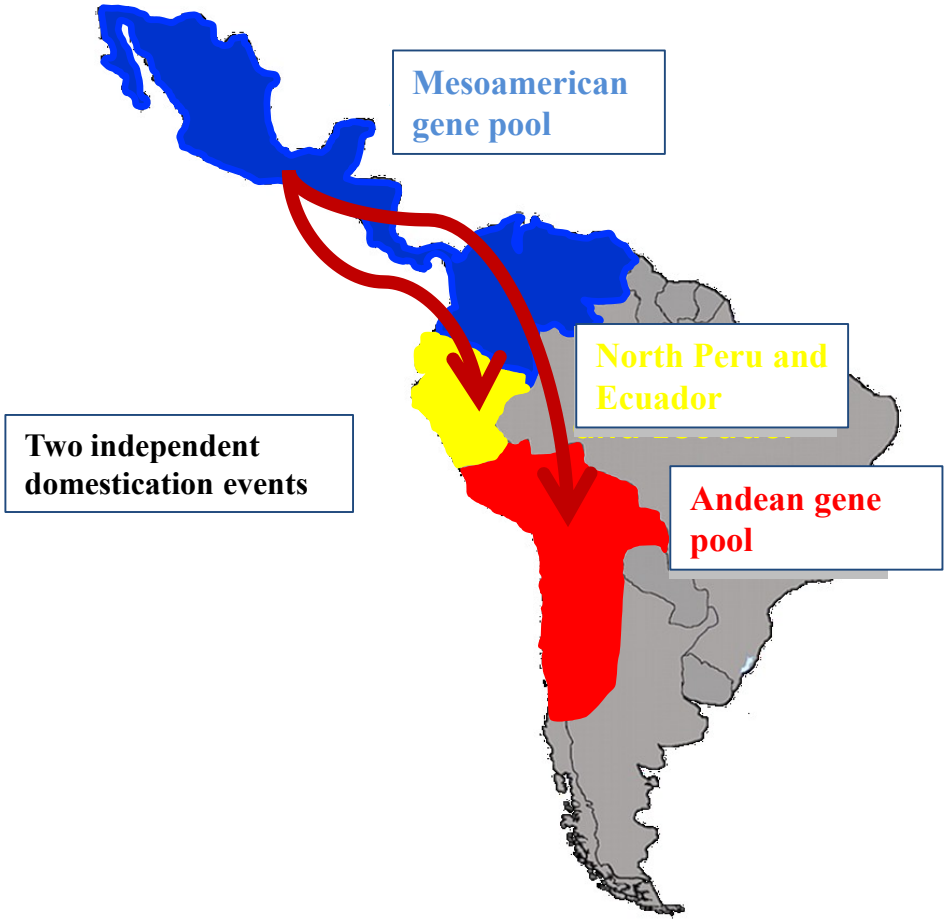
Like other legumes, beans are a good source of micronutrients, more than cereals (Welch et al., 2000), as well as an important source of iron, phosphorus, magnesium; to a lesser extent, they are a source of zinc, copper and calcium (Broughton et al., 2003). For these reasons, in the last decades the cultivation of *Phaseolus vulgaris* showed a significant increase (Akibode and Maredia, 2011). The production is about of 12 million

tons per year (http://faostat3.fao.org/home/E), and the major production areas are Latin America and sub-Saharan Africa (Akibode and Maredia, 2011).

The domestication of *Phaseolus vulgaris*

The common bean is characterized by a particular evolutionary history. This species is native to South America and most likely its centre of origin is Mexico (Bitocchi et al., 2012). Several evidences lead to this outcome: the geographical distribution of the closest relatives, the highest genetic variability of wild Mesoamerican related species than wild species from South America, the occurrence of a heavy bottleneck in the Andes prior to domestication, and finally the presence of wild beans in Mesoamerica closely linked to the wild beans from South America both from the Andes and Ecuador-Peru (Bitocchi et al., 2012). The expansion in South America led to the development of two major wild gene pools: the Mesoamerican and the Andean (Figure 2.3).

Figure 2.3 – Domestication routes in common bean.



Domestication occurred afterwards with two independent events one in Mesoamerica and the other in the Andes (Bitocchi et al., 2012). The occurrence of these two events are supported by different studies based on different approaches: morphological traits (Delgado-Salinas et al., 1988; Sing et al., 1991), agronomic data (Sing et al., 1991), protein seed (Gepts et al., 1986), alloenzymes (Koening et al., 1989) and molecular markers (Bitocchi et al., 2013). The two independent domestication events have given rise to two domesticated gene pools sharply differentiated both at morphological, biochemical and molecular level (Papa et al., 2006; Acosta-Gallegos et al., 2007; Schmutz et al.,2014). The existence of these two geographically distinct and isolated, evolutionary routes of common bean is a unique pattern among the crops of particular interest.

Domestication syndrome in *Phaseolus Vulgaris*

Domesticated and wild forms show marked phenotypic differences (Hawkes et al., 1983; Harlan et al.,1992). Two main characters selected during the domestication process are the reduction of seed dormancy and the loss of dehiscence of the pods; both these traits were crucial for adaptation to cultivated environments (Koinange et al., 1996). Other selected characters are: the determinate growth habit, the increase in size of the seeds and the pods and the reduced sensitivity to photoperiod. In general, it was also observed that the domesticated forms mature earlier than wild progenitors (Koinange et al., 1996).

The domestication of indehiscent fruit in legume crops

The fruit of the Leguminosae is called legume or pod. In the Leguminosae, seed dispersal occurs as a consequence of the shattering (or dehiscence) of the pod.

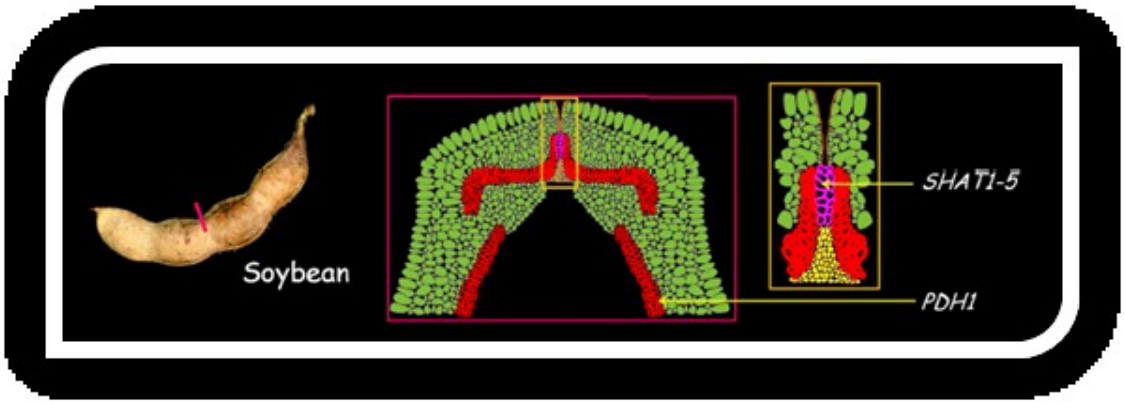
In *Lotus corniculatus* pod shattering is a major problem as high seed loss occur due to the scalar flowering and pod maturation (Grant et al., 1996) and at present, management practices have not been successful to control pod shattering. It has been observed that the orientation of the cells in the pericarp and the degree of lignification may affect this process. Environmental conditions are also implicated, having the relative humidity (between 35% and 49%) a major role at the time of harvest (Grant et al., 2006). Pod dehiscence under greenhouse and open field condition are uncorrelated. The trait is probably controlled by more than one gene and, despite it has a high heritability, in *Lotus* breeding the attempt to reduce shattering through recurrent selection has been unsuccessful. However, interspecific hybridization is suggested as a promising approach (Grant et al., 2006).

In cultivated soybean [(*Glycine max* Merr. (L.)] , the indehiscent pod is a major domestication trait that has been targeted by selection under domestication (Hymowitz, 1970; Harlan, 1992;Dong et al. 2014; Funatsuky et al., 2014). Pod shattering happens after the maturation stage, i.e., when humidity levels are low and dehydration of tissues occurs (Tiwari and Bhatia, 1995). Indeed, pod dehiscence is due to the formation of an

abscission layer along the suture between the two pod valves. After the maturation, dehydration leads to the distortion of pod tissues that accumulated elastic energy. When this latter overcomes the force that binds the two valves along the ventral suture, shattering occurs (Liljegren et al., 2004; Mitsuda and Ohme-Takagi, 2008; Ogawa et al., 2009).

Cultivated soybean is more resistant to shattering than wild soybean (*Glycine soja* Sieb. and Zucc.). In a previous study, the genetic analysis conducted using a mapping population derived from the cross between wild and domesticated soybean did not allow the identification of major QTLs (Quantitative Trait Loci) for the shattering trait (Liu et al. 2007). However, more recently, a domestication shattering gene SHATTERING1-5 (*SHAT1-5*) has been mapped on chromosome 16 of soybean (Dong et al., 2014). This is an homolog of the *A. thaliana AtNST1/2* that acts as a master transcriptional activator of secondary cell wall biosynthesis and it promotes the thickening of the "*fiber-cap cells*" along the ventral suture of the pod valves (Dong et al., 2014). The expression of *SHAT1–5* is specifically localized in the developing fiber cap cells. Interestingly, the abscission layer is functional in both cultivated and wild soybeans (Dong et al., 2014) but *SHAT1-5* is differentially expressed in wild and cultivated soybeans, being up-regulated in the wild, so that in indehiscent pods the fiber cap cells in the ventral suture show an extreme lignification which hinders pod shattering after maturation (Figure 2.4; Dong and Whang, 2015).

Figure 2.4 – Ventral section of soybean pod and tissues where SHAT1-5 and PDH1 act (from Dong and Wang, 2015).



In *Arabidopsis* double mutants plants (nst1-1 and nst3-1) characterized by indehiscent pods were developed (Mitsuda et al., 2005, 2007). These mutants are not able of thickening the secondary cell walls in the interfascicular fibers. But, when both the wild and the domesticated alleles of soybean *SHAT1-5* are used to transform these *Arabidopsis* mutants, this ability is restored. Moreover, wild and domesticated soybeans did not show fixed amino acid difference between protein products of this gene. This suggests that the differential expression of *SHAT1–5* could be due to regulatory changes. In addition, it has also been shown that up-regulation of *SHAT1–5* is due to the disruption of a repressive *cis*-regulatory element in the 5′-promoter region (Dong et al., 2014). It has also been argued that selection under domestication favored changes in this part of the gene as null-mutants show undesirable pleiotropic effects.

Variability for the degree of shattering was also documented within the cultivated gene pools of soybean. The cultivated varieties less resistant to shattering were more resistant than the wild species though these were also not well adapted to the cultivation under conditions of low humidity, particularly for the mechanical harvest (Tsuchiya, 1987).

In the cultivated soybean a *major* QTL (qPDH1, QTL for Pod Dehiscence 1) on chromosome 16 has been identified (Kang et al., 2009; Yamada et al., 2009). Anatomic analysis on NILs (*near-isogenic lines*) for this QTL did not evidence differences in the morphology of ventral suture included the formation of secondary cell wall (Suzuki et al., 2009). Moreover, the genomic region corresponding to this QTL did not contain homologs to *Arabidopsis* associated with pod dehiscence (Suzuki et al., 2010). Recently, the QTL qPDH1 has been cloned (Funatsuki et al. 2014) and it has been shown that it codifies for a "*dirigent-like*" protein whose expression is correlated with the lignin deposition in the inner sclerenchyma of the pod valves. This gene regulates the dehiscence increasing the twisting force in the pod wall at low humidity (Funatsuki et al., 2014). Moreover, the study of the geographical distribution of the allele conferring shattering resistance showed its relevance for the global expansion of soybean crop (Funatsuki et al. 2014).

A SNP that leads to a premature stop codon in *PDH1* generates a truncated and non-functional protein and is responsible of the indehiscence of fruit in cultivated soybean (Funatsuki et al., 2014). However, molecular mechanisms leading to indehiscent pod by *qPDH1* remain not fully resolved.

The results of Dong et al. (2014) and Funatsuki et al. (2014) as a whole suggest that domestication and breeding when acted to reduce shattering also indirectly acted upon several underlying genes, among which *SHAT1–5* and *PDH1*. This suggests that the interaction among genes might be relevant to fine-tune the indehiscence degree of cultivated soybean that is adapted to different environments.

Similar to other legume crops, the reduction of pod dehiscence represents a key domestication syndrome in the domesticated common bean. Indeed, as for soybean and other crops the natural seed dispersal of the common bean became an undesirable trait when the plant was brought under cultivation. Pod dehiscence mechanism has been attributed to the presence of fibers surrounding the vascular bundle in the pod walls as well to the oblique orientation of the fiber in parchment layer lining the pod cavity (Roth, 1977). Decreased fiber content in the pod walls would reduce or retards the dehiscence (Wilke, 1972). After domestication selection for reduced pod fiber content pursued even further, eventually leading to fibreless pod type ("snap" or "stringless" bean in English, "mangetout" in French, "habichuela" or "ejote " in Spanish) which do not open at all, even after maturity.

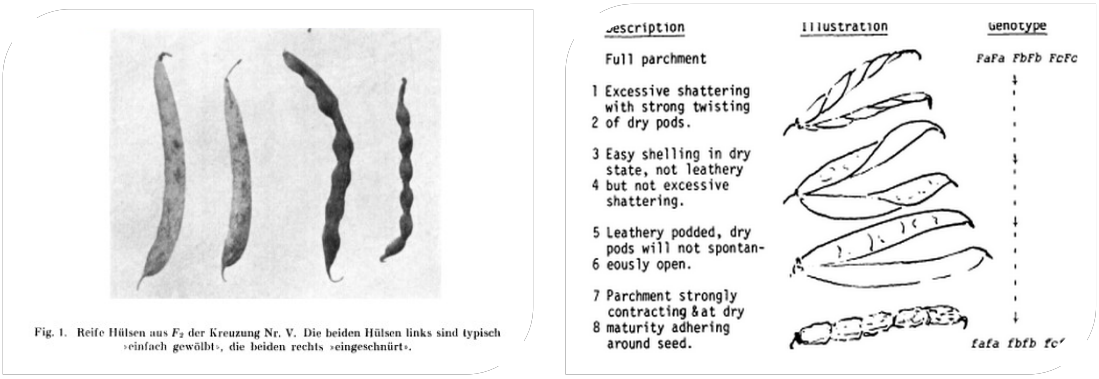
The first studies (sometimes contradictory) on the genetic control of pod wall characteristics in common bean can be traced back to more than one-hundred years ago. These studies distinguished two main types of pod wall 1) tough, parchment or fibrous, and 2) tender, non-parchment or fibreless.

Tshemark (1901, 1902) mentioned dominance of tough, parchment or fibrous over tender, non-parchment or fibreless. On the other hand, Emerson (1904) reported on crosses between stringy and stringless podded varieties. The pod of F1 progenies were sometimes intermediate between the two parents but overall found a strong tendency toward dominance of tenderness (stringless). Wellensiek (1922) crossed the stringy variety Wagenaar with three different stringless varieties. In all of the three F1 generations, stringless appeared to be dominant with clear-cut monohybrid segregation in F2. Tjebbes and Kooiman (1922) suggested the dominance of the parchment type. Joosten (1924) examined many stringless varieties and the degree of their stringlessness and point out the high variability of the character and the probable influence of external factors. Currence (1930), in crossing stringy and stringless varieties reported that crosses between fibrous and fibreless types gave F1 stringless in two case and intermediate in an another one. He explained his results on the basis of a model with

two genes interacting epistatically. Specifically, it assumed that the dominant gene "S" produces stringlessness and a second factor, T, when present inhibits the action of the first factor, i.e. only "SStt" and "Sstt" are stringless.

Lamprecht (1932) distinguished two main types "einfach gewolbt" (tough, parchment or fibrous) and "eigenshnurt" "tender, non-parchment or fibreless Lamprecht (1932) suggested that the quantity of fibers in pods valves (*parchment*) is controlled by a major gene (termed F_A) and three additional genes with minor effects. The progressive accumulation of recessive alleles at these genes would be correlated to the reduction of shattering ability. (Figure 2.5)

Figure 2.5 – Left: the two main types of common bean recognized by Lamprecht (1932) based on pod wall characteristics. Right: schematization of the hypothesis of Lamprecht (1932) on the genetic control of pod shattering in common bean.



Subsequently, Prakken (1934) hypothesized that the genetic control of the traits parchment and stringness (i.e. absence of filamentous fiber along the pod ventral suture) is independent. More in particular the study of Prakken (1934) provided several insights on the string strenght and toughness of the pod wall.

String strenght

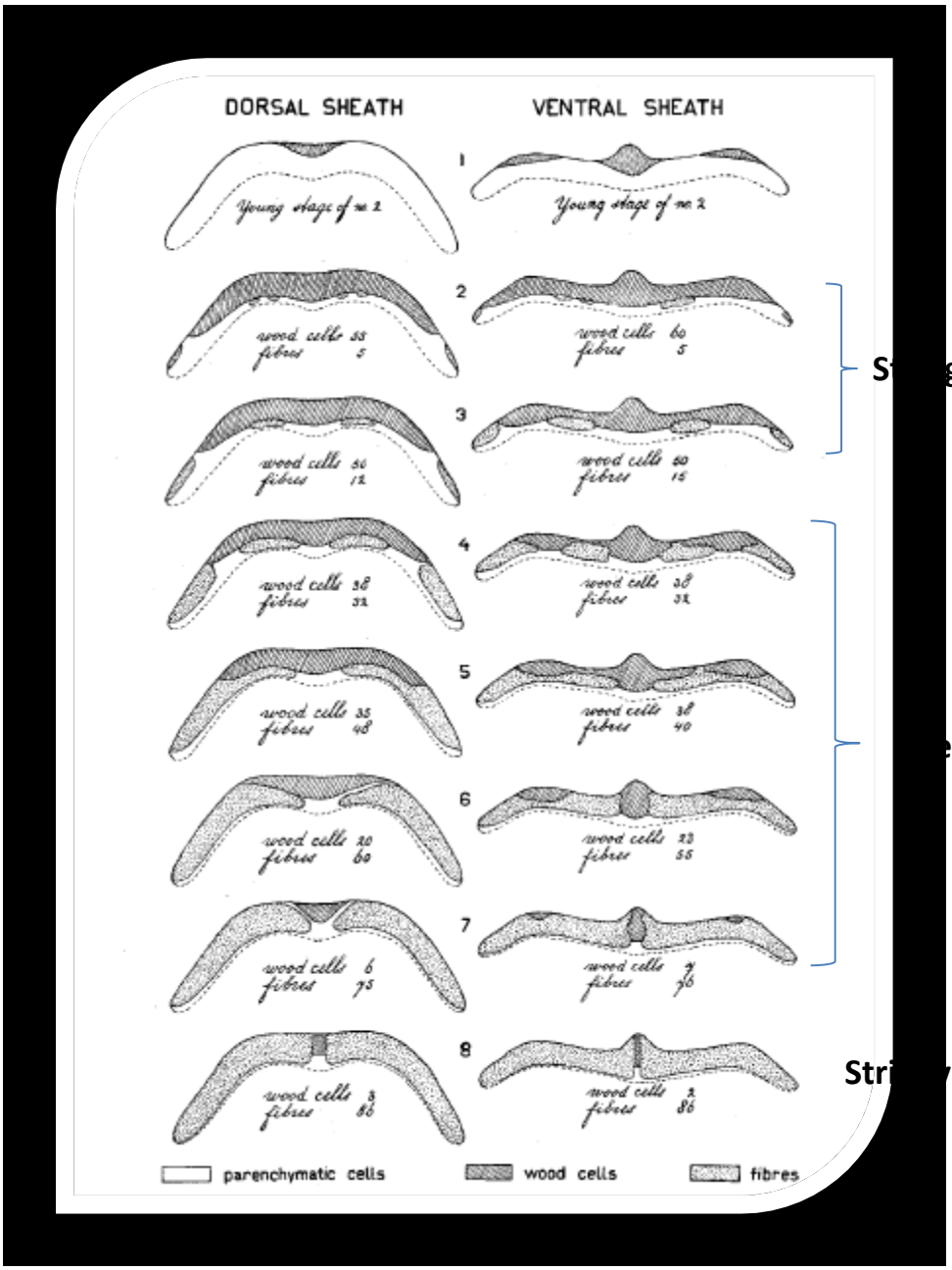
Prakken (1934) observed that the strength of string mainly depends upon the percentage of fibers in the sheath of the vascular bundles in dorsal and ventral sutures. In all the analyzed plants a) the ventral sheath consisted of 4-to-7, the dorsal of 5-to-8 cell layers, b) in each sheath occured parenchymatic cells, wood cells and fibers; c) these three cell

types always have the same relative position. He observed that in the pure stringy type variety Wagenaar (Figure 2.6, n. 8) the sheath consists entirely of fibres, except for a narrow part in the middle. In the stringless type (Fijne tros) (Figure 2.6, n. 2 and 3) the sheath mainly consists of wood cells with the fibers forming four small groups. The F1 generation (Figure 2.6, n. 4, 5 and 6) was intermediate, though similar to the stringless type. In F2 the pure stringy Wagenaar type ("stst") reappeared in a clearly unifactorial way. The percentage of fibres and the strength of string in homozygous and heterozygous "St" plants were influenced by other factors. In a sheath without or nearly without fibers the percentage of wood cells may slightly influence the strength of string.

Toughness of the pod wall

A tough or parchmented pod wall was due to a fibrous layer between inner and outer parenchyma. In the stringless type Fijne tros variety the fibrous layer was totally lacking; the dry pod is conspicuously constricted and shrivelled. The stringy Wagenaar variety had a rather weak fibrous layer; only parts of its cells were lignified. In F1 the fibrous layer was extremely thick and hard; all its cells were lignified. The dry pod was not constricted or shrivelled at all. The segregation into non-parchmented ("toto") *versus* parchmented ("ToTo") was unifactorial. Prakken also provided the genotype for Fijne tros (stringless/non-parchmented) "StSt toto" and Wagenaar (stringy/parchmented) "stst ToTo".

Figure 2.6 – Picture showing the distribution of the different cell types in the dorsal and ventral sheaths of the pod valves in common bean as a function of the different degree of stringlessness (from Prakken, 1934).



Overall the data collected by Prakken (1934) suggests some parallelism with what has been observed in soybean (Dong et al. 2014; Funatsuki et al., 2014).

More than forty years later Drijfhout (1978) analysed string formation in F_1, F_2 and F_3 generations of crosses between cultivars without string, with incomplete string and with complete string, and in BC_1 and BC_2 generations of the hybrid between a stringless and a stringy cultivar backcrossed to the stringless parent. Segregation ratios suggested that two genes were involved in string development:

- stringlessness being controlled by the dominant alleles of "St" ("StSt"; "Ststst")

- complete string being controlled by the recessive alleles of “St” (“stst”)
- incomplete string development, at high temperature and in genotypes with the dominant St allele, being controlled by the dominant allele of the second gene (“Ts”) (“StSt” or “Stst” with “TsTs” or “Tsts”). These results are somewhat similar to those found from Currence (1930).

Shelmidine and Hartmann (1984) collected evidences for two recessive genes and two dominant genes controlling string development in one bean population.

Taken all together, the data point toward a relatively simple determination of shattering in bean with probably not less than 2 and up to 3-4 genes involved.

It has been affirmed that the indehiscent fruit results from the loss of fibers in the sutures (“stringless”), which is under the control of a major QTL, *St* locus on linkage group 2. (Koinange et al., 1996).

Nanni et al. (2011) characterized a genomic sequence in *Phaseolus vulgaris* of 1,200 bp (*PvSHP1)* that is homologous to *SHATTERPROOF-1* (At*SHP1*), a gene involved in the control of fruit shattering in *Arabidopsis thaliana.* The gene AtSHP1 is involved with the *Arabidopsis thaliana*, *INDEHISCENT* gene (At*IND*) in silique shattering. However, the *PvSHP1* fragment was mapped to chromosome Pv06 in *P. vulgaris* (nor on chromosome Pv02 like the *St* locus) and it is linked to gene *V*, responsible for flower and seed color. Moreover, the polymorphisms in this gene are not associated with the shattering (Nanni et al. 2011). In *Arabidopsis thaliana*, the gene *INDEHISCENT* (*IND*) is the primary factor required for silique shattering. More recently, Gioia et al. (2012) mapped *PvIND1*, a homolog of *AtIND*, in common bean in a region near to the *St* locus on chromosome Pv02. Although *PvIND* maps near the *St* locus, it do not completely co-segregates with the *St* trait nor with the dehiscent/indehiscent phenotype. Gioia et al. (2012) concluded that *PvIND* may not be directly involved in pod shattering and may not be the gene underlying the *St* locus. These authors also called for a more precise phenotyping method to characterize shattering in common bean.

Recently, Dong and Whang (2015) underlined that while Gioia et al. (2012) postulated *PvIND* as the *AtIND* homolog based on sequence homology in the conserved b-HLH domain and it is possible that polymorphisms in other *AtIND* homologs in common bean may be associated with pod indehiscence. They based their two observations on: a) the *IND*-related transcription factors are specific to Brassicaceae; b) the role of *AtIND* in the shattering of *A. thaliana* may have been acquired since the duplication event

recently occurred in the Brassicaceae *HECATE3* (*HEC3*) gene clade (Liljegren et al., 2004; Girin et al., 2011). However, as fibers are mainly composed of sclerenchyma cells with well-developed secondary cell walls, it is also likely that genes involved in the regulation of secondary cell wall deposition or fiber cell differentiation may have contributed to control pod dehiscence in the *St* locus (Dong and Whang, 2015).

OBJECT

The main objectives of the present activity were:

1) the phenotypic and molecular characterization of a collection of introgression lines of common bean (*Phaseolus vulgaris* L.), developed *ad hoc* to study the domestication syndrome in particular for pod and pod-related traits.
2) to specifically focus on the study of pod shattering trait in common bean to decipher its genetic bases.

The achievement of the first task will allow building phenotypic and molecular datasets that will serve marker-traits association studies that ultimately will lead to the identification and the cloning of genes underlying very important agronomic traits.

The second task wants to directly assess the potentiality of the collection for gene mapping studies focusing on pod shattering, one among the most relevant domestication traits in common bean. Deciphering its genetic bases is important for both evolutionary studies and breeding purposes. Indeed, the comparisons with other plant species will allow answering questions about the existence and the relevance of molecular mechanisms of parallel evolution in the Leguminosae species. A detailed understanding of the shattering mechanism and the identification of the underling genes will also provide breeders with key information to manipulate this trait and reduce yield losses.

MATERIALS AND METHODS

Plant materials: the collelction of introgression lines (ILs)

A population of 287 introgression lines (ILs) was phenotyped. This is a representative fraction of a larger set of about 600 ILs developed by the research group supervised by Prof. R. Papa from the Università Politecnica delle Marche (Ancona, Italy) in collaboration with the research group supervised from the Prof. G. Attene from the Università degli Studi di Sassari (Sassari, Italy).

The population has been developed starting from the cross between the line MG38 and the Andean variety MIDAS. The line MG38 is a *Recombinant Inbred Line*, RIL, obtained from the cross between a wild Mesoamerican genotype, G12873, and the Andean variety MIDAS (Figure 2.7). The genotype MG38 was selected to have a wild phenotype for growth habit (climbing) and pod (small size, curved shape, pigmented valves and not-shattering) and seed characteristics (very small size). Based on AFLP analysis MG38 possess 55% of the genome attributable to the wild Mesoamerica parent G12873 (Papa, personal communication). To obtain the ILs, MG38 was backcrossed with MIDAS as recurrent parent. MIDAS is characterized by big and relatively straight, yellow and not-shattering pod with big seeds.

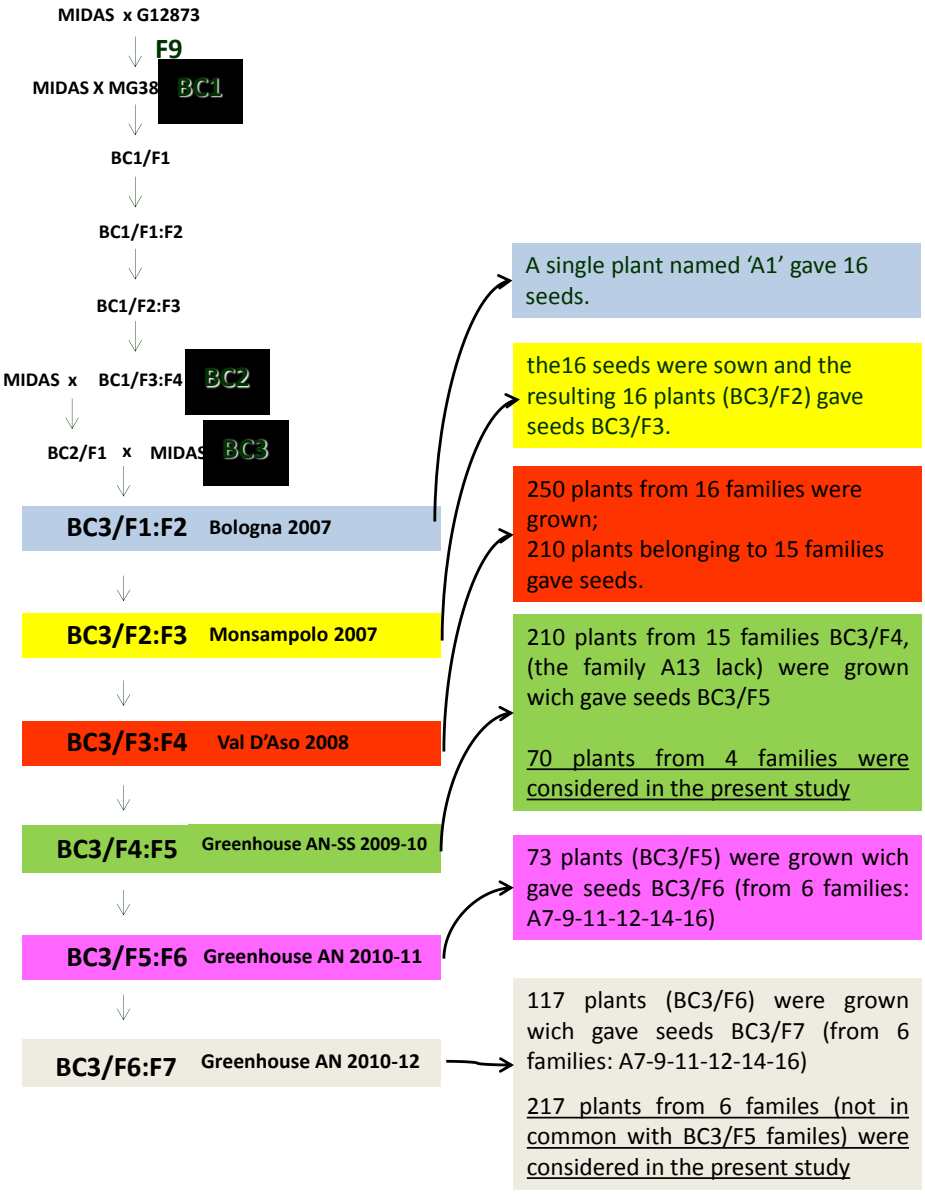
To obtain the introgression lines, different cycles of backcross and selfing were accomplished together with the selection for wild characteristics of the pod and seeds (Figure 2.8). Among the 287 lines analyzed in this study, 70 belong to $BC_3/F_4:F_5$ families while 217 to $BC_3/F_6:F_7$ families. For these latter, based on AFLP marker it was estimated that about 5-10% of the genome can be attributed to the wild line G12873 (Papa, personal communication). Thus, it is expected a high homozigozity level within each family. However, it should be also expected that within each family some genetic segregation is detectable because of the residual heterozigosity (expected lower in $BC_3/F_6:F_7$ than in $BC_3/F_4:F_5$).

Figure 2.7 –Differences between MG38 (on the left) and MIDAS (on the right) for pod traits (photos by M.L. Murgia and D. Rau).

The characteristics of the ILs are of particular interest as they were obtained to allow overcoming some of the most important and common problems that limit the use of wild germplasm into breeding programs: 1) the dominance of the majority of the wild alleles at domestication loci, and 2) the presence of deleterious mutations that can mask the useful variant (Rick, 1974; Eshed and Zamir, 1994; Dwivedi et al., 2008). Moreover, the relatively high number of generations of recombination should allow mapping of the genes with an acceptable resolution. This would facilitate the valorization of the diversity present in the exotic germplasm for bean *breeding* as also the identification of useful genes.

Figure 2.8 – Scheme used to develop the bean introgression lines analyzed in this study.

.

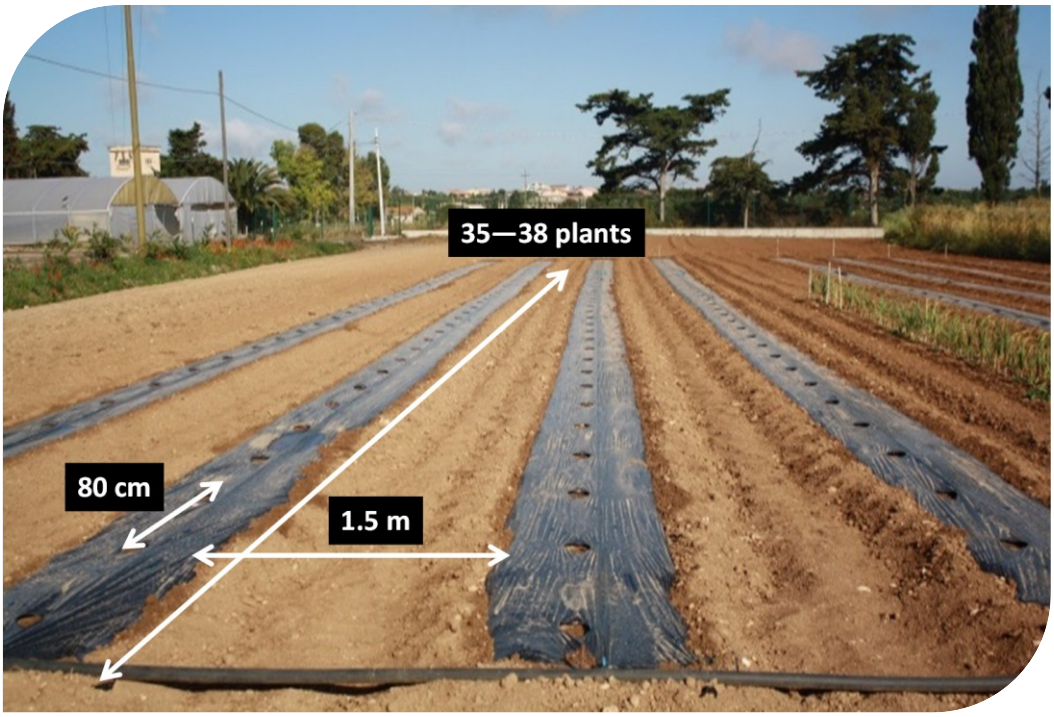


Phenotyping under field conditions

The phenotyping under field conditions was conducted in 2014 between May and October (sowing date May 19th). Per each line, a seed rate of two seeds per hole was used. A field scheme of eight rows with 35-38 holes per row was chosen, with a distance between rows of 1.5 m and 0.8 m between holes within row. The positions of

the lines in the field were randomized. Along each row, a plastic sheet was used to facilitate weeds control (Figure 2.9).

Figure 2.9 – Field trial design.



The following traits were collected in the field: number and orientation of the cotyledonary leaves, stem color, flower color; growth habit, flowering ad pod setting time (days from the 26th of June), plant height (cm), plant vigor (as product of the height x width, cm^2) and pod color.

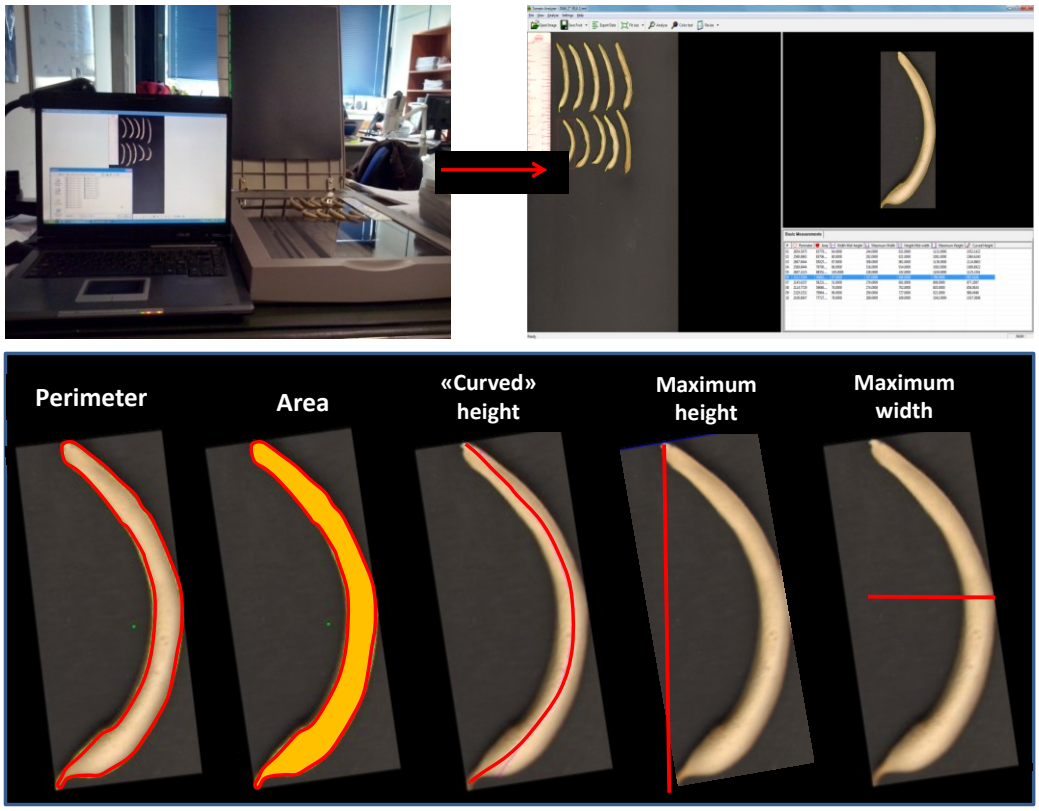
Precision phenotyping of pod traits

At maturation stage, 10 randomly sampled not-shattered pods were collecctd per line. These were then scanned, and the acquired image (600 dpi) was processed with the *Tomato Analyzer* software (Rodriguez et al., 2010) to determine the following traits: perimeters, area, curved height, maximum height and maximum width (Figure 2.10). We also calculated the ratio curved length/maximum height, to synthetically describe the shape of the pod: a ratio = 1 indicates a perfectly straight pod while ratios < 1

indicate a “C” shape, more or less pronounced according to the curvature. All these variables must be referred to the projection area of the pod on the scanner glass.

We first set the procedure on the parental lines MG38 and MIDAS which are markedly different between each other. When different “readings” of the same image were consistently stable we extended this analysis to all of the other bean lines. Statistical analyses were performed on the mean data obtained from 10 pods per line.

Figure 2.10 – Step of pod precision phenotyping. In the order: scan of 10 pods per bean genotypes and automatic determination of five pod traits by tomato analyzer software.

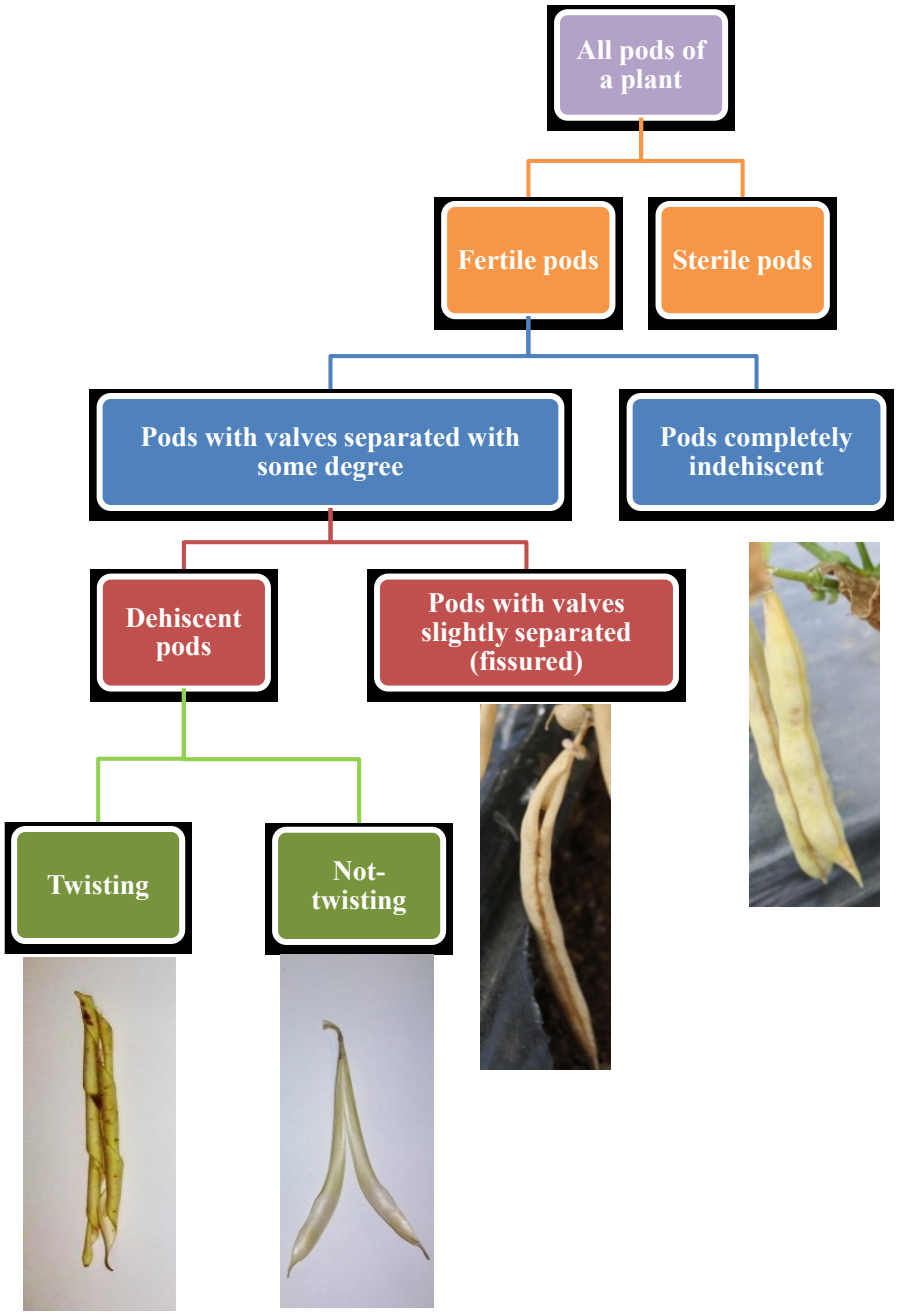


.

A zoom on pod shattering traits

For each plant we counted the number of naturally shattered and non-shattered pods. Shattered pods were further classified into different typologies as exemplified in Figure 2.11.

Figure 2.11 – Method used to classify the pods produced by each plant in order to phenotype the shattering trait.



We considered four categories: indehiscent, with valves "fissured" along the ventral suture between the pod valves, dehiscent with not-twisting valves and dehiscent with twisting valves. It should be noted that the last categories were sometimes difficult to call because of the presence of intermediate cases. At the end of this classification, for each plant the number of pods falling into each category was counted (Figure 2.12). For statistical analyses the number of pods was referred to the number of fertile pods produced by the plant.

Figure 2.12 – Example of pod classification for a bean line

Furthermore, indehiscent pods were trashed by hands per each line; this allowed evaluating the "resistance to manual shattering" based on a scale from 1 (very low resistance = valves separated and abruptly shuttered after a very light pressure on the distal part of the pod) to 9 (very strong resistance = valves did not separate and it is necessary "to break" them). To avoid bias when scoring resistance, this resistance to manual shattering was recorded independently from the pod classification and after that the pods of all of the plants were classified.

After that seeds were separated from the valves, the following traits were determined per plant: pods weight per plant (g), valves weight per plant (g), seeds weight per plant (g), pods per plant (n.), seeds per plant (n.), average pod weight (g), weight of 100 seeds (g), seeds per pod (g), average valves weight (g), seeds per pod (no.), H.I. at pod level (as g of seeds per pod/average pod weight).

Chemical characterizations

This was endeavoured to understand if pod shattering mechanism is correlated to the chemical characteristics of the pod valves

Elementary composition analysis (Carbon, Hydrogen and Nitrogen) of pod valves. For each bean line 2g of dried pod valves were considered pod valves were weighted. Valves were pulverized mechanically using a grinder at a speed of 18000 rpm for 1 min. The pulverized tissues were transferred in FALCON tubes (50 ml) and stored for few days at room temperature in a dry and cool place. The analysis was performed using 0.080 g of pulverised tissue per each line. The samples were combusted at 1000 °C in an excess of oxygen using the LECO CHN 628 (Leco Corporation, St. Joseph, MI) instrument to determine carbon (C), hydrogen (H) and nitrogen (N). This was calibrated using the forage standard "oat meal 502276" with a percentages of 46.43 and 2.64. For every run, three different samples of the standard were included.

The analysis was first performed considering the two parental lines MG38 an MIDAS, for which three biological replicates (i.e. three plants for each of them were available). For each biological replicates, three technical replicates (i.e. three independent analyses) were performed. As we observed highly significant differences between the two parents (See Results section), the analysis was then extended to all the lines. For each lines we analysed

Fiber Analysis

For each plant 6g of dried valves were ground using the Retsch SM 100 mill for 10 minute. Then, the procedure proposed by Van Soest (Van Soest and Wine, 1967) was used.

NDF

To obtain the Neutral Detergent Fiber (NDF), for each sample, 0.50 g of pulverized tissue was used and placed into filter bags. The bags were then accommodated in the ANKOM220 Fiber Analyzer. This instrument performs the digestion at 100 ± 0.5°C at a pressure between 10 and 25 psi. After adding 2 litres of neutral detergent solution, samples were agitated in this solution at 100°C for 80 minutes. The neutral solution is composed by sodium lauryl sulphate and EDTA, with a pH ~7.0. This allows the removal of the soluble fractions present in the plant cell (pectins, sugars and starch, proteins and lipids) retaining the insoluble fraction of the vegetable cell wall. After agitation, samples were washed twice with hot water and twice with deionized water, dried at 100°C and weighted. They were finally incinerated at 500°C. The NDF was expressed as percentage (%) of dried organic matter after subtracting the weight of the ashes. The NDF thus represents the total content in cell wall of the analysed sample.

ADF

To obtain the Acid Detergent Fiber, for each sample, 0.50 g of pulverized tissue were used and placed in another filter bag to be analysed with ANKOM220 Fiber Analyser. In this case, 2 litres of acid detergent solution were added [20g cetyl trimethylammonium bromide (CTAB) added to 1 liter 1N of H_2SO_4]. Samples were agitated at 100°C for 60 minutes. After that, samples were washed twice with hot water and twice with deionized water, dried at 100°C and weighted. The ADF was expressed as percentage (%) of dried organic matter. The ADF is mainly ad intermediate step necessary to obtain the ADL (see here below). However, the difference NDF-ADF roughly estimates the content in hemicelluloses of the analysed sample (Van Soest and Wine, 1967).

ADL

To obtain the Acid Detergent Lignin (ADL) The ADF is treated with sulphuric acid 72% for 3 hours mixing every 30 minutes. The samples were washed twice in hot water twice with deionized water, dried at 100°C and weighted. They were finally incinerated at 500°C. The ADL was expressed as percentage (%) of dried organic matter after

subtracting the weight of the ashes. The ADL was expressed as percentage (%) of dried organic matter and it is an estimation of the lignin content of the analysed sample. Moreover the difference ADF-ADL represent an estimation of the cellulose content of the analysed sample (Van Soest and Wine, 1967).

The determination of the content in cell wall, hemicellose, cellulose and lignin was first obtained for theMG38 and MIDAS for both of which three biological replicates were available. For each biological replicates, three technical replicates were obtained. As the analysis of variance showed clear-cut differences between the parent lines, the analysis was extended to 15 indehiscent lines and 15 lines characterize by a very high shattering level.

Anatomical and histological study of pod valves

This investigation had the specific aim to investigate, in particular differential pattern in cell wall characteristics, with emphasis for lignin deposition.

This analysis was conducted considering five- and 20-days-old pods. Fresh pods were kept in a solution 5:2 of EtOH 95% and glacial acetic acid for three days, and then stored at 4°C in EtOH 70%.Cross-sections of pod bean ventral and dorsal structure were obtained. They were treated Javell water for approximately 10 minutes to eliminate cellular content an maintain cell walls. After washings, cross-sections are immersed in sections Acetic Acid 50% for a few minutes. Cross-sections were the stained with different methods.

Toluidine Blue O staining 1%

The Toluidine Blue O (TBO) is used to differentially stain polysaccharides and lignin (Mitra et al., 2014).It has been used 1g of toluidine blue O in 100 ml of water. Thickened wall of lignin that has not taken up the toluidine blue stain and appears sky blue, the cellulose and hemicellulose to the contrary are stained dark blue (Mitra et al., 2014).

Cochineal carmine and iodine green

Sections were first treated for 2 minutes with green iodine acidified before the use. Sections were then washed several times with ethanol 70% and, finally, with water. Then sections were treated with the cochineal carmine for 30 minutes. Three washes of 10minutes each with ethanol 75%, ethanol 95% and ethanol 100% were performed. Sections are then subjected to a treatment with xylene for 10 minutes is. Lignin is coloured of green by iodine green while cellulose is coloured of pink by cochineal carmine.

Paraffin inclusions

To allow paraffin to penetrate into plant tissues it is first necessary to dehydrated with three treatments with ethanol (at 50%, 75%, and 100%) Ethanol was then replaced by an organic solvent (BioClear, xylene) with three treatments (3/4 ethanol and 1/4 BioClear, 1/2 Ethanol and 1/2 BioClear, and 1/4 ethanol and 3/4 BioClear) each of 30 minutes. Tissues were then treated with BioClear over-night. The BioClear was gradually replaced by the paraffin with three treatments (3/4 BioClear and 1/4 liquid paraffin, 1/2 BioClear and 1/2 liquid paraffin, 1/4 BioClear and 3/4 liquid paraffin) each of 1 hour, finishing with pure paraffin for 1 hour. The samples were included in paraffin over-night. Paraffin inclusions were then cut and sections of 10µm were obtained using a sliding microtome (Reicher). Paraffin was then removed with three treatments with ethanol (ethanol 100%, ethanol 95%, ethanol 75%, Et-OH 50%) and final washing with water. Sections were stained with Toluidine Blue O at concentration of 1%.

Statistical analysis

For each of the registered variable we determined the frequency distribution. The association between variables was quantified by the Pearson 'r' coefficient (quantitative traits) or by contingency analysis (qualitative traits). Differences among groups of lines for the various phenotypic and chemical traits were tested using one way analysis of variance, considering each line as a replicate of the group. The correlations structure among the phenotypic quantitative variables was studied by using Principal component analysis and by clustering correlations based on their pairwise correlation matrix.

To investigate the relationships between resistance to manual shattering and other measure of shattering we adopted the method of recursive partitioning also known as decision trees analysis. This is particularly indicated to investigate relationship among variable without having an *priori* model. to find combinations of genes whose expression distinguished the samples based on histology. In medicine it has been show it is particularly useful to convert a series of symptoms data (X variable) and diagnoses (Y variable) of diseases into a hierarchal system of (dycotomic) questions to be asked to new patients in order to make possible an early and rapid diagnosis. This analysis is particular powerful as it consider a very high number of possible partitions take into consideration only the best one (rif). In our case the categorical x variable was the degree "resistance to manual shattering" while all the other 5 indicators of shattering were considered as possible explanatory Y variables. Differences among group of genotypes were tested using the one-way Analysis of Variance. Statistical analysis on phenotypic data were performed using JMP ver. 7 software (SAS Insitute, Inc., 2012).

Molecular analysis

DNA extraction

The genomic DNA was extracted from young leaves by taking approximately 100-300 mg of tissue from each plant for a total …(Inserire numero individui). The frozen leaf tissue was grinded with a TissueLyserII (Qiagen s.r.l., Milano, Italy) and DNA extracted using the DNeasy 50 Mini Plant Kit (Qiagen GmbH, Hilden, Germany). The quantity and purity of DNA were assessed by spectrophotometric measurement

(GENEQUANT II, Pharmacia Biotech LTD, Cambridge, England). DNA stocks were stored at -20°C.

Pool-seq analyisis

Construction of segregating pools

Two DNA pools for sequencing were made by selecting extreme individuals from the mapping population of 267 plants with the percent of pod shattering ranging from 0–80%. The pool with not shattering individuals ($Pool_{SH}$) was made by mixing equal amounts of DNA from 27 completely indehiscent plants (with percent shattering equal to zero).The second pool ($Pool_{SH}$) was made by mixing equal amounts of DNA from 30 highly shattering plants (more than MG38) i.e. with a percent of shattering from 65% up to 80%. DNA quality and concentration were measured by 0.8% agarose gel electrophoresis, and adjustments were made for a final DNA concentration of 100 ng/mL. Genomic DNA digestion and amplification Fragment selection, extraction and amplification and sequencing were performed by NGS Service of the center of functional genomics of the University of Verona. Divergence between pool was estimated by the using the Δ SNPs index = SNP index of the pool of the highly shattering lines *minus* the SNP index of the pool of not-shattering lines. The SNP index is calculated as fraction of reads per position that are attributable to the MG38 (highly shattering parental line).

GBS analysis

Library preparation and raw data of sequence (Hiseq3000, 2 x 150 bases) were obtained at NG6, INRA, Monpellier,France, as services. Ro calling variants the **Tassel 5 GBS** v2 **Pipeline** was used. Assoiation mapping was perfomed by Using Tassel ver.5, with or without considering kinship. As we retained 4913 SNPs for the analysis with a 20 X coverage, we correct the level of significance byusing Bonferroni method (0.05/n. of markers = 0.5/4913 ~10^{-5}). To further investigate the geneti bases of shattering traits we also adopted the multi-locus mixed-model approach of Segura et al. (2012). As a complement of the associatin analysis we also investigate the co-ooccurence of candidate gene involved in shattering process (table) or in lignin bosythstes within the mapped QTLs.

References	Species	Gene(s)	Gene category	Molecular function	Phenotypic effect	Specie	Number omologue	Ortologues	Chromosome
Liljegren et al.,(2000)	*Arabidopsis Thaliana*	Shatterproof 1	Trascriptor factor	Transcriptional regulator (MADS)	Indehiscent pod	*Phaseolus vulgaris*	26	2	5,8
	Arabidopsis Thaliana	Shatterproof 2	Trascriptor factor	Transcriptional regulator (MADS)	Indehiscent pod	*Phaseolus vulgaris*	24		
Liljegren et al.,(2004)	*Arabidopsis Thaliana*	Indehiscent AT4g00120	Trascriptor factor	Transcriptional regulator (bHLH)	Indehiscent pod	*Phaseolus vulgaris*	9		
Rajani and Sundaresan (2001)	*Arabidopsis Thaliana*	Alcatraz	Trascriptor factor	Transcriptional regulator (bHLH)	Partially Indehiscent pod	*Phaseolus vulgaris*	23	1	
Gu et al., (1998)	*Arabidopsis Thaliana*	Fruitfull	Trascriptor factor	Transcriptional regulator (MADS)	Premature bursting pod	*Phaseolus vulgaris*	26	5	8,9,10
Roeder et al.,(2003)	*Arabidopsis Thaliana*	Replumless	Trascriptor factor	Transcriptional regulator (homeo domain)	Partially indehiscent pod	*Phaseolus vulgaris*	33	2	8,10
Mitsuda and Chme-Takagi (2008)	*Arabidopsis Thaliana*	NST1/NST3 AT1G32770	Trascriptor factor	Transcriptional regulator (NAC)	Indehiscent pod	*Phaseolus vulgaris*	26		
	Arabidopsis Thaliana	NST1 AT2G46770	Trascriptor factor	Transcriptional regulator (NAC)		*Phaseolus vulgaris*	27	1	10
	Arabidopsis Thaliana	NST1 AT1G12260	Trascriptor factor	Transcriptional regulato		*Phaseolus vulg*	24	6	2,3,5,11

				r (NAC)		*aris*			
	Arabidopsis Thaliana	NST1 AT2G13650	Trascriptor factor	Transcriptional regulator (NAC)		*Phaseolus vulgaris*	23		
	Arabidopsis Thaliana	NST1 AT3G61910	Trascriptor factor	Transcriptional regulator (NAC)		*Phaseolus vulgaris*	28		
	Arabidopsis Thaliana	NST1 AT4G36160	Trascriptor factor	Transcriptional regulator (NAC)		*Phaseolus vulgaris*	24		
	Arabidopsis Thaliana	NST3 AT1G76340		Golgi nucleotide sugar transporter 3		*Phaseolus vulgaris*	18	1	10
Arnaud et al., (2010)	*Arabidopsis Thaliana*	GA3ox1	Catalityc enzyme	GA biosynthesis	Partially Indehiscent	*Phaseolus vulgaris*	25	5	5,10,1,9
Ogawa et al.,(2009)	*Arabidopsis TThaliana*	ADPG1	Endo-polygalacturonase	Degrade cell wall matrix	Indehiscent pod	*Phaseolus vulgaris*	18		
Funatsuki et al., (2014)	*Gycine max*	PDH1	Dirigent-like protein	Lygnin biosynthesis	Indehiscent pod	*Phaseolus vulgaris*	28		
Dong et al., (2014)		Shattering1-5	Trascriptor factor	Transcriptional regulator (NAC)	Indehiscent pod	*Phaseolus vulgaris*	26	1	10
Mao et al.,(2000)	*Solanum Lycopersicum*	JOINTLESS	Trascriptor factor	Transcriptional regulator (MADS)	Non-shedding fruit	*Phaseolus vulgaris*	22		
Nakano et al., (2012)	*Solanum Lycopersicum*	MACROCALYX	Trascriptor factor	Transcriptional regulator (MADS)	Non-shedding fruit	*Phaseolus vulgaris*	25	1	3
Schumacher et al., (1999)	*Solanum Lycopersicum*	Lateral Suppressor	Trascriptor factor	Transcriptional regulator (GRAS)	Non-shedding fruit	*Phaseolus vulgaris*	29		
Li et	*Oryza*	Shattering4	Trascript	Transcri	Non	*Phas*	7	2	7

al., (2006)	*sativa*		or factor	ptional regulator (Myb)	shattering seed	*eolus vulgaris*			
Konishi et al., (2006)	*Oryza sativa*	qSH1	Trascriptor factor	Trascriptor regulator (homeo domain)	Non shattering seed	*Phaseolus vulgaris*	34		
Yoon et al.,(2014)	*Oryza sativa*	SH5	Trascriptor factor	Trascriptor regulator (homeo domain)	Non shattering seed	*Phaseolus vulgaris*	34	1	8
Zhou et al., (2012)	*Oryza sativa*	Shattering Abortion1	Trascriptor factor	Trascriptor regulator (AP2)	Non shattering seed	*Phaseolus vulgaris*	27	1	3
Lin et al., (2012)	*Oryza sativa*	Sh1	Trascriptor factor	Trascriptor regulator (YABBY)	Non shattering seed?	*Phaseolus vulgaris*	9	4	2,3,5,11
Lin et al., (2012)	*Sorghum bicolor*	Shattering1	Trascriptor factor	Trascriptor regulator (YABBY)	Non shattering seed	*Phaseolus vulgaris*	9		
Lin et al., (2012)	Zea mays	Shattering1 GRMZM2G085873	Trascriptor factor	Trascriptor regulator (YABBY)	Non shattering seed?	*Phaseolus vulgaris*	9		
	Zea mays	Shattering1 GRMZM2G074124	Trascriptor factor	Trascriptor regulator (YABBY)		*Phaseolus vulgaris*	9	5	2,3,5,11
Simons et al., (2006) Zhang et al., (2011)	Triticum aestivum	Q Traes1ALA9FB6BF52	Trascriptor factor	Trascriptor regulator (ERF)	Free-threshing character	*Phaseolus vulgaris*	33		
	Triticum aestivum	Q Traes3B82E1F5484	Trascriptor factor	Trascriptor regulator	Free-threshing chara	*Phaseolus vulgaris*	30		

				(AP2)	cter				

References	Species	Gene(s)	Molecular function	Specie	Number homologs	Ortologues	Chromosome
	Arabidopsis Thaliana	At5g09330	Transcriptional regulator (NAC)	*Phaseolus vulgaris*	22	1	(9)
	Arabidopsis Thaliana	AT5G13180	Transcriptional regulator (NAC)	*Phaseolus vulgaris*	33	4	(2,8,9)
	Zea mais	Pal3a GRMZM2G074604	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	Pal3b GRMZM2G029048	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	Pal3c GRMZM2G081582	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	Pal3d GRMZM2G160541	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	Pal3e GRMZM2G063917	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	Pal3f GRMZM2G334660	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6	2	(1,7)
	Zea mais	Pal3g GRMZM2G170692	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6	3	(1,7)
	Zea mais	Pal2a GRMZM2G441347	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	Pal2b GRMZM2G118345	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		

	Zea mais	Pal more distant GRMZM2G153871	Phenylalanine ammonia-lyase	*Phaseolus vulgaris*	6		
	Zea mais	C4H1	Trans-cinnamate 4-monooxygenase	*Phaseolus vulgaris*	23		
	Zea mais	C4H2	Trans-cinnamate 4-monooxygenase	*Phaseolus vulgaris*	29	2	(6,8)
	Arabidopsis Thaliana	MYB43 AT5G16600	Trans-cinnamate 4-monooxygenase	*Phaseolus vulgaris*	29		
	Arabidopsis Thaliana	MYB83 AT3G08500	myb domain protein 83	*Phaseolus vulgaris*	29	3	(1,11)
	Arabidopsis Thaliana	MYB63 AT1G79180	myb domain protein 63	*Phaseolus vulgaris*	30	1	(5)
	Arabidopsis Thaliana	MYB58 AT1G16490	myb domain protein 58	*Phaseolus vulgaris*	23		

RESULTS

Phenotyping under field conditions

Qualitative traits

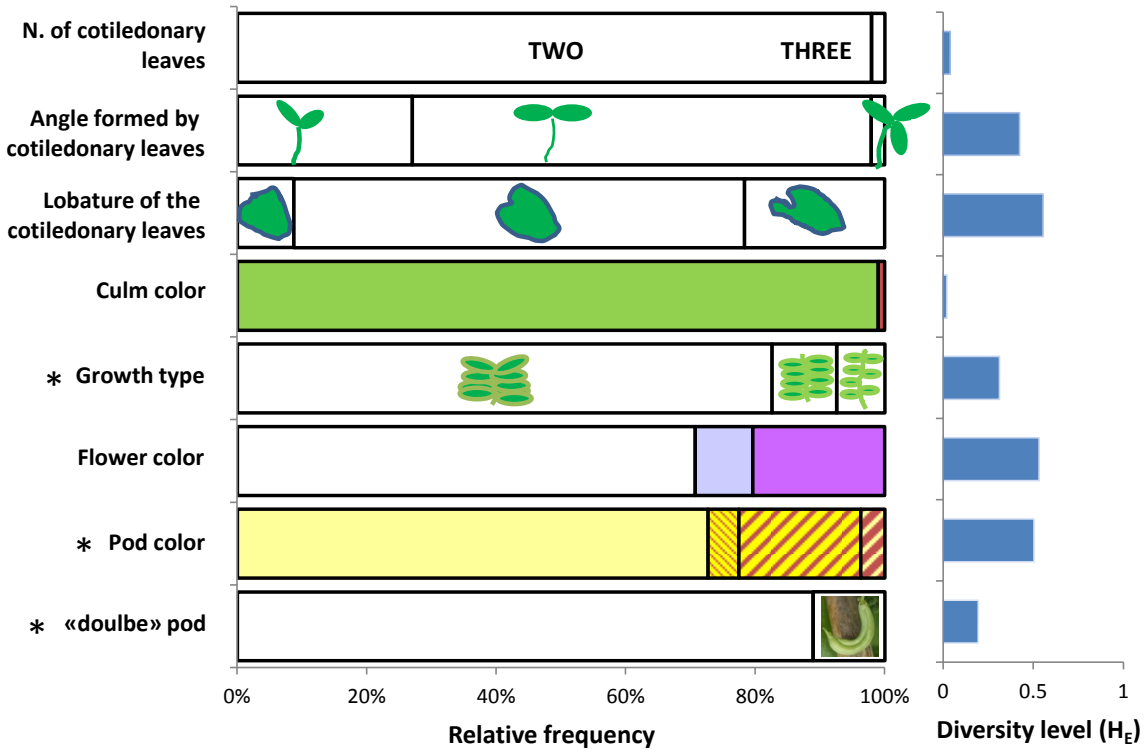
The results obtained for the eight qualitative traits examined are summarized in Figure 2.7. Overall, it has been observed high diversity levels for the traits "angle formed by cotyledonary leaves" and "leaf lobatures type", for "flower color" and "pod color", and with lesser extent for "growth habit" and "double pod". The remaining two traits ("number of cotyledonary leaves" and "stem color") were almost monomorphic with nearly all the plants showing two cotyledonary leaves and green stem.

The majority of lines, 71.1%, showed a 180° angle formed by the cotyledonary leaves while 25% of the plants showed an angle between 90°and 180°, with only six plants having three cotyledonary leaves forming 120°; the most frequent lobature type was the intermediate one (68.8%), followed by "deep lobature" (21.4%) and "absence of lobature" (8.6%).

Three flower colors were observed: white (as in MIDAS, 70.0%), light purple (8.9%), and purple (as in MG38, 20.3%) For the variable "pod color" two main types were recognized: uniform yellow (as in MIDAS; 72.7%) and yellow with red stripes (27. 3%). However, it was also noted that plants differed for the number and intensity of the stripes from low (4.8%), to medium (18.8%) and high stripped (3.7%).

The vast majority (82.6%) of the plants was classified as "non-climbing" (as MIDAS); however climbing types cumulatively represented the 18%, with 10% of the individuals classified as intermediate and 7.4% as climbing (as MG38). The 11%of plants displayed at least one double pod.

Figure 2.7 – Frequencies and diversity of eight qualitative traits recorded within the collection of bean lines. *traits that were mapped



The association between variables was tested. It was observed a strong association between flower color and pod color (Pearson χ^2=206.5, d.f.=6, P<10^{-4}; R^2 = 0.47).The majority of uniform yellow pods derived from white flowers while almost all with stripes pods were from purple and light purple flowers. The classification into light purple and purple was not predictive of the pod colors. Other associations of small strength were detected between :

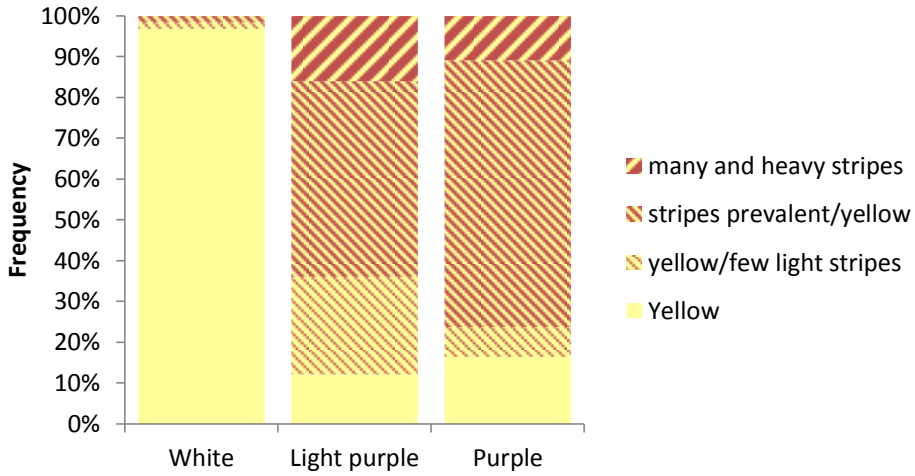
1) "lobature of the cotyledonary leaves" and "flower color" (Pearson χ^2=14.0, d.f.=6, P=0.030; R^2 = 0.03) with the sample of plants with deep leaf lobature enriched for light purple flowers;

2) "growth habit" and "variable angle formed by cotyledonary leaves" (Pearson χ^2 = 12.24, d.f. = 4; P = 0.016; R^2=0.038) with the class of plant having a 180° angle between cotyledonary leaves enriched for non-climbing types and the other two classes enriched for climbing types;

3) “stem color” and “flower color” (Pearson χ^2=11.9, d.f.=2, P=0.003; R^2 = 0.02) and “stem color” with “pod color” (Pearson χ^2=79.2, d.f.=3, P < 10^{-4}; R^2 = 0.05). Specifically, plants with colored stem always produced purple flower and pods with a high number of stripes. Figure 2.8

4) “double pod” and “growth habit” (Pearson χ^2=9.4., d.f.=2, P=0.009; R^2 = 0.04) with the climbing types enriched for double pod compared to the non-climbing type.

Figure 2.8 – Association between flower and pod color.



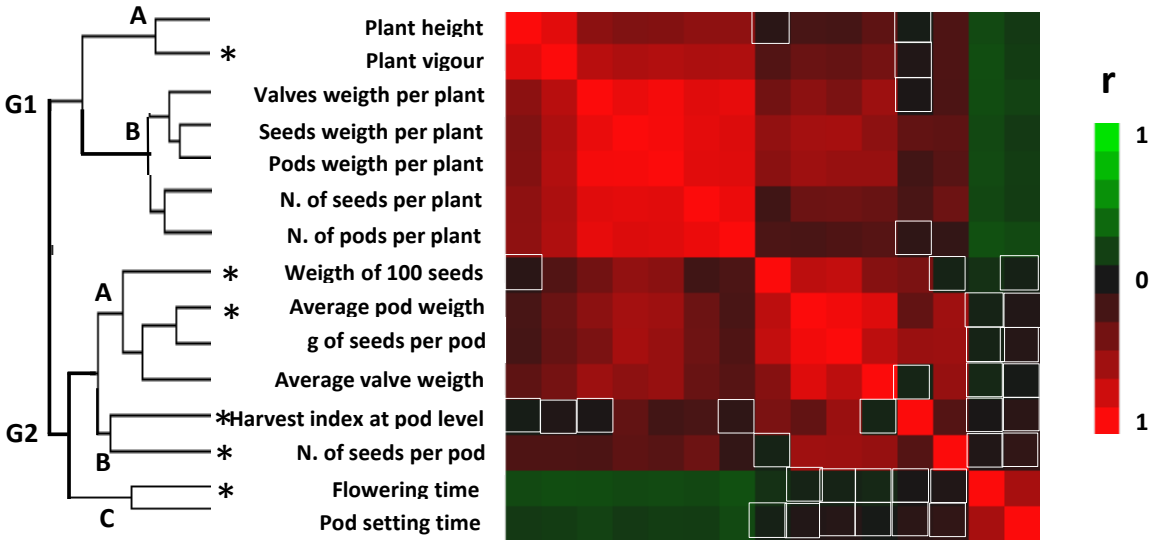
The successive mapping analyses will concentrate on pod color, “double pod”. This decision is based upon the observed level of genetic variation, the pattern of correlation among variables and the *a priori* decision of focusing the present research on the effects of domestication on pod characteristic. It will be also mapped “growth habit” to investigate the role of this trait when studying the genetic bases of the pod related traits.

Quantitative traits

The 15 morpho-phenological and productive quantitative traits showed a strong correlation structure (Figure 2.9). Indeed, among the 105 pairwise correlations among variables, 68 were significant with P < 0.0001, nine with P < 0.001, eight with P <0.01, two with P < 0.05 and only 18 (17.1%) were not significant (P>0.05). After correcting for a false discovery rates of 5% (Benjamini and Hochberg, 1995), 85 variables were still significant. Moreover, principal component analysis resulted in five PCs with eigenvalue > 1 that cumulatively explained the 89.6% of the total variance, with PC_1

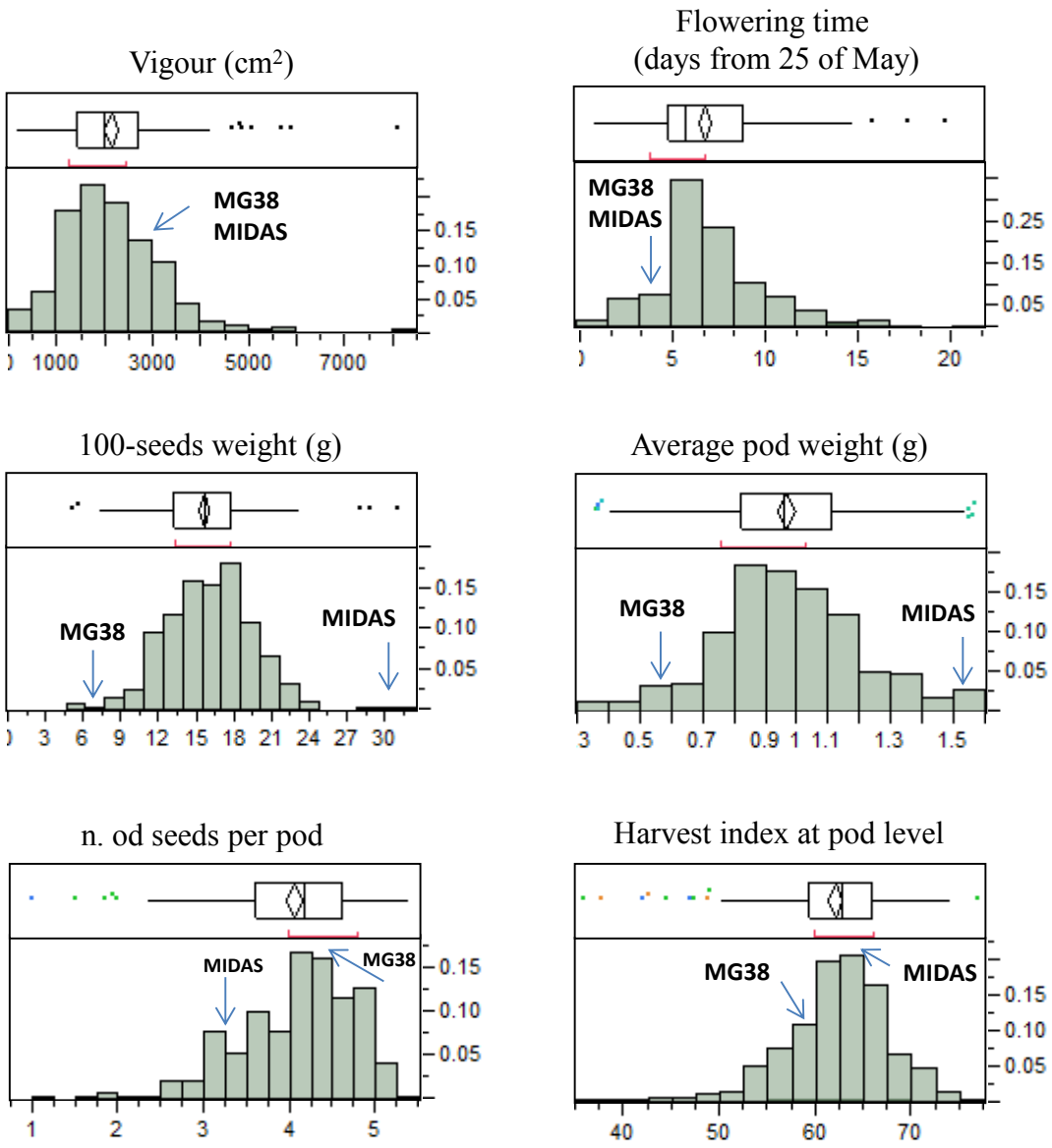
alone explaining 46.8% of the total variance. The clustering of the variables based on pairwise-correlations resulted in two main groups, G1 and G2, which further split into two (G1A and G1B) and three (G2A, G2B and G2C) subgroups for a total of five clusters (Figure 2.9). The majority of the significant correlations was positive in sign. Negative correlations (in general quite weak; Figure 2.9) were almost exclusively found for the phenological variables (G2C) *versus* plant vigor (G1A) and plant vigor-related traits (G1B) indicating that earlier lines tended to be more vigorous than the later ones. The other groups of variables, G2B and G2C, mainly comprised pod-related traits that were not significantly influenced by phenology.

Figure 2.9 – Correlation structure among the 15 morpho-phenological and productive traits recorded across 267 bean lines**.** Left: cluster analysis of the variables based on correlations. *traits that were mapped. Right: color map representing the pairwise correlation matrix among the variables.



Considering the present results and the objectives of the PhD thesis, the mapping analyses will be limited to the following four traits: average pod weight, weight of 100 seeds, number of seeds per pod, harvest index at pod level. It has been decided to also map plant vigor and flowering time to take into account these more general characteristics of the plant when studying the genetic bases of the pod-related traits. In Figure 2.10 are reported the distribution of these six quantitative traits.

Figure 2.10 – Frequency distribution for the six quantitative traits subjected to QTL analysis.



Precision phenotyping of pod size and shape

As expected the two parental lines showed contrasting characteristics for all of the six traits (Table 2.3). We observed good levels of variation for both shape and size (Table 2.3).

Table 2.3 – Results of the precise phenotyping of pod shape (measures are in pixels) for the two parental lines MG38and MIDAS. For each of them 10 pods of the three replicates were measured. The t tests were conducted considering for each replicate the means of 10 pods. *traits that were mapped

	MG38		MIDAS			
	mean	St. dev.	mean	St. dev.	Range	P (t test)
Perimeter	1600.6	20.6	3255.2	35.6	1654.6	10^{-4}
Area*	38149.0	2044.6	120873.0	1990.6	82724	10^{-4}
Maximum Width	226.3	26.6	274.2	54.2	47.9	0.046
Maximum Height	617.2	21.2	1445.3	38.1	828.1	10^{-4}
Curved Height	667.6	7.8	1437.4	37.5	769.8	10^{-4}
Maximum/curved height*	0.9	0.0	1.0	0.005	0.1	0.005

Based on the coefficient of variation (CV, Table 2.4) the trait that showed the lowest variation was the ratio maximum/curved height while the remaining five traits displayed relatively similar levels of variation.

Table 2.4 – Results of the precise phenotyping of pod shape (measures are in pixels). For each of the 267 lines, 10 pods were measured the means obtained.

	Mean	St. dev.	Min.	Max.	Range	C.V.
Perimeter	2351	286	1331	3218	1887	12.2
Area	72220	12040	21854	108928	87074	16.7
Maximum width	286	51	132	494	363	17.9
Maximum height	937	135	444	1290	846	14.5
Curved height	974	123	524	1318	794	12.6
Maximum/curved height	0.96	0.04	0.79	1.00	0.21	4.6

The highest values were observed for pod area that also displayed the largest difference between the minimum and the maximum values. The comparison between the ranges of variation observed for the parental lines with and the introgression lines evidenced that the parental lines occupy the opposite tails of the frequency distributions for all of the traits except for maximum width and maximum/curved height. However, these six variables showed a high level of redundancy. Indeed, the two principal components with eigenvalues > 1 capture the 94.5% of the total variance (PC_1=64.7% and

PC_2=29.8%). The PC_1 was strongly and positively correlated with perimeter, area, maximum height and curved height ($r > 0.95$ and $P<10^{-4}$ in all of the cases). The PC_2 was positively correlated with the pod maximum width ($r=0.89$, $P<10^{-4}$) and negatively correlated with the ratio maximum/curved height ($r = -0.81$; $P<10^{-4}$).Among the precision traits it has been decided to limit the mapping to pod area (as a measure of size) and maximum/curved height ration (as a measure of shape).

A close up on pod shattering

The percentage (%) of pods with valves that were, to some degree, separated was on average 49.6% with a range between 0% and 96.1% (Figure 2.11; Table 2.5). The indehiscent pods was on average 50.4%, varying from a minimum of 3.9% to a maximum of 100%. Twenty-nine individuals (about 10% of the total) had a percentage of non-shattering pods of 100%. It also appeared that the distribution of these two variables tends to be bimodal. The level of shattering (as % of shattered pod per plant) was on average 31.6% with a very high variation being comprised between 0% (all pods of the plant not shattered) and 82.6% (almost all pods shattered). The % of fissured pods was on average 18.0 % spanning from 0% to 71.7%. Dehiscent pods were further subdivided into twisting and not-twisting with mean values of 11.1% for the former and 20.1% for the latter though showing similar ranges of variation (from 0% to about 60% in both cases).

Figure 2.11 – Frequency distribution of the variables used to measure shattering in the analyzed population of common bean. All variable are expressed as percentage (%) of pod the fertile pods. Sample size is n=267 for all of the distributions.

Table 2.5 – Descriptive statistics for the variables used to measure pod shattering trait in the population of bean lines. All variables were expressed as percentage of the fertile pods.

Percentage of pods	Mean	Std. dev.	Mean SE	$L_{95\%}$	$U_{95\%}$
Completely indehiscent	50.4	26.7	1.6	53.6	47.2
Valves separated to some degree	49.6	26.7	1.6	52.8	46.4
Valves very slightly separated	18.0	13.4	0.8	19.7	16.4
Dehiscent	31.6	21.3	1.3	34.1	29.1
Twisting	11.5	10.9	0.7	12.8	10.2
Not-twisting	20.1	14.3	0.9	21.8	18.4

The % of shattered pods was more strongly correlated with the frequency of the not-twisting than with the frequency of the twisting types (Figure 2.12). However, in both cases the correlations were highly significant ($P<10^{-4}$). In particular, it has been noted that while a low number of twisting corresponded to quite different level of shattering, a low number of non-twisting pods was less predictive of an overall low level of shattering.

Figure 2.12 – Relationships between level and modality of shattering. Left: relationship with the frequency of twisting pods. Right: relationship with the frequency of not-twisting pods. Either linear and smoothing spline ($\lambda = 10000$) fits are presented. These analyses were performed excluding completely indehiscent plants.

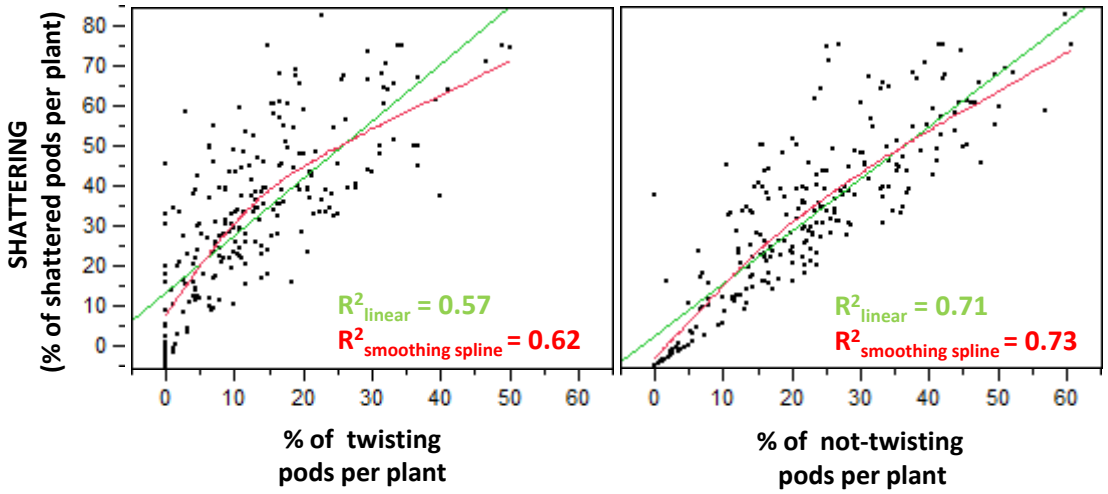
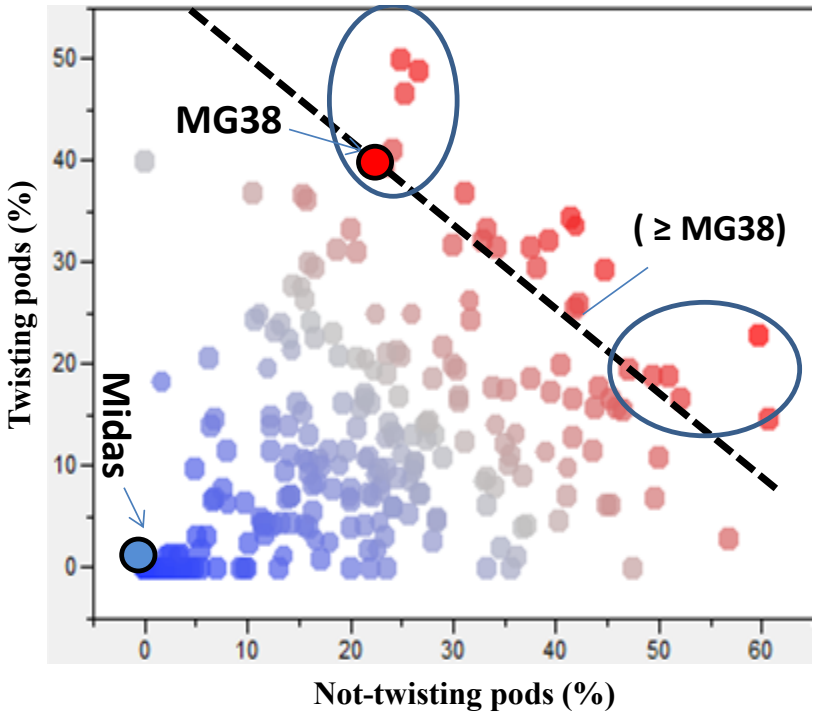


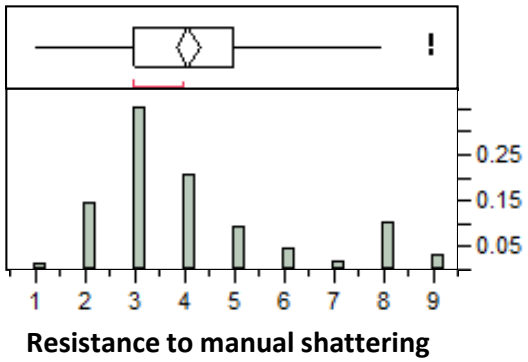
Figure 2.13 emphasizes that the analyzed segregant population was highly variable not only for the level but also for the modality of pod shattering and that bean lines with the same level of shattering could display very different proportion of twisting/not-twisting types. In particular, it was evident that with very high % of shattering (>70%) the ratio between not-twisting and twisting varied from 25:50 (1:2) to 60:15 (4:1). Moreover, the figure also suggests that transgressive variation occurred as about 10% of lines displayed higher shattering compared to MG38, the strongly shattering parental line.

Figura 2.13 – Relationship between level of shattering and frequency of twisting and not-twisting types. The color represents the level of shattering: deep blue = absence of shattering; red = high shattering. The dotted line separates bean lines with a level of shattering lower than or higher than MG38, the highly shattering parental line (~65%).



The traits "resistance to manual shattering" was characterized by the distribution depicted in Figure 2.14. The mean value for the trait was 4.12 (St. Dev. = 1.96; Mean S.E.=0.12). About 15% of the genotypes belong to the classes 1 and 2 (i.e. equal or even lower than MG38, strongly dehiscent) while about the 10 % of the genotypes fall in the classes 8 and 9 (equal to or even higher than MIDAS, completely not dehiscent). It is also interesting to note that this distribution appears bimodal.

Figure 2.14 – Frequency distribution of the resistance classes to manual shattering. Classes are from 1 (very low resistance) to 9 (very high resistance).



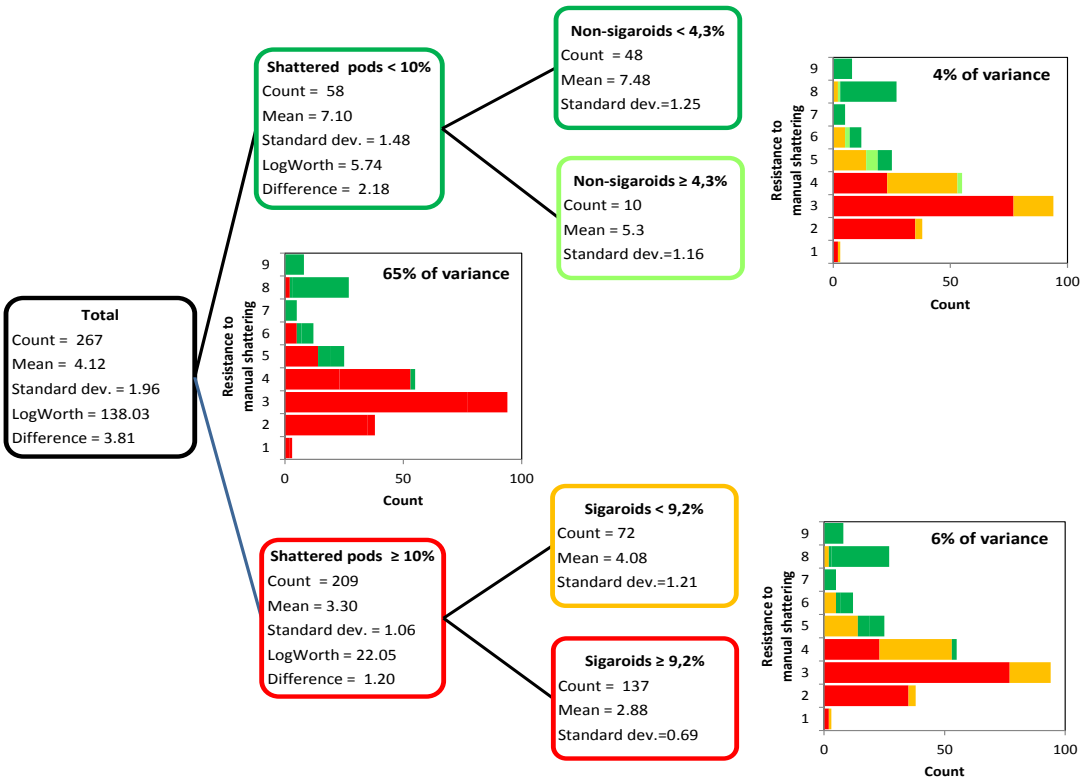
We also modeled the resistance to manual shattering based on the frequency and the modality of shattered pods. With this aim, the recursive partition analysis was applied. This showed that a first criterion for the classification of the bean lines into medium-

low or medium-high resistant to manual shattering was based on a frequency of shattered pods higher or lower than 10% (Figure 2.15). This partition explained 65.0% of the variance for the manual resistance to shattering trait.

The second partition indicated a role for the variable frequency of twisting pods. Indeed, among the genotypes with more than 10% of shattered pods, the analysis showed that a second criterion for the classification of the bean lines into very low or medium-low resistant to manual shattering was to have a level of twisting pods higher than or lower than 9.2%, respectively. This partition explained an additional significant but small portion (6%) of the resistance to manual shattering.

The third partition indicated a role for the variable frequency of not-twisting pods. Indeed, among the genotypes with less than 10% of shattered pods, a third criterion for the classification of the bean lines into medium or highly-very highly resistant to manual shattering was based on a level of not-twisting pods respectively higher than or lower than 4.3%. In this case, the partition explained an additional 4% of the total variance for the manual resistance to shattering. Cumulatively, the three partitions explained 75% of the total variance. The fourth partitions (not shown) explained only 0.8% of the total variance indicating that considering more complex models to explain the manual resistance to shattering was not necessary.

Figure 2.15 – Results of the recursive partition analysis Resistance varies from 1 (low resistance) to 9 (high resistance).



Relationships between pod shattering and other plant characteristics

In Table 2.6 were reported the correlations between the % of shattered pods with the 8 recorded qualitative traits and 15 quantitative traits of which 2 phenological and 13 morpho-productive. Overall, shattering was very poorly correlated with all of the 23 traits. Significant associations were detected for four qualitative traits (stem, flower and pod color and double pod) and for four productive traits namely valves weight per plant, average valves weight, weight of 100 seeds, and harvest index at pod level. After sequential Bonferroni correction for a FDR of 0.05 (Benjamini and Hochberg, 1995) all associations remained significant except for the color of the stem.

Specifically, for what concerns the associations with qualitative traits

- plants with red stem had higher shattering than those with green stem (59.6% ± S.E. 12.66% vs 31.25% ± 1.30).
- plants with white flowers had higher shattering (36.52 ± 1.44) than those with purple (22.85 ± 2.67) and light purple (13.66 ± 3.96) color
- yellow pod was associated to high shattering (35.89 1.44) compared to stripped/yellow (17.58 ± 2.8), with stripped (30.55 ± 6.33) and yellow/stripped (24.92 ± 5.77) in an intermediate positions.
- the 30 plants characterized by the occurrence of double pod showed higher shattering (43.1 ± 3.81) than the others (30.11 ± 1.36)

For what it concerns the associations with the quantitative traits, despite the effects are overall weak, it was observed that the valves weight per plant and the average valve weight tended to increase with shattering levels while the weight of 100 seeds and the harvest index at pod level tended to decrease with the increase of shattering levels. More specifically, the contrasts between the 236 dehiscent and the 29 indehiscent plants evidenced that the first group had a production of valves per plant and an average valve weight respectively 35.4% (t test: P=0.0085) and 26.2% (P=0.0004) higher than the second; to the opposite, the weight of 100 seeds and the harvest index at pod level were respectively 15.9% (P=0.0002) and 8.9% (P<0.0001) lower in the dehiscent than in indehiscent plants.

Table 2.6 – Correlations (Pearson r) between the percentage (%) of shattered pods of a plant with 23 morphological and productive traits. In **bold** are indicated the traits selected for mapping analyses. **P<0.01; ****P<$10^{-4.}$

TRAITS	ASSOCIATIONS			
QUALITATIVE TRAITS	**R^2_{adj}**	**M.S.**	**F**	**P**
No. of cotyledonary leaves	0.000	106.86	0.235	0.628
Angle of the cotyledonary leaves	0.000	99.21	0.21	0.81
Lobature of the cotyledonary leaves	0.012	672.75	1.481	0.229
Stem color	**0.020**	**2446.21**	**5.50**	**0.02**
Growth type	0.000	15.59	0.035	0.95
Flower color	**0.140**	**8396.23**	**21.45**	**<10^{-4}**
Pod color	**0.120**	**4708.32**	**11.75**	**<10^{-4}**
"double pod"	**0.037**	**4473.24**	**10.25**	**0.0015**
QUANTITATIVE TRAITS	**R^2**	**r**	**n**	**P**
Plant height	0.010	0.102	264	0.099
Plant vigor	0.010	0.099	263	0.107
Flowering time	0.000	0.00	266	0.998
Pod setting	0.001	0.034	264	0.058
Pods weight per plant[1] (g)	0.003	0.056	267	0.366
Valves weight per plant[2] (g)	**0.029**	**0.169**	**264**	**0.0061**
Seed weight per plant[3] (g)	0.000	-0.003	265	0.949
No. of pods per plant[4]	0.014	0.117	265	0.055
No. of seeds per plant[5]	0.010	0.100	266	0.103
Average pod weight[1/4]	0.001	-0.031	267	0.611
Average valves weight [2/4]	**0.071**	**0.267**	**264**	**<10^{-4}**
Weight of 100 seeds[3/5x100]	**0.045**	**-0.212**	**265**	**0.0005**
g. of seeds per pod[3/4]	0.014	-0.119	265	0.0523
No. of seeds per pod[5/4)]	0.006	0.77	266	0.211
Harvest index at pod level[3/1]	**0.106**	**-0.327**	**265**	**<10^{-4}**
Average across traits	0.028			

Correlations between shattering and pod size and shape.

All correlations between the % of shattered pods and the six morphometric variables describing size and shape of the pod were negative in sign and, despite not particularly strong (r between -0.132 and -0.452), they were all significant (P comprised between <0.05 and $P<10^{-4}$; Table 2.7). Thus, when shattering levels increased, pods tended to become smaller (with lower area, perimeter, shorter and narrower) and more curved (lower ratio curved/maximum height). Accordingly, when shattering level increased pods tended toward the characteristics of MG38, the parental lines carrying wild derived pod traits.

Table 2.7 – Correlations between the % of shattered pods of a plant and other pod traits measured with the software *Tomato analyser*.

Trait	R^2	r	n	P
Perimeter	0.089	-0.298	259	<0.0001
Area	0.022	-0.148	259	0.0171
Maximum Width	0.181	-0.425	259	<0.0001
Maximum Height	0.020	-0.143	259	0.0213
Curved Height	0.071	-0.266	259	<0.0001
Maximum/curved height	0.120	-0.347	259	<0.0001

Chemical elementary composition analysis of pod valves (CHN)

The analysis of variance (ANOVA), evidenced that the pod valves of the two parental lines MG38 and MIDAS had a different carbon content and that this difference was statistically highly significant (P<0.0001; Table 2.8 and 2.9). In particular, MG38 (highly dehiscent) had a carbon content (43.87 % of dry weight) that was on average 6.6% higher than for the indehiscent MIDAS (41.01; Table 2.9). The analysis also evidenced a marginally significant difference for hydrogen content (P<0.047; Table 2.8), again in favor of MG38 (6.69) compared to MIDAS (6.53; Table 2.9). The difference was not significant for nitrogen content albeit in favor of MIDAS (P=0.502; 0.58 vs 0.64 for MG38 and MIDAS, respectively; Table 2.8 and 2.9).

Table 2.8 – Results of ANOVA performed on chemical elementary analysis (CHN) data obtained for the pod valves of the two parental lines MG38 (dehiscent) and MIDAS (indehiscent).

	Mean Square	$F_{1,4}$	P
Carbon	12.304	256.00	<.0001
Hydrogen	0.037	8.08	0.047
Nitrogen	0.004	0.54	0.502

Table 2.9 – Mean content (% of dry weight) of carbon, hydrogen and nitrogen determined for the pod valves of the two parental lines (MG38) (dehiscent) and MIDAS (indehiscent).

	Level	Number	Mean	Std Error	$L_{95\%}$	$U_{95\%}$
Carbon	MIDAS	3	41.01	0.13	40.66	41.36
	MG38	3	43.87	0.13	43.52	44.22
Hydrogen	MIDAS	3	6.53	0.04	6.42	6.64
	MG38	3	6.69	0.04	6.58	6.80
Nitrogen	MIDAS	3	0.64	0.05	0.50	0.77
	MG38	3	0.58	0.05	0.45	0.72

These differences were tested and the same analyses were extended to the introgression lines (Table 2.10 and 2.11). The contrast indehiscent vs. dehiscent lines was highly significant. More in detail the ILs with shattering displayed 6.92% more carbon than those characterized by no shattering. Moreover, the group of shattering lines always showed a carbon content ~2% higher than MG38 while the difference between MIDAS and indehiscent or "nearly" indehiscent lines was lower. In Figure 2.16 the frequency distributions for carbon, hydrogen and nitrogen content are given. It is apparent that the distribution for carbon content was bimodal.

Table 2.10 – Results of one way ANOVA performed on chemical elementary analysis (CHN) data obtained for the comparison indehiscent vs. dehiscent lines.

	Indehiscent vs. dehiscent			
	Mean Square	$F_{1,227}$	P	R^2_{adj}
Carbon	191.93	105.22	<.0001	0.31
Hydrogen	0.72	6.69	0.01	0.02
Nitrogen	0.1	2.34	0.13	0.01

Table 2.11 – Mean content (% of dry weight) of carbon, hydrogen and nitrogen determined for the contrasts indehiscent vs. dehiscent lines.

	Level	Number	Mean	Δ (%)	S.E.	$L_{95\%}$	$U_{95\%}$
Carbon	Not-dehiscent	26	41.68	6.92	0.26	41.15	42.21
	dehiscent	203	44.57		0.1	44.38	44.75
Hydrogen	Not- dehiscent	26	6.33	2.79	0.07	6.19	6.47
	dehiscent	203	6.51		0.02	6.46	6.55
Nitrogen	Not- dehiscent	26	0.54	-11.91	0.04	0.45	0.63
	dehiscent	203	0.48		0.01	0.45	0.5

The study of the relationship between carbon content and frequency of shattered pods per plant (Figure 2.17) revealed that there was an abrupt transition in carbon content between 5% - 10% of shattered pods per plants (Figure 2.17 A).

Figure 2.16 – Frequency distribution for carbon, hydrogen and nitrogen content (% of dry weight) within the population of bean introgression lines.

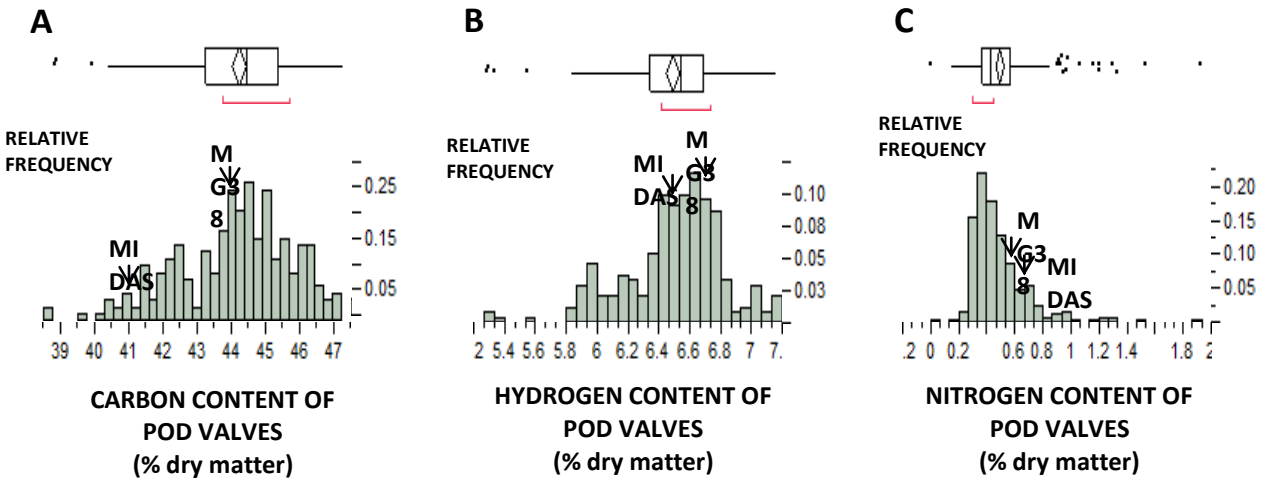
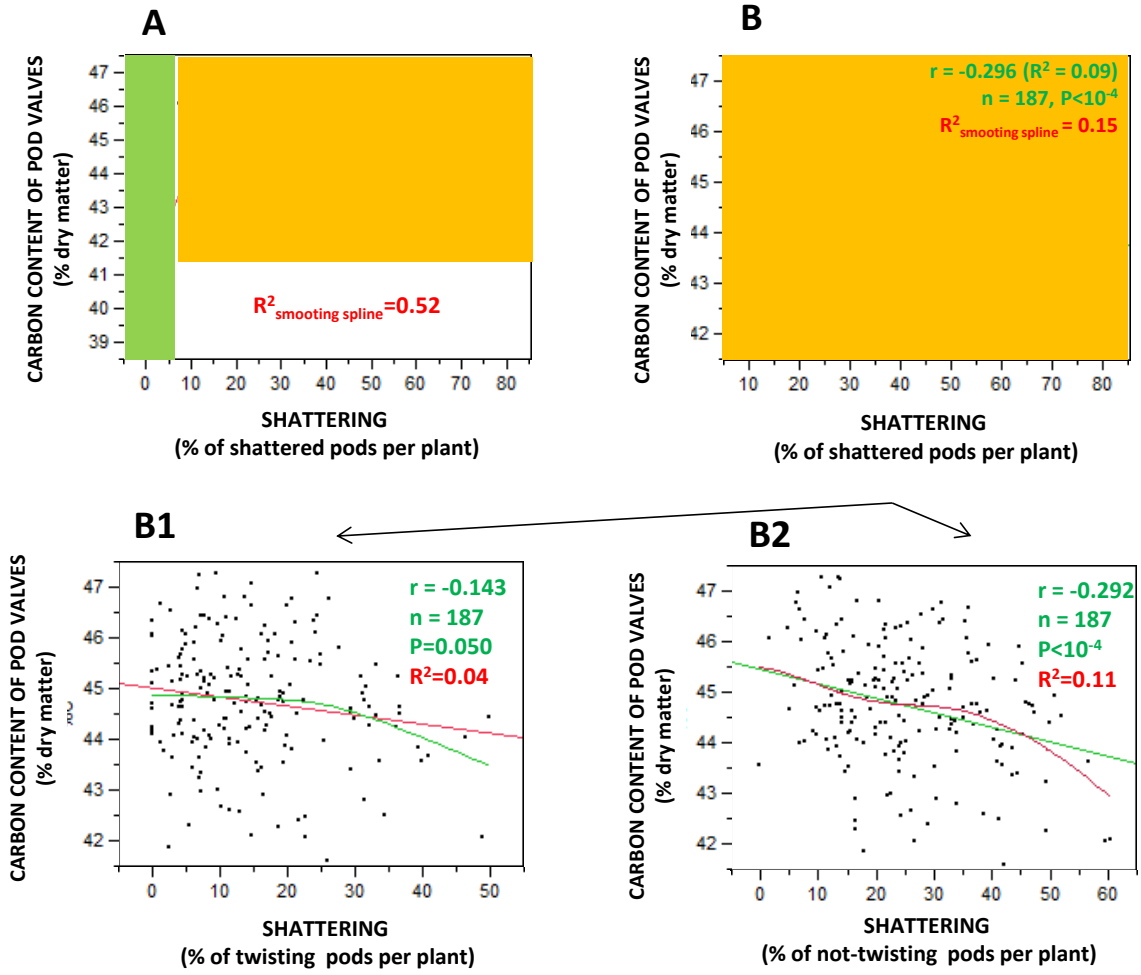


Figure 2.17 – A) Relationships between carbon content and frequency of shattered pods per plant. The green area comprises individuals for which shattering is ≤7.14% while the orange area those individuals for which shattering is > 7.14%. **B)** Relationship between carbon content and shattering level excluding individuals with low or no

shattering, also distinguishing between twisting (**B1**) and not twisting (**B2**) pods. R^2 values are given for smoothing spline ($\lambda = 10000$) (red) and for linear fits (green).



Partition analysis revealed that this value is indeed 7.14%. The classification of the bean lines into < 7.14% vs. ≥ 7.14% of shattering pods lead to an increase of the variance explained for carbon content, from 31% (Table 2.10) to 47% (Table 2.12) as also to an increase of the differences between groups from 6.92 (Table 2.11) to 6.97 (Table 2.13). Furthermore, if the lines with <7.14 of shattering pods are excluded from the analysis, a weak but significant negative correlation between shattering level and carbon content was observed (Figure 2.17 B). Moreover this tendency was stronger for the % of not-twisting than for the % of twisting pods (Figure2.17 C and D).

Table 2.12 – Results of one way ANOVA performed on chemical elementary analysis (CHN) data obtained for the lines with <7.14% of shattered pods vs. lines with > 7.14% of shattered pods.

	<7.14% vs. ≥ 7.14% of shattered pods			
	Mean Square	**$F_{1,227}$**	**P**	**R^2_{adj}**
Carbon	286.04	202.92	<.0001	0.47
Hydrogen	0.29	2.63	0.11	0.01
Nitrogen	0.05	1.34	0.25	0.00

Table 2.13 – Mean content (% of dry weight) of carbon, hydrogen and nitrogen determined for the pod valves of 240 introgression lines.

	Level	**Number**	**Mean**	**Δ (%)**	**S.E.**	**$L_{95\%}$**	**$U_{95\%}$**
Carbon	< 7.14%	41	41.84	6.97	0.19	41.48	42.21
	> 7.14%	188	44.76		0.09	44.59	44.93
Hydrogen	< 7.14%	41	6.41	1.44	0.06	6.3	6.53
	> 7.14%	188	6.51		0.02	6.46	6.55
Nitrogen	< 7.14%	41	0.52	-7.81	0.03	0.46	0.58
	> 7.14%	188	0.48		0.01	0.45	0.51

Fiber analysis of pod valves.

The analysis of variance (ANOVA), evidenced that the pod valves of the two parental lines MG38 and MIDAS had a different fiber content (NDF) and that this difference was statistically very highly significant (P<0.0001; Table 2.14 and 2.15). In particular, MG38 (highly dehiscent) had a fiber content (62.0%) that was on average ~48% higher than for MIDAS (indehiscent) (42.0%; Table 2.15). The analysis also evidenced strong differences for the three cell wall components, in the decreasing order: lignin > hemicelluloses > cellulose again in favor of MG38 compared to MIDAS (Table 2.14 and 2.15).

Table 2.14 – Results of ANOVA performed on data from fiber analysis of pod valves of the two parental lines MG38 (dehiscent) and MIDAS (indehiscent).

Cell wall components	Mean Square	$F_{1,4}$	P	R^2_{adj}
NDF (fibers)	600.20	98.74	$<10^{-3}$	0.95
ADF	172.40	173.06	$<10^{-3}$	0.97
ADL (lignin)	9.51	163.04	$<10^{-3}$	0.97
NDF-ADF (emicellulose)	129.25	14.01	0.02	0.72
ADF-ADL (cellulose)	100.91	134.07	$<10^{-3}$	0.96

Table 2.15 – Mean content (% of dry weight) of neutral detergent fiber (NDF), acid detergent fiber (ADF), ADL (lignin), hemicellulose, cellulose, determined for the pod valves of the two parental lines (MG38) (dehiscent) and MIDAS (indehiscent).

Cell wall components	Level	Number	Mean	Δ (%)	S.E.	$L_{95\%}$	$U_{95\%}$
NDF (fibers) (%)	MIDAS	3	42.01		2	33.41	50.62
	MG38	3	62.01	47.62	0.23	61.03	63.00
ADF (%)	MIDAS	3	30.33		0.47	28.3	32.37
	MG38	3	41.06	35.34	0.66	38.2	43.91
ADL (lignin) %	MIDAS	3	3.08		0.19	2.25	3.92
	MG38	3	5.6	81.68	0.03	5.47	5.73
NDF-ADF (emicellulose) %	MIDAS	3	11.68		2.32	1.68	21.68
	MG38	3	20.96	79.5	0.87	17.23	24.69
ADF-ADL (cellulose) %	MIDAS	3	27.25		0.3	25.95	28.55
	MG38	3	35.45	30.1	0.64	32.7	38.21

The analysis was then extended to two groups each constituted of 12 lines one completely indehiscent and one with a % of shattered pods higher than MG38 (> 65%). Again, it was observed that the two groups were different for all of the chemical

fractions (Table 2.16). However, compared to the two parental lines (Table 2.15) differences in NDF, hemicellulose and cellulose were all strongly reduced. In sharp contrast, the difference in lignin content registered a sharp increase (Table 2.17).

Table 2.16 – Results of ANOVA performed on data from fiber analysis of pod valves. Two groups of 12 lines one completely indehiscent and one with a % of shattered pods higher than MG38 (>65%) were compared.

Cell wall components	Mean Square	$F_{1,22}$	P	R^2_{adj}
NDF (fibers)	825.15	273.04	<.0001	0.92
ADF	269.61	102.00	<.0001	0.81
ADL (lignin)	115.74	305.35	<.0001	0.93
NDF-ADF (emicellulose)	151.43	140.95	<.0001	0.86
ADF-ADL (cellulose)	32.05	14.64	0.0009	0.37

Table 2.17 – Mean content (% of dry weight) of neutral detergent fiber (NDF), acid detergent fiber (ADF), ADL (lignin), hemicellulose, cellulose, determined for two groups of 12 lines one completely not dehiscent (ND) and one highly dehiscent (HD; with a % of shattered pods higher than MG38).

Cell wall components	Level	Number	Mean	Δ (%)	S.E.	$L_{95\%}$	$U_{95\%}$
NDF (fibers) (%)	ND	12	45.33		0.5	44.29	46.37
	HD	12	57.05	25.87	0.5	56.01	58.1
ADF (%)	ND	12	30.52		0.47	29.54	31.49
	HD	12	37.22	21.97	0.47	36.25	38.19
ADL (lignin) %	ND	12	2.43		0.18	2.06	2.8
	HD	12	6.82	180.84	0.18	6.45	7.19
NDF-ADF (emicellulose) %	ND	12	14.81		0.3	14.19	15.43
	HD	12	19.84	33.92	0.3	19.21	20.46
ADF-ADL (cellulose) %	ND	12	28.09		0.43	29.51	31.28
	HD	12	30.04	7.6	0.43	27.2	28.97

Figure 2.18 – Comparison between two groups of 12 lines one completely not dehiscent and one highly dehiscent (with a % of shattered pods higher than MG38) for cell wall components. Green: MIDAS; Red: MG38. Ligh blue: introgression lines.

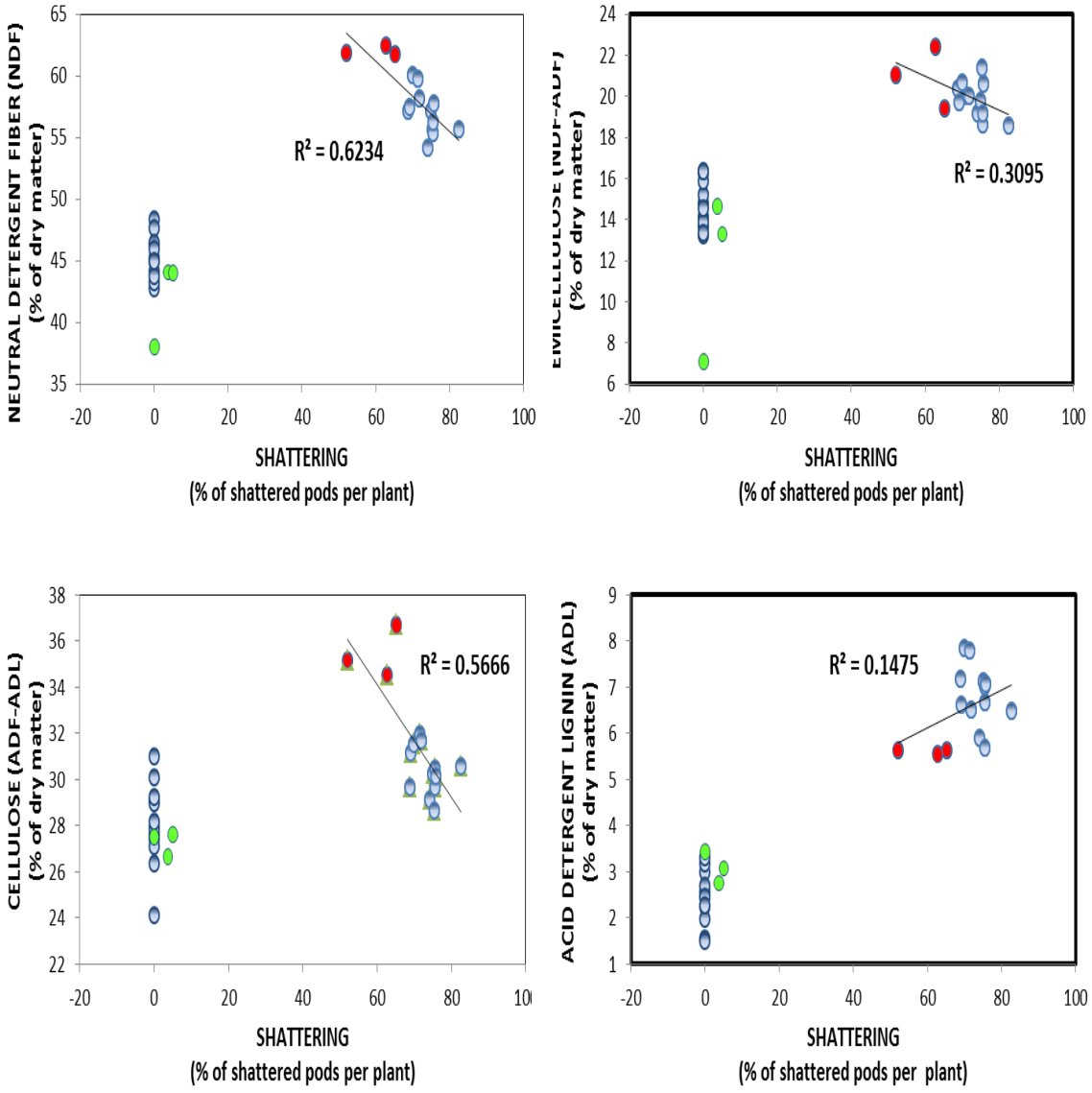
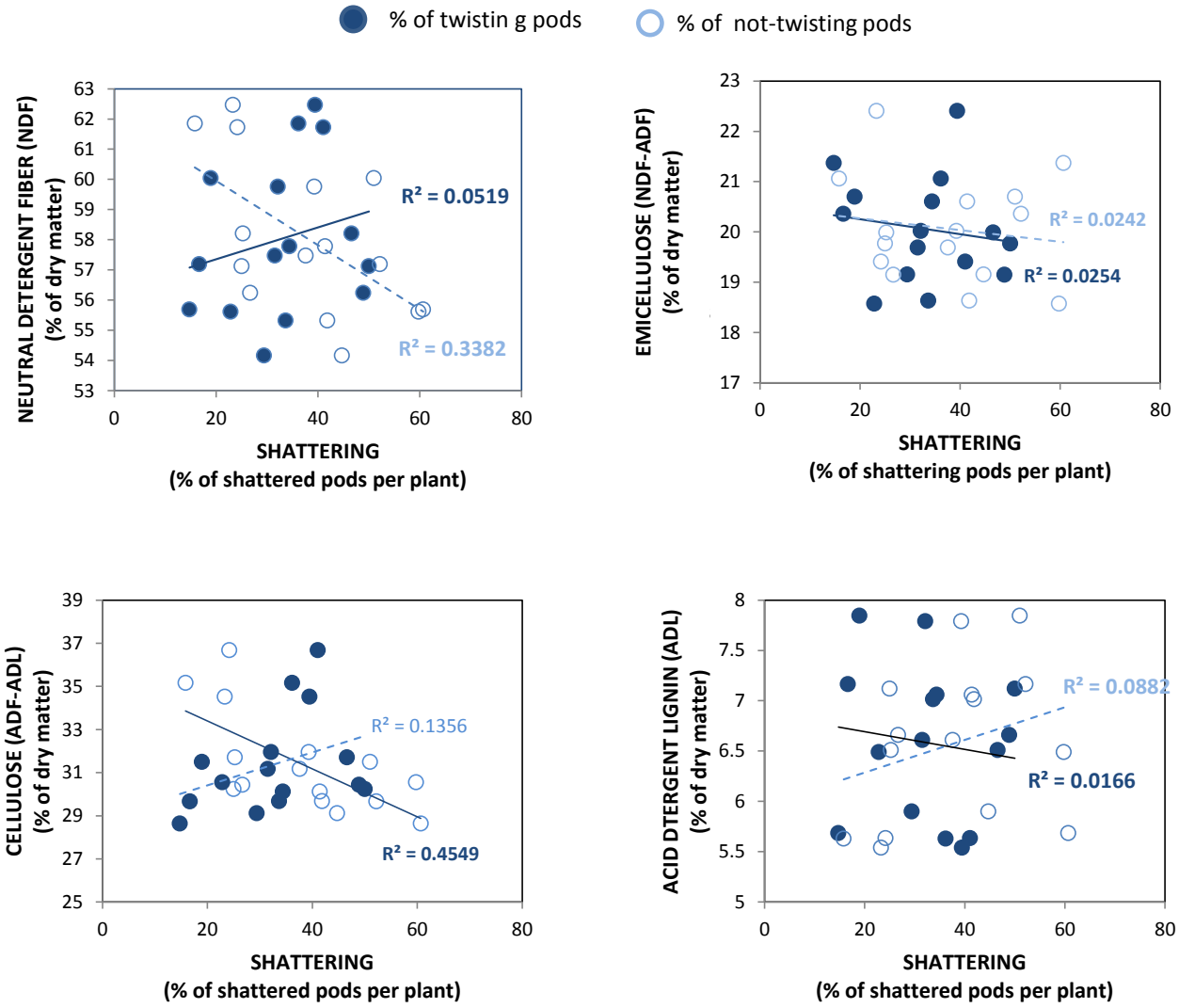


Figure 2.18 clear-up the relationships between parental lines and introgression lines for cell wall components: the achievement of very high pod shattering (even higher than MG38) is characterized by a high ADL/NDF ratio, i.e. a strong fraction of lignin in the cell wall. Furthermore, the frequency of twisting pods and of not-twisting pods responded to the variation in NDF, cellulose and lignin in opposite directions. In

particular, it resulted that the frequency of not-twisting pods, but not the frequency of twisting pods, was associated with the reduction of cellulose content (Figure 2.19).

Figure 2.19 – Dependency of shattering modality from cell wall composition.



Finally we also found that carbon content is strongly correlated with the total fiber content (NDF) of pod valves (Table 2.18). Moreover, among the three components (lignin, hemicellulose and cellulose) the strongest correlation is with lignin.

Table 2.18 – Correlation between elementary composition (CHN) and cell-wall fiber content in pods valves (N=30).

	%C	%H	%N
NDF (fibers) (%)	0.745****	0.108	0.122
ADL (lignin) %	0.697****	0.097	0.19
NDF-ADF (hemicellulose) %	0.682****	0.15	0.029
ADF-ADL (cellulose) %	0.577***	0.022	0.141

When a stepwise multiple regression analysis was performed considering carbon content as dependent variable and lignin, hemicellulose and cellulose as independent variables, the only variable entering in the model was lignin content (Table 2.19). This indicates that hemicellulose and cellulose are correlated to carbon content mainly because of their correlation with lignin content.

Table 2.19 – Results of the stepwise multiple regression (method: forward selection) considering carbon content as dependent variable and lignin, hemicellulose and cellulose as independent variables.(d.f.= degree of freedom, S.S. Sum of Square)

Parameter	Estimate	d.f.	S.S.	F	P
Intercept	40.545	1	0	0	1
NDF-ADF (hemicellulose) %	0	1	2.63	2.60	0.118
ADF-ADL (cellulose)	0	1	3.25	3.29	0.081
ADL (lignin) %	0.459	1	28.31	26.51	1.85×10^{-5}

Anatomical and histological analysis of pod valves

The analysis was conducted with five-days and 21-days old pods.

Five-days-old pods (Figure 2.20– Panel n.1). The ventral sheath did not display very evident lignin deposition. Only few cells showed modest levels of lignification. Similarly there was not lignin deposition on the fiber along the inner layer of pod valves. No obvious differences between the two parental lines were observed, albeit the size and the shape of the cell of the future ventral sheath were bigger and rounded in MG38 than in MIDAS.

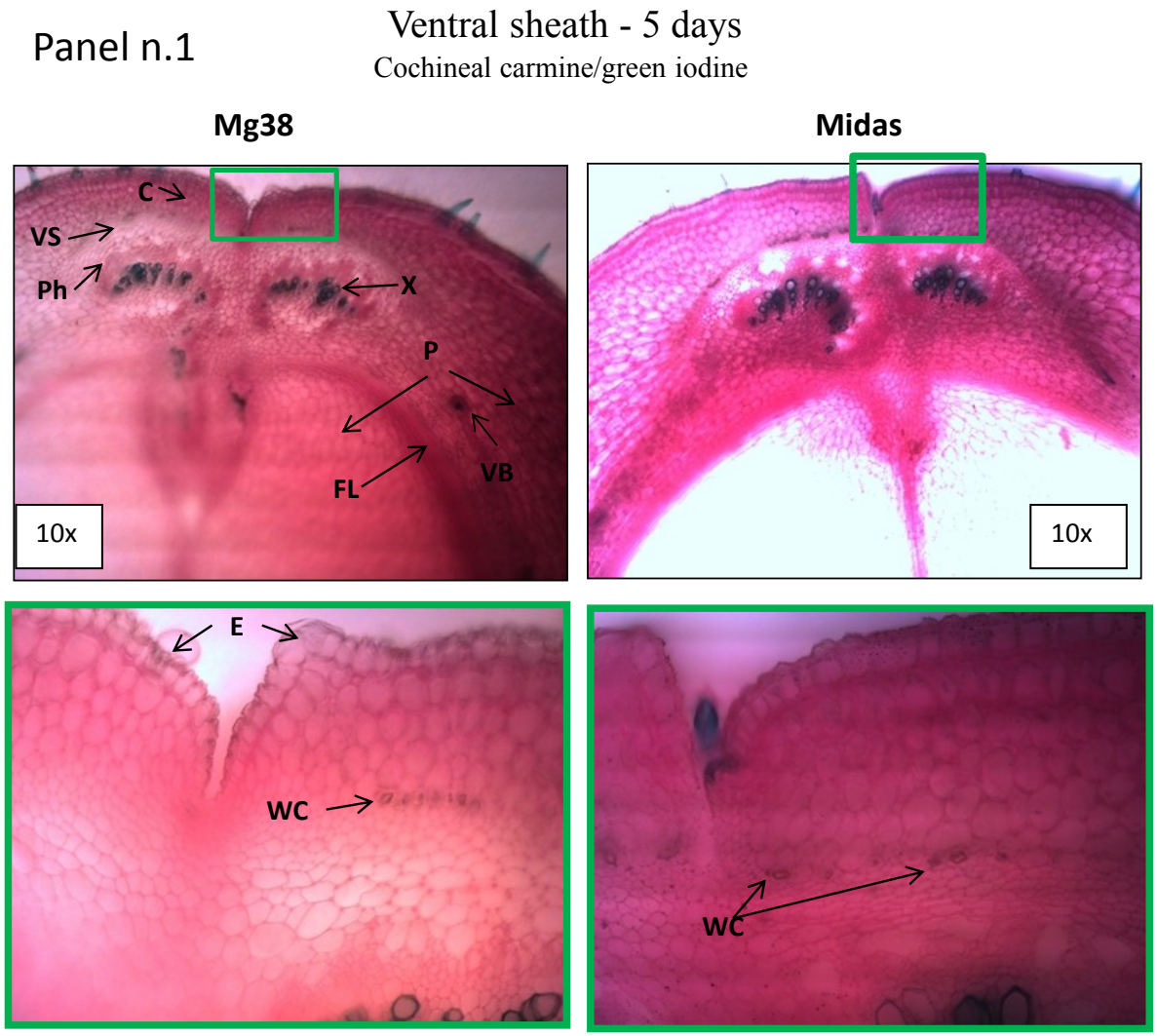
20-days-old pods. The ventral sheath displayed very evident lignin deposition (Panel 2 and 4). The proportion between the number of lignified fiber and number of wood cells was clearly in favor of MG38 compared to MIDAS. In MG38 wood cells were limited to the external part of the ventral sheath and to the dehiscence zone, while in MIDAS almost all of the cells of the ventral sheath were wood cells. In MG38 the thickness of the cell-wall tended to reduce when moving from the sheath to the dehiscence zone (Panel n. 2 and Panel n. 5), where an easy tendency to "fracture" was also observed (Panel n. 3). A difference in the number of fibers compared to wood cells was also observed for the dorsal sheath (Panel n. 7). A clear-cut difference among the two parental lines was observed for the degree of lignification in the inner layer of fibrous cell with very strong lignification in MG38 and the complete absence of lignin deposition in MIDAS.

Coherently, similar patterns were observed when two introgression lines (one with no shattering ability and one with a shattering ability higher than MG38) were compared (Panel n. 6, n. 7 and 8).

It has been difficult to obtain good sections at the maturation stage because of the fragility of the tissues. However, it has been noted that at this stage the ventral sheath of MIDAS had more mechanical resistance while that of MG38 appeared very fragile (Panel n. 9).

Overall this indicates that the differences in carbon content and more in particular in fiber content (as measured by NDF), and lignin content (as measured by ADL) were attributable to a differential degree of lignification of the structures in the correspondence of the ventral (and the dorsal) sheath and, even more evidently, in the inner layer of the pod valves.

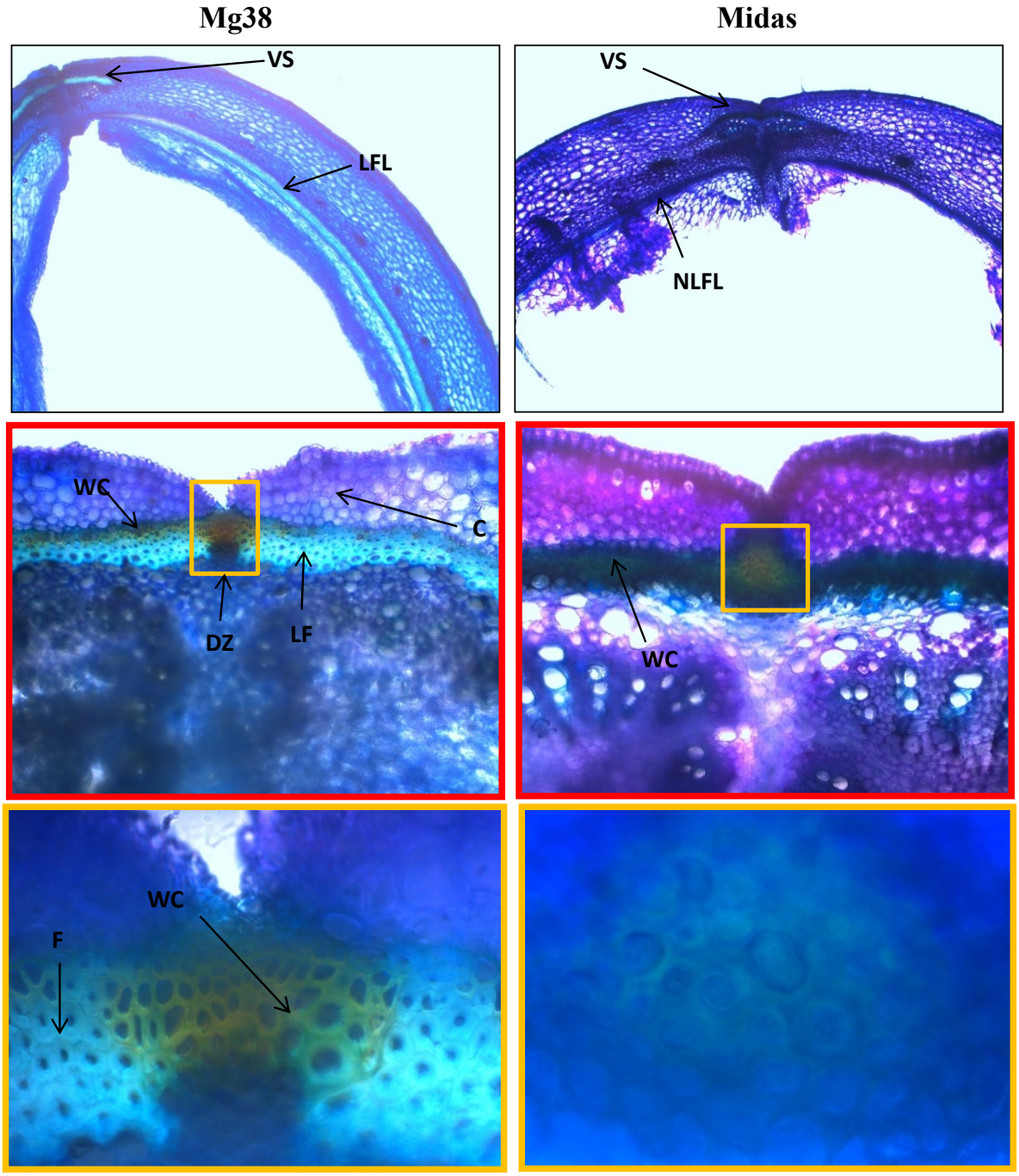
Figure 2.20 – Panel 1-9. Results of the anatomical and histological study. Legend: C=collenchyma; DS= dorsal sheat; DZ= dehiscence zone; E =epidermis; FL = fibrous layer; LF= lignified fibers LFL = lignified fibrous layer; NLFL= not lignified fibrous layer; P= parenchyma; Ph=phloem; VB = vascular bundle; VS=ventral sheath; WC= wood cells; X = xilem;



Inner fibrous layer and Ventral Sheath - 20days

Toluidine blu staining

Panel n.2

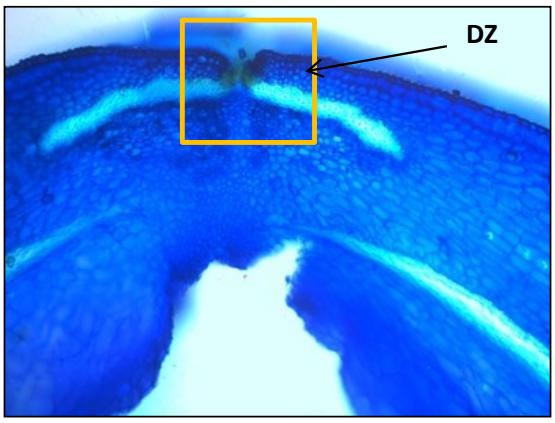


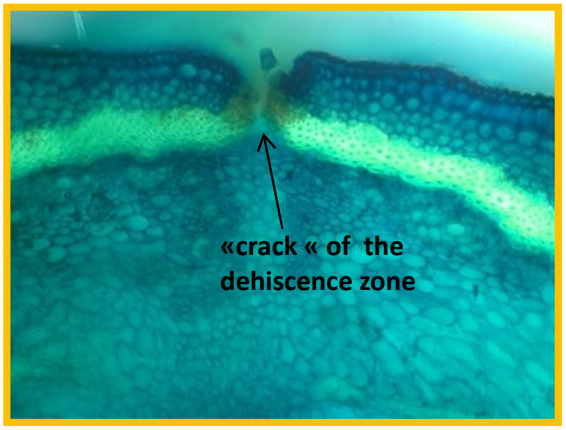
Panel n.3

Ventral Sheath - 20days

Toluidine blu staining

Mg38

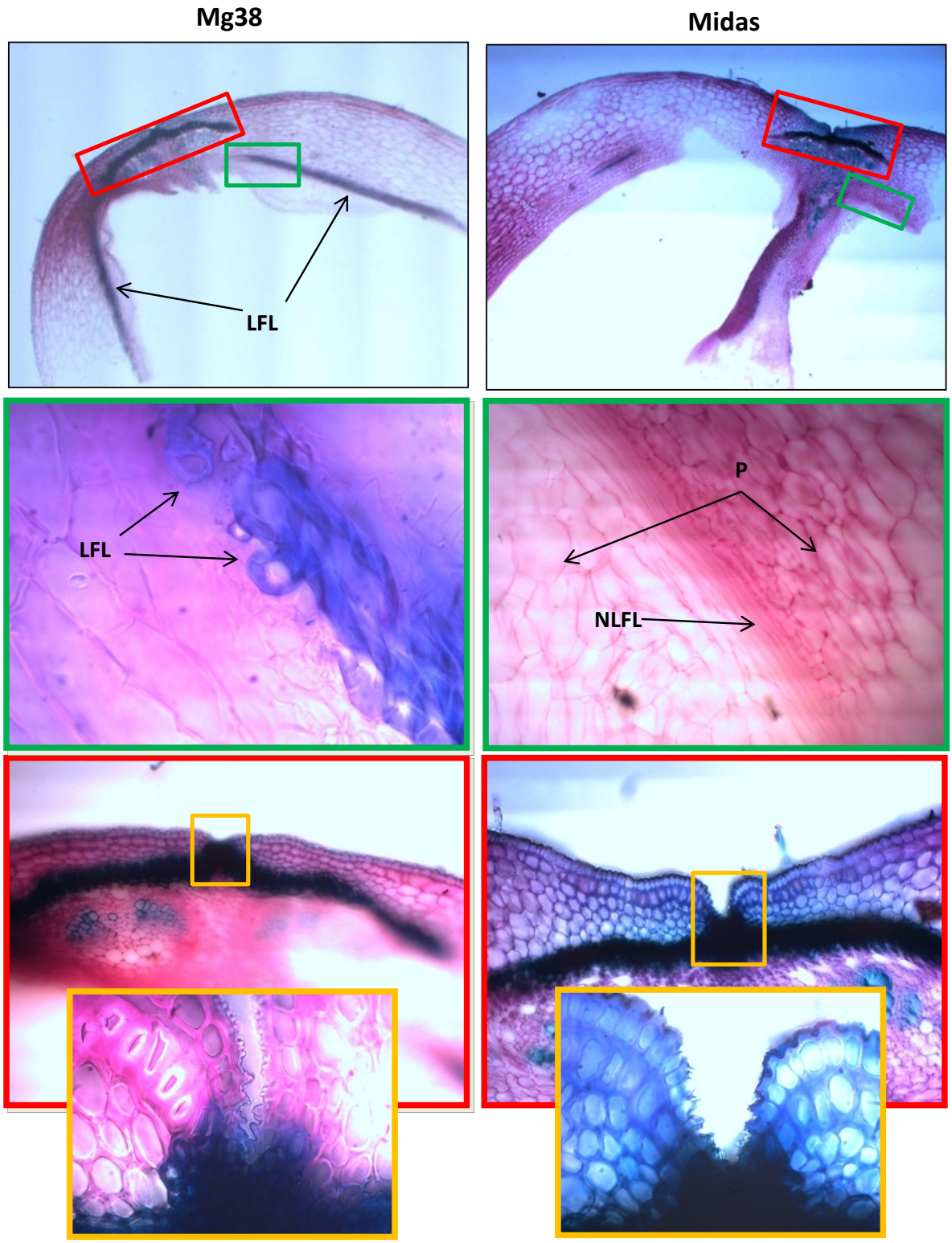




Inner fibrous layer and Ventral Sheath - 20days

Cochineal carmine/green iodine

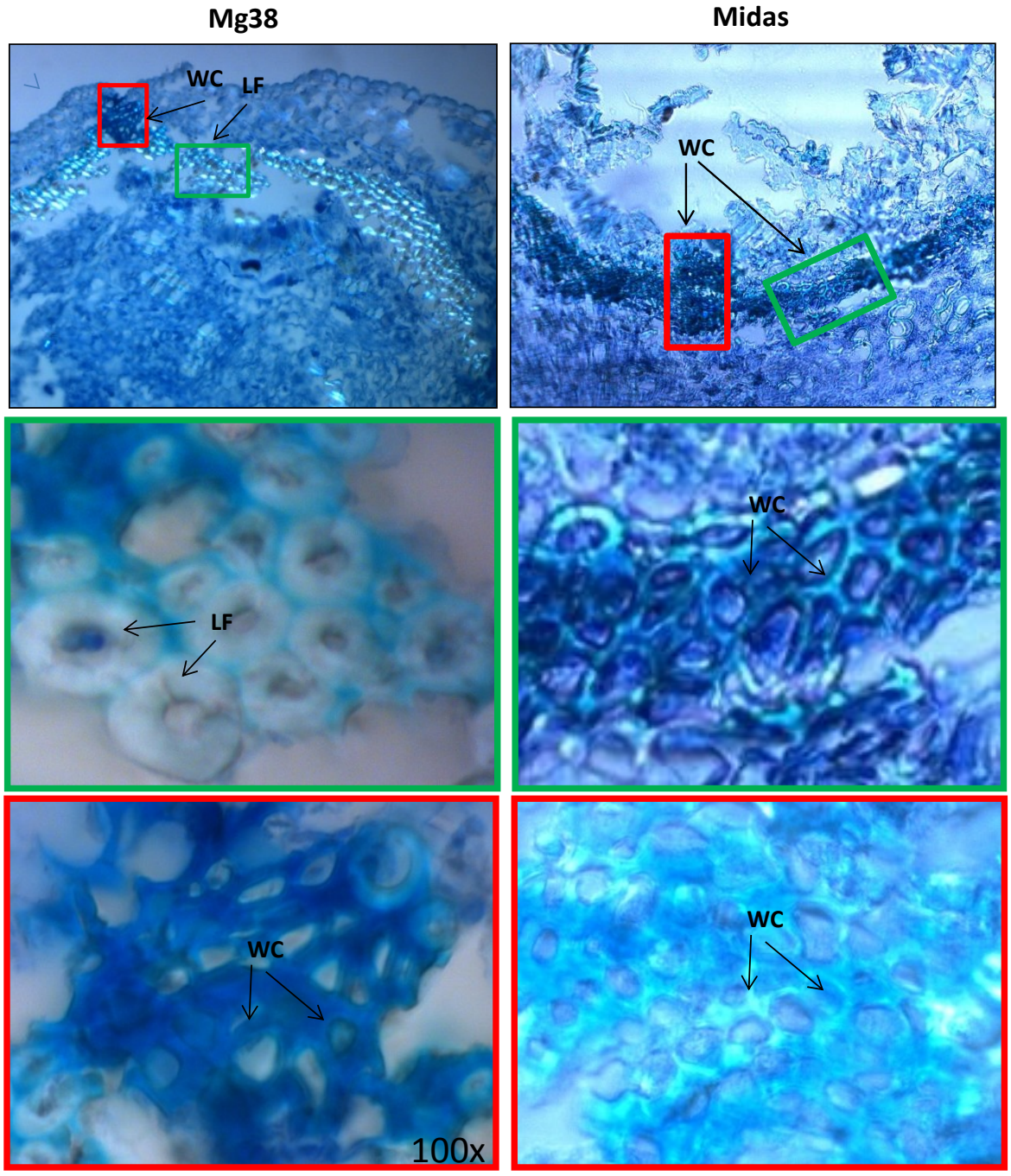
Panel n.4



Ventral Sheath - 20days

Paraffin inclusions - Toluidine blu staining

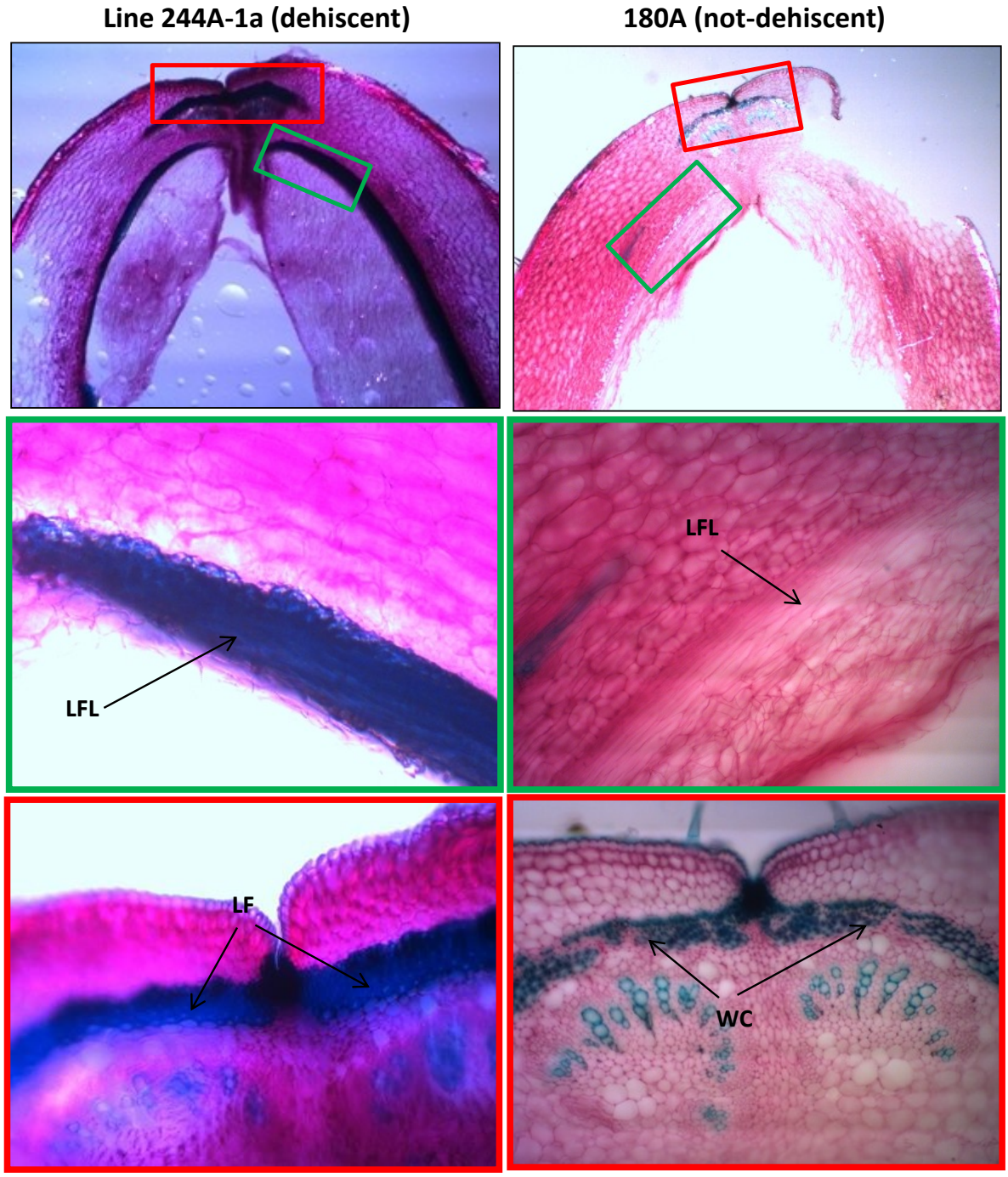
Panel n.5



Inner fibrous layer and Ventral Sheath - 20days

Cochineal carmine/green iodine

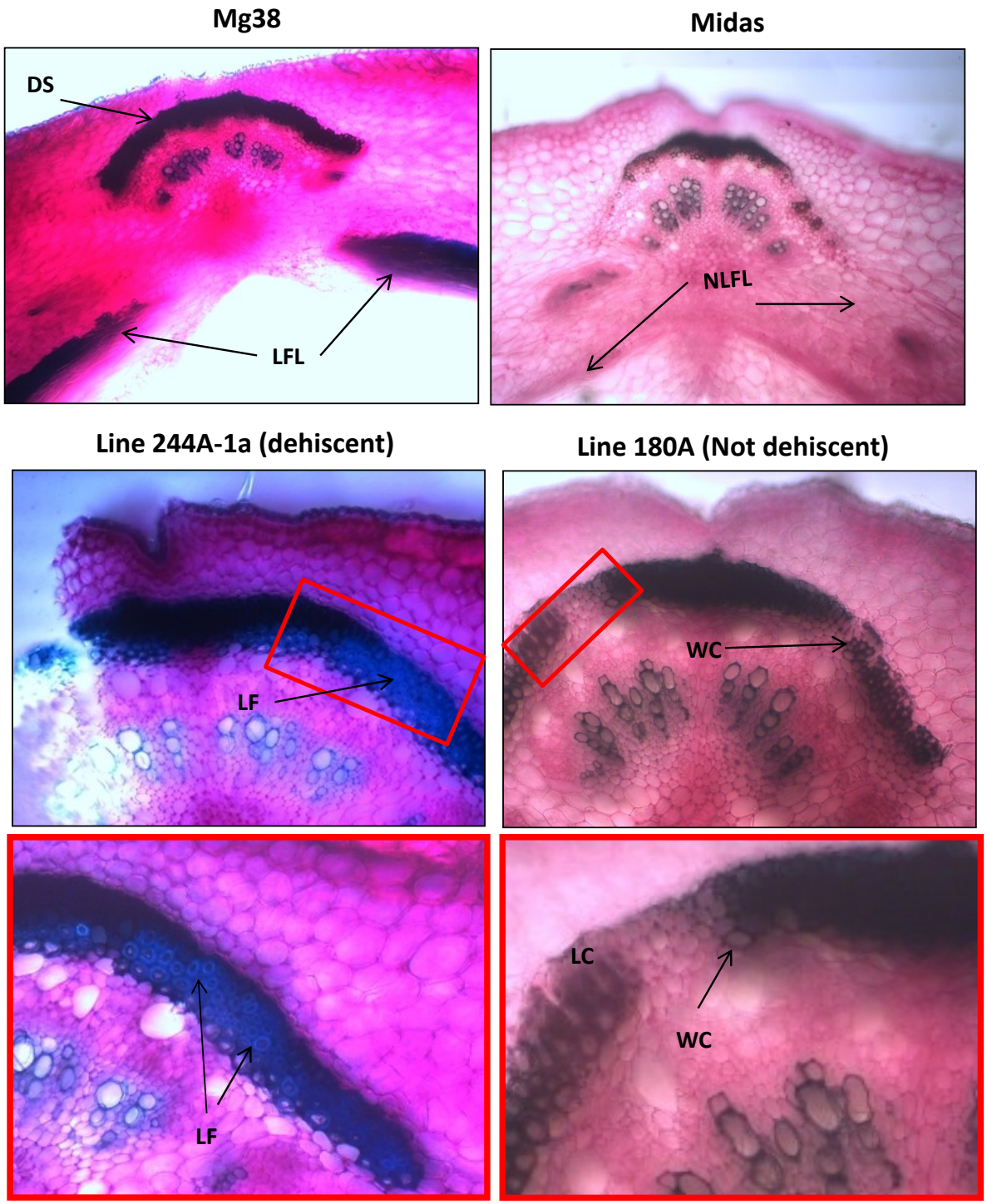
Panel n.6



Inner fibrous layer and Dorsal SheatH - 20days

Cochineal carmine/green iodine

Panel n.7

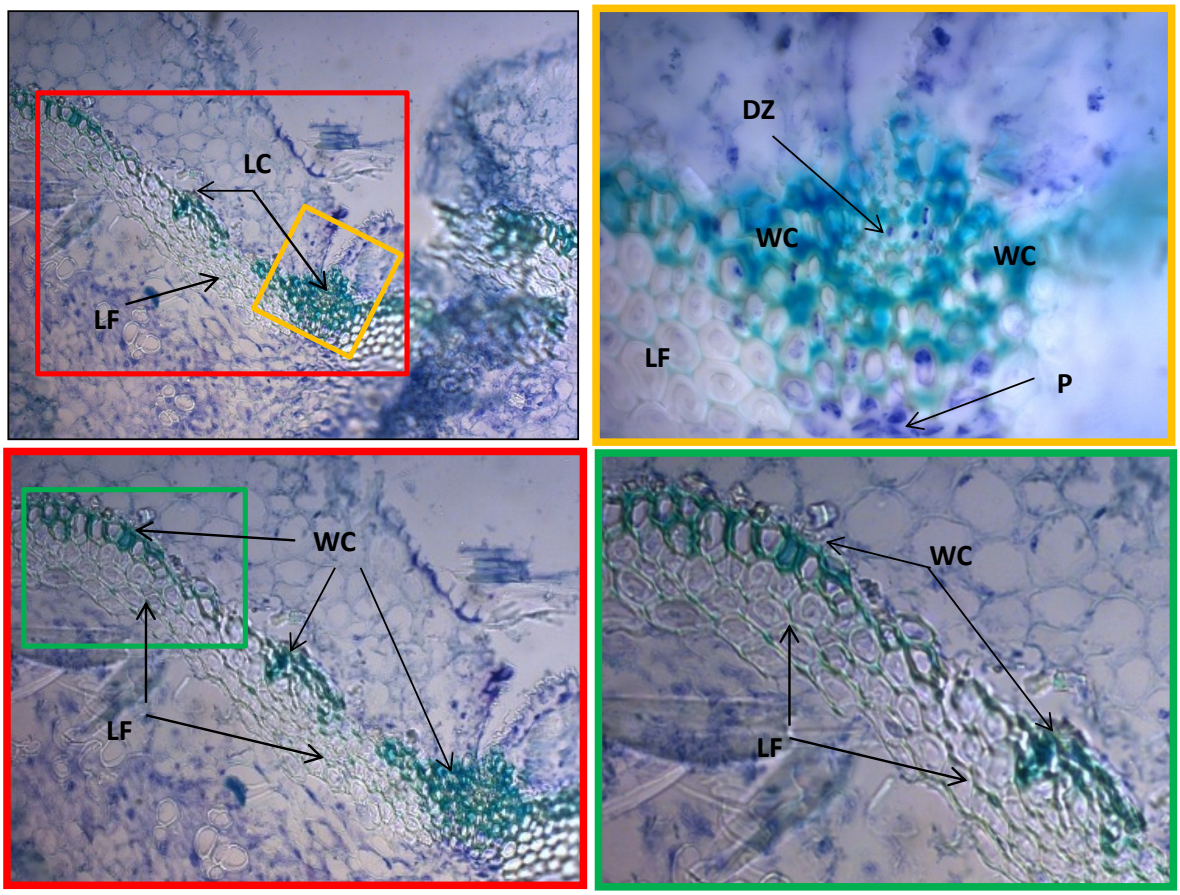


Ventral Sheath - 20days

Paraffin inclusions-Toluidine blu staining

Panel n.8

244A-1a (higly dehiscent)

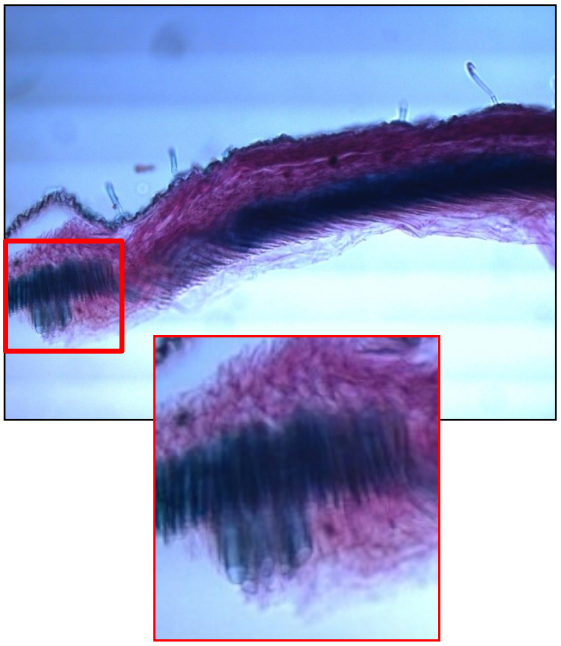


Panel n.9

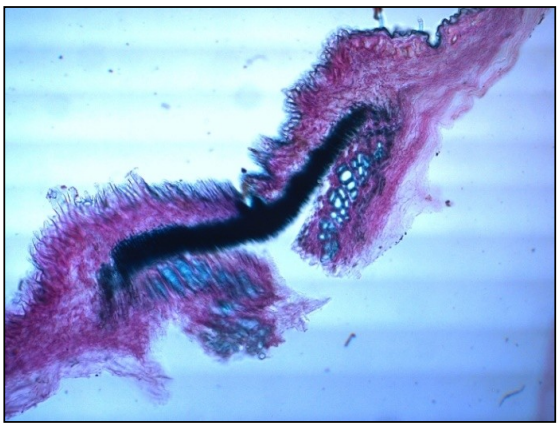
Ventral Sheath – Maturation stage

Cochineal carmine/green iodine

Mg38



Midas

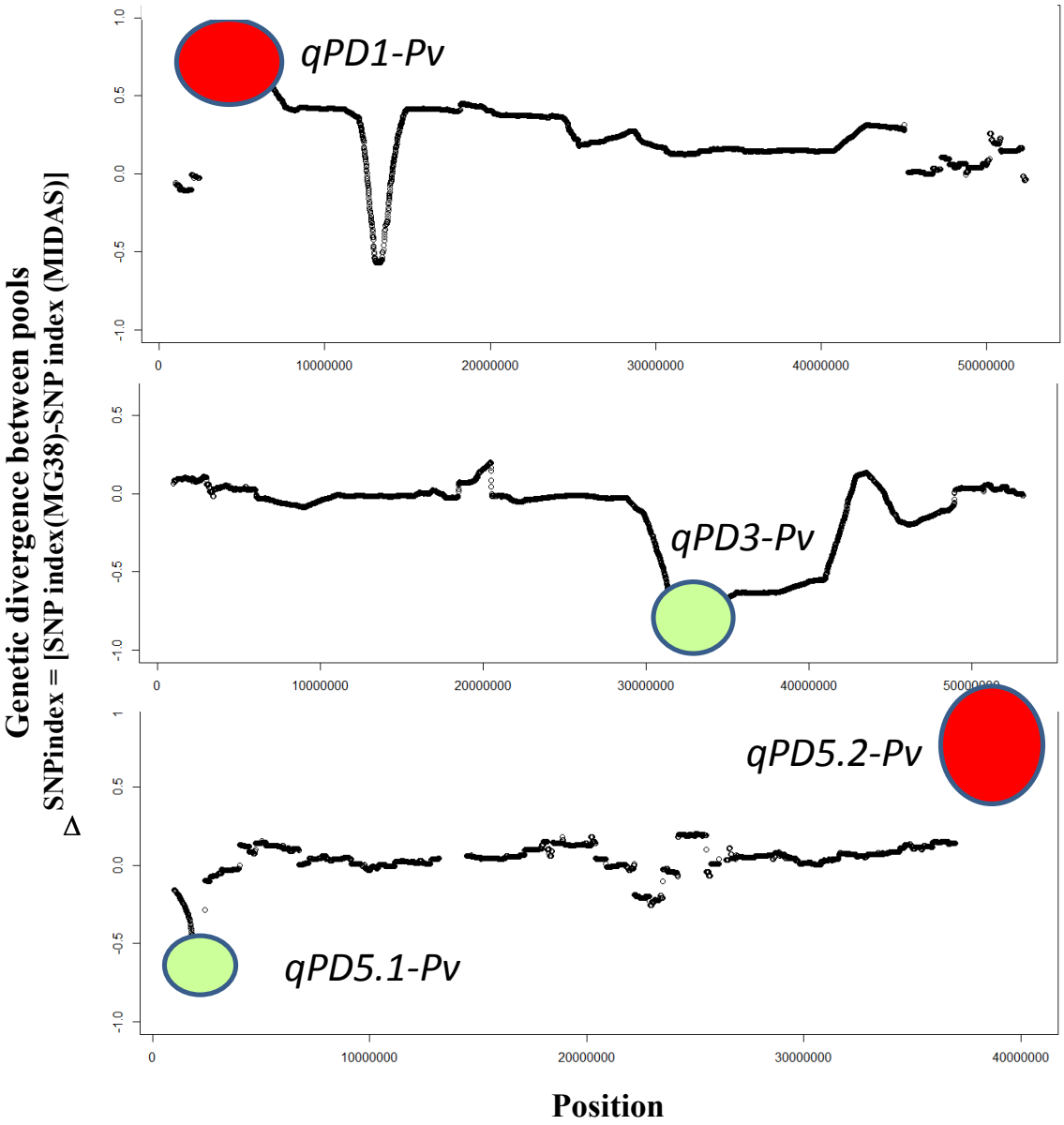


Mapping the shattering locus in *Phaseolus vulgaris*

We first compared two pools of lines with very contrasting level of shattering; the first pool ($Pool_{SH}$) comprised 29 genotypes that were completely not-dehiscent while the second pool ($Pool_{SH}$) comprises 30 individuals with a shattering level equal or even higher than MG38 (≥65%).

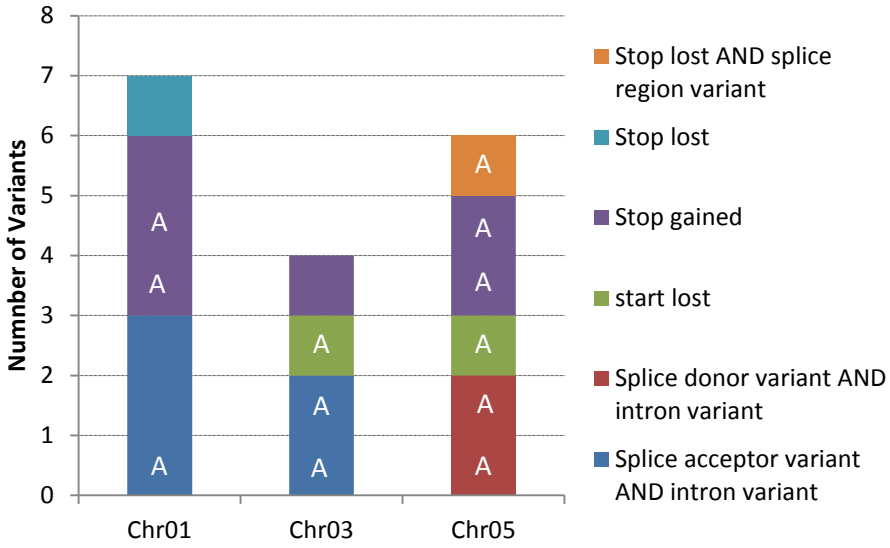
The analysis revealed four genomic regions that are highly divergent between pools, i.e. are associated with the occurrence/not-occurrence of shattering (Figure 2.21. These regions were located one chr.1, one on chr.3 and two on chr.5. These were designated as QTLs affecting the shattering trait: qPD1-Pv (QTL for pod dehiscence on chr.1 of *Phaseolus vugaris*), qPD3-Pv, qPD5.1-Pv, and qPD5.2-Pv, respectively. Two of these genomic regions were characterized by positive values of the SNP index (qPD1-Pv and qPD5.1-Pv), while two by negative values (qPD3-Pv and qPD5.1-Pv). Positive values meant that the pool composed by shattering lines ($Pool_{SH}$) inherited the genomic segment from MG38 (the parent characterized by high shattering) and, similarly, the pool containing non-shattering lines ($Pool_{SH}$) inherited the genomic segment from MIDAS (the parent characterized by completely absence of shattering). Differently, negative values indicated that lines of $Pool_{SH}$ inherited the segment from MIDAS (not-shattering) while the lines of $Pool_{SH}$ inherited the segment from MG38 (highly shattering). Based on the delta SNP index represented in the plot depicted in Figure2.21, the most impacting QTL was qPD5.2-Pv for which the SNP index almost reached the maximum value of 1, fFollowed by qPD1-Pv (SNP index ~ 0.9), qPD3-Pv (SNP index ~ -0.75) and qPD5.1-Pv (SNP index ~ -0.6). On average, the size of the four identified QTLs was of 3282500 bp. The longest segment was that of qPD1-Pv (4400000bp) followed, in decreasing order, by qPD5.2-Pv (4140000 bp), qPD3-Pv (3560000 bp) and qPD5.1-Pv (1030000 bp). The segments contained 326 genes (qPD1-Pv), 191 genes (qPD3-Pv), 103 genes (qPD5.1-Pv) and 491 genes (qPD5.2-Pv), respectively. The number of SNP was 14493 (Chr1), 11787 (chr3) and 9785 (Chr5), respectively.

Figure 2.21 – Genetic divergence (Δ SNPs index) along chromosomes 1, 3 and 5 between shattering and not-shattering pools of lines. Δ SNPs index = SNP index of the pool of the highly shattering lines *minus* the SNP index of the pool of not-shattering lines. The plot represents the average values of sliding windows of 2Mb with a step of 10Kb. The SNP index is calculated as fraction of reads per position that are attributable to the MG38 (highly shattering parental line. Lines of $Pool_{SH}$ inherited the segment from MG38f (red) or from MIDAS (green).



Among all of these 36065 SNPs, 33213 were classified as "modifier", 1657 of "low" impact, 1178 of "moderate impact", and 17 of "high" impact. Variants with high impact were distributed seven on chr.1, four on chr.2 and six on chr.5. All these 17 variants were in transcript and in coding region and 12 were attributed to annotated genes. They can be classified in six variant types among which the prevalent was "stop gained" that was also observed across the three chromosomes.

Figure 2.22 – Number, distribution across QTLs, and types of the variants with high impact on the shattering/highly shattering phenotypes based on pool-seq analysis. A = annotated.

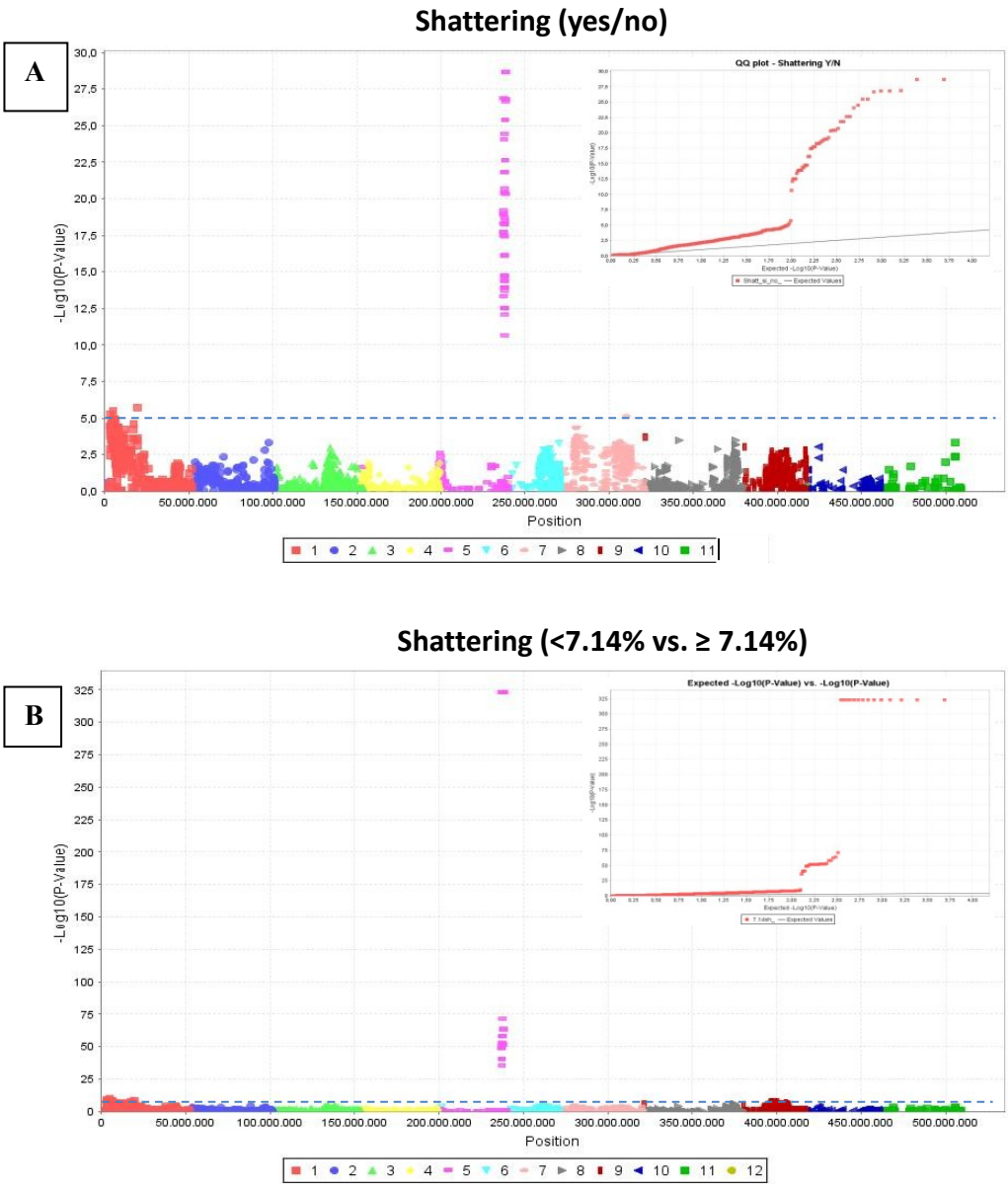


Mapping by using GBS

A first run of mapping was conducted considering shattering as a qualitative trait with two possible states: shattering and non-shattering without correcting for kinship. A QTL with a major effect popped up on chromosome 5 (Figure 2.23A). This was characterized by very high probability level ($P<10^{-28}$) and a notable R^2 of 0.68. We repeated the analysis by contrasting the lines with a level of pod shattering < 7.14% against those having ≥ 7.14% of shattered pods (Figure 2.23B). The evidences in favor of this QTL on chr.5 were strongly reinforced with a sharp increase of its statistical significance (from 10^{-28} to 10^{-325}) and with a maximum $R^2 = 0.99$ (data not shown). The association

remained significant at permutation test ($P<10^{-4}$) and after Bonferroni correction ($P < 0.05/4913$). Ather significant ($P <10^{-5}$) peaks were also identified on chr.1 with an R^2 value for the best associated markers of 0.42 ($P<10^{-4}$ at permutation test).

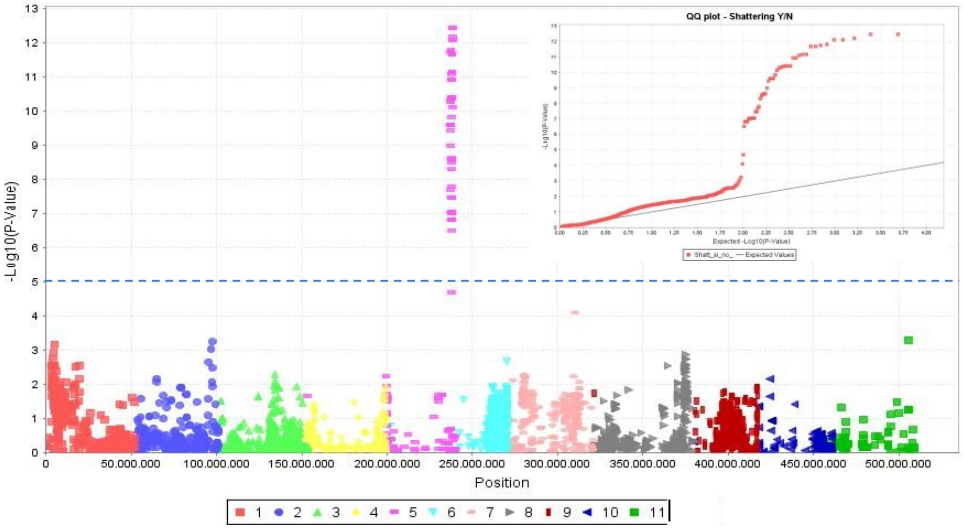
Figure2.23 A-B – Manhattan plots for the pod shattering trait using 143 bean lines. The negative log_{10}-transformed p-values of each test are plotted against the marker position in the genome for the shattering as qualitative trait (above) and the shattering level <> 7.14 (below). Dotted line: $P = 10^{-5}$. Embedded: quantile-quantile plots for general linear model (GLM) (without correction for kinship).



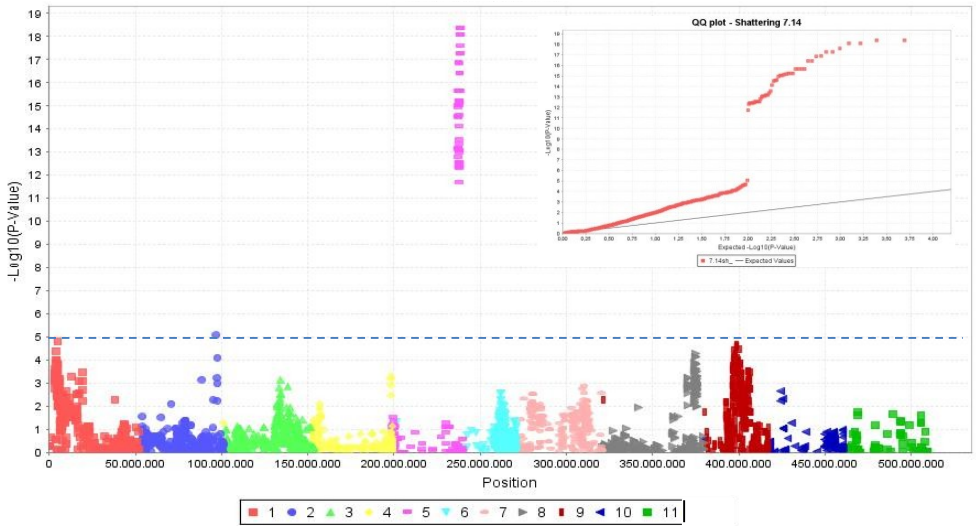
When the analysis took into account kinship, the statistical significance of the observed peaks was reduced albeit it remained significant ($P<3.65x10^{-13}$ and $P<3.65x10^{-19,}$ respectively; (Figure 2.24) also after Bonferroni correction ($P<10^{-5}$), with the overall picture substantially unchanged. Indeed, the analysis identified a genomic region spanning 1044648 bp containing 41 markers with an average R^2 of 0.95. In particular, there were nine markers with R^2 between 0.80 and 0.90 which belong to a single genomic stretch of 72221 bp; 17 markers with R^2 between 0.91 and 0.95; and 23 markers with an $R^2 > 0.99$. The 23 with $R^2>0.99$ were organized in four distinct segments that summed up to 199425 bp. Moreover, another significant ($P<10^{-5}$) peak was identified in the terminal part of the chr.2 with an $R^2 = 0.23$ ($P < 8.48x10^{-6}$). A peak marginally not significant ($P=1.51x10-^{5}$) was also observed at the proximal end of chr.1.

Figure 2.24 – Manhattan plots for the pod shattering trait using 143 bean lines. The negative log_{10}-transformed p-values of each test are plotted against the marker position in the genome for the shattering as qualitative trait (above) and the shattering level <> 7.14 (below).. Dotted line: $P= 10^{-5}$. Embedded: quantile-quantile plot for MLM (including correction for kinship).

Shattering (yes/no)



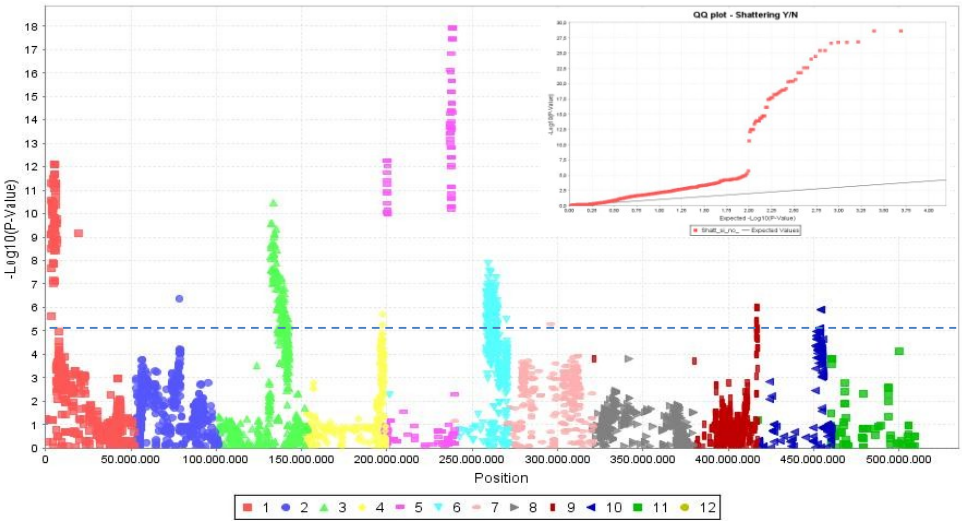
Shattering (<7.14% vs. ≥ 7.14%)



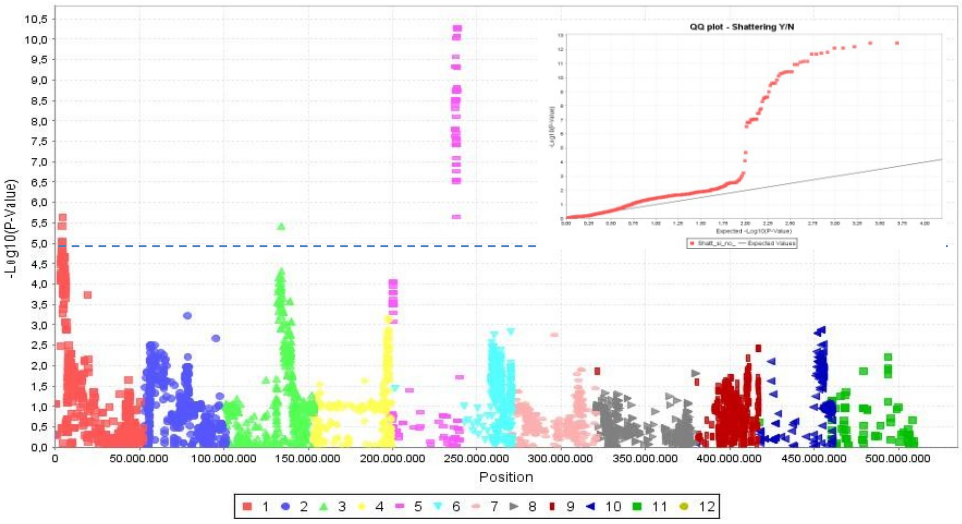
In a further step we mapped the trait "frequency of shattering pods" (per plant) using all 143 lines. The GLM association mapping analysis evidenced eleven peaks of which four (two on chr.5, one on chr.1 and one on chr.3) had higher significance levels compared to the others. These four peaks roughly corresponded to the four QTLs previously identified by pool-seq analysis. The other QTLs, in decreasing order of relevance, were on chrs.1, 6, 2, 9, 10, 4 and 7. Moreover, when the MLM was applied, none of these seven QTLs reached the significance after Bonferroni correction. In this case only the peaks corresponding to qPD1-Pv, qPD3-Pv and qPD5.2-Pv of the pool-seq analysis were significant ($P<10^{-5}$).

Figure 2.25 – Manhattan plots for the trait "frequency of shattered pods" when using 143 bean lines and for both GLM (above) and MLM (below) analyses. The negative $\log_{10}$-transformed p-values of each test are plotted against the marker position in the genome. Dotted line: $P= 10^{-5}$. Embedded: quantile-quantile plots.

Frequency of shattering pods (GLM)



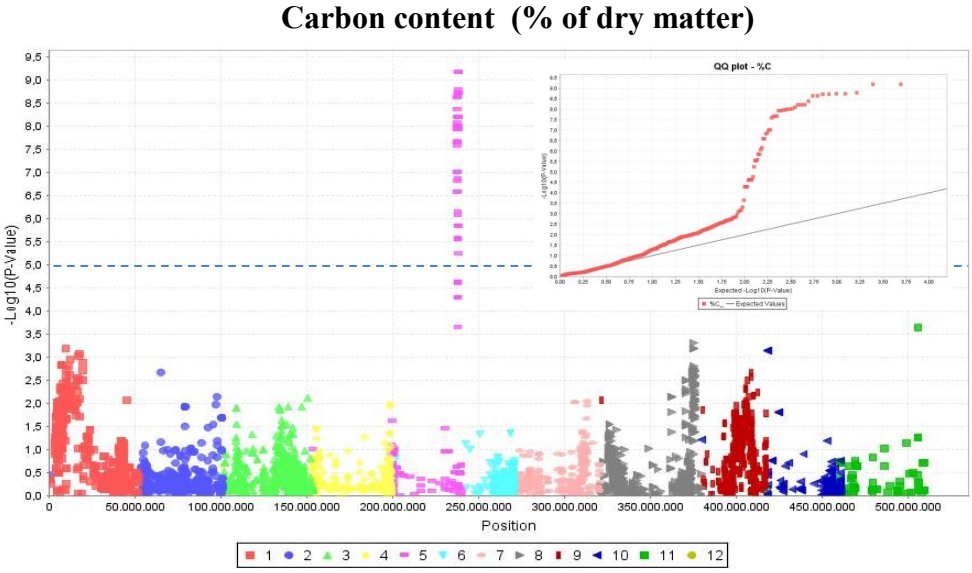
Frequency of shattering pods (MLM)



Taken into account the results of the MLM analysis, and adopting a more conservative approach to reduce the risk of false positives, from here after only the results of association mapping conducted correcting for kinship (MLM approach) are presented.

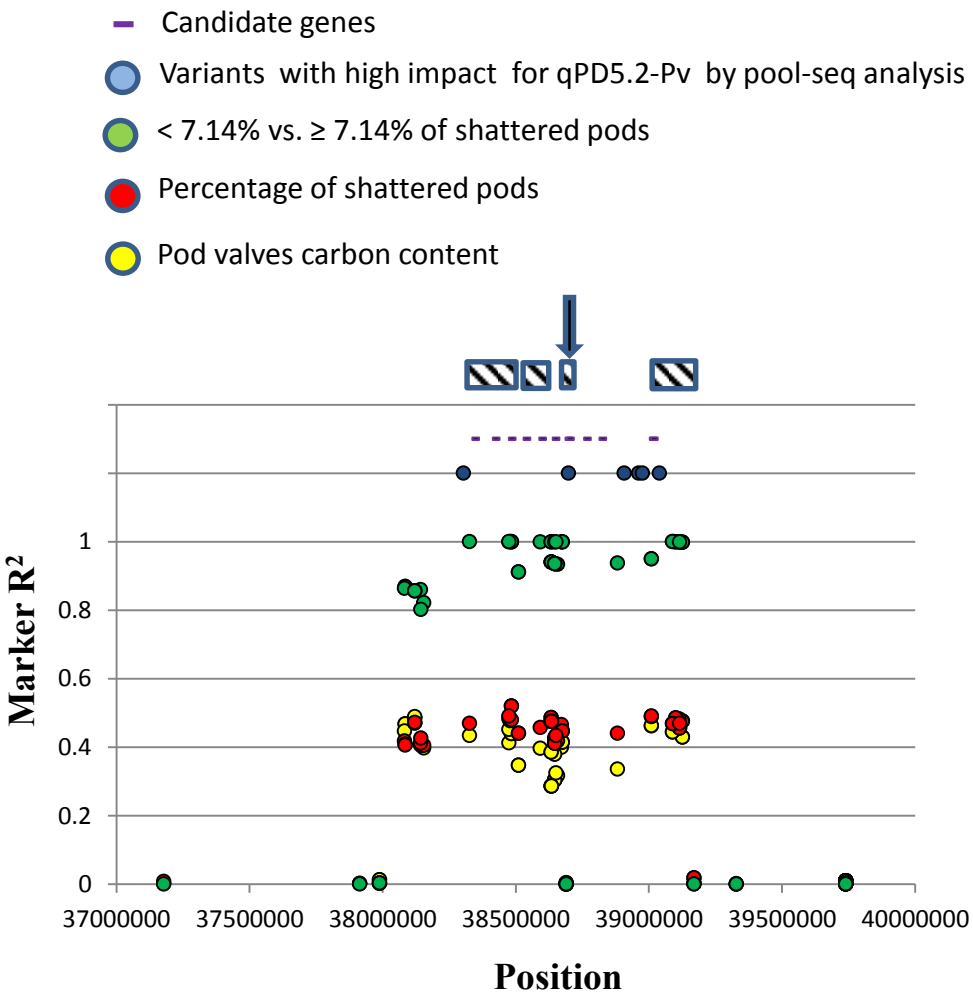
When the trait “carbon content” of the valves was considered, a single major locus explaining the 45% of the total variance for the trait, was mapped on chr.5 (Figure 2.26).

Figure 2.26 – Manhattan plot for the trait carbon content (% of dry weight) using 143 bean lines. The negative log_{10}-transformed p-values of each test are plotted against the marker position in the genome. Dotted line: P= 10^{-5}. Embedded: quantile-quantile plot for MLM (including correction for kinship).



This co-mapped with the peak found for presence/absence of shattering and for the level of shattering (Figure 2.27). It was also evident a fitting overlap between the results of pool-seq analysis and GBS analyses for this locus (figura 2.27). Indeed, all the six high impact variants found at the qPDH5.1-Pv locus (of ~4 Mb) also belong to the genomic stretch (of ~ 1 Mb) where shattering and not-shattering individuals are differentiated by GBS analysis.

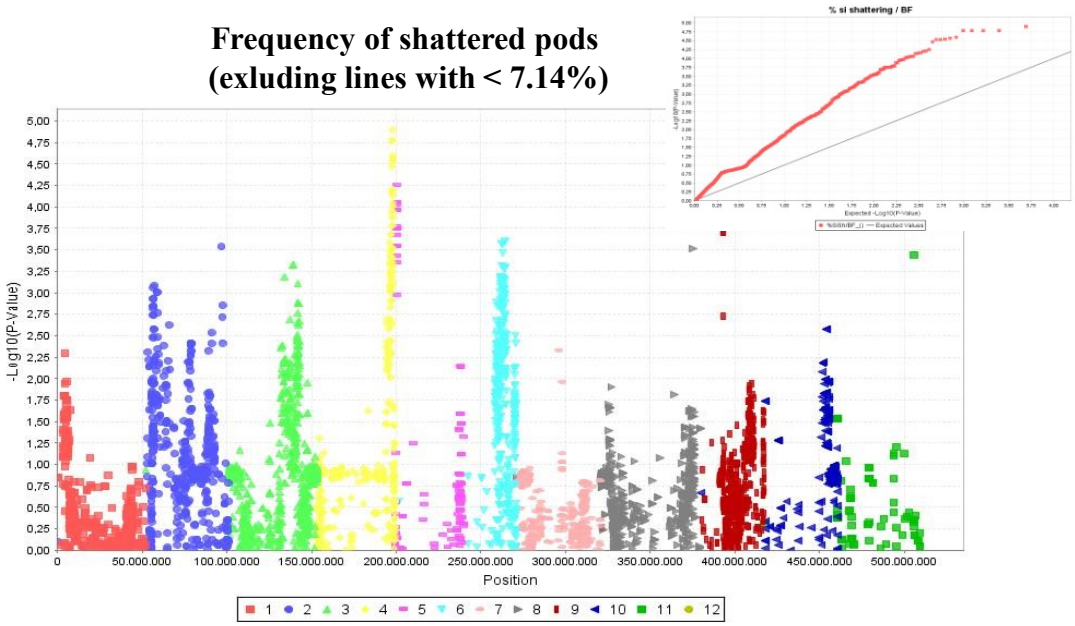
Figure 2.27 – Relationships between the positions associated to pod shattering variables, carbon content and candidate genes (see legend). Dashed boxes and the arrow indicate the genomic regions deserving further analyses for the cloning of the shattering gene in *Phaseolus vulgaris*.



We observed that this region on chr.5, includes 130 genes; among these, 106 are annotated and, as also emphasized in the Figure 2.27, eleven can be considered as promising candidates involved in the shattering processes and/or in the cell-wall biochemistry (see table2.1 an 2.2 of the materials and methods section). However, this genomic segment did not contain orthologs nor homologs of the shattering genes found in soybean. Moreover, the detected genomic region did not contain any genes with a NAC domain (which characterize the soybean *SHAT1-5*) nor genes for proteins belonging to the dirigent-like family (which is the protein family encoded by the soybean *PDH1*).

As a further step, we focused on the lines displaying a variation from 7.14% to 82% for the level of shattering. Thus, mapping was performed excluding lines with a shattering level < 7.14%. Any significant associations was found ($P>10^{-5}$) (Figure 2.28)

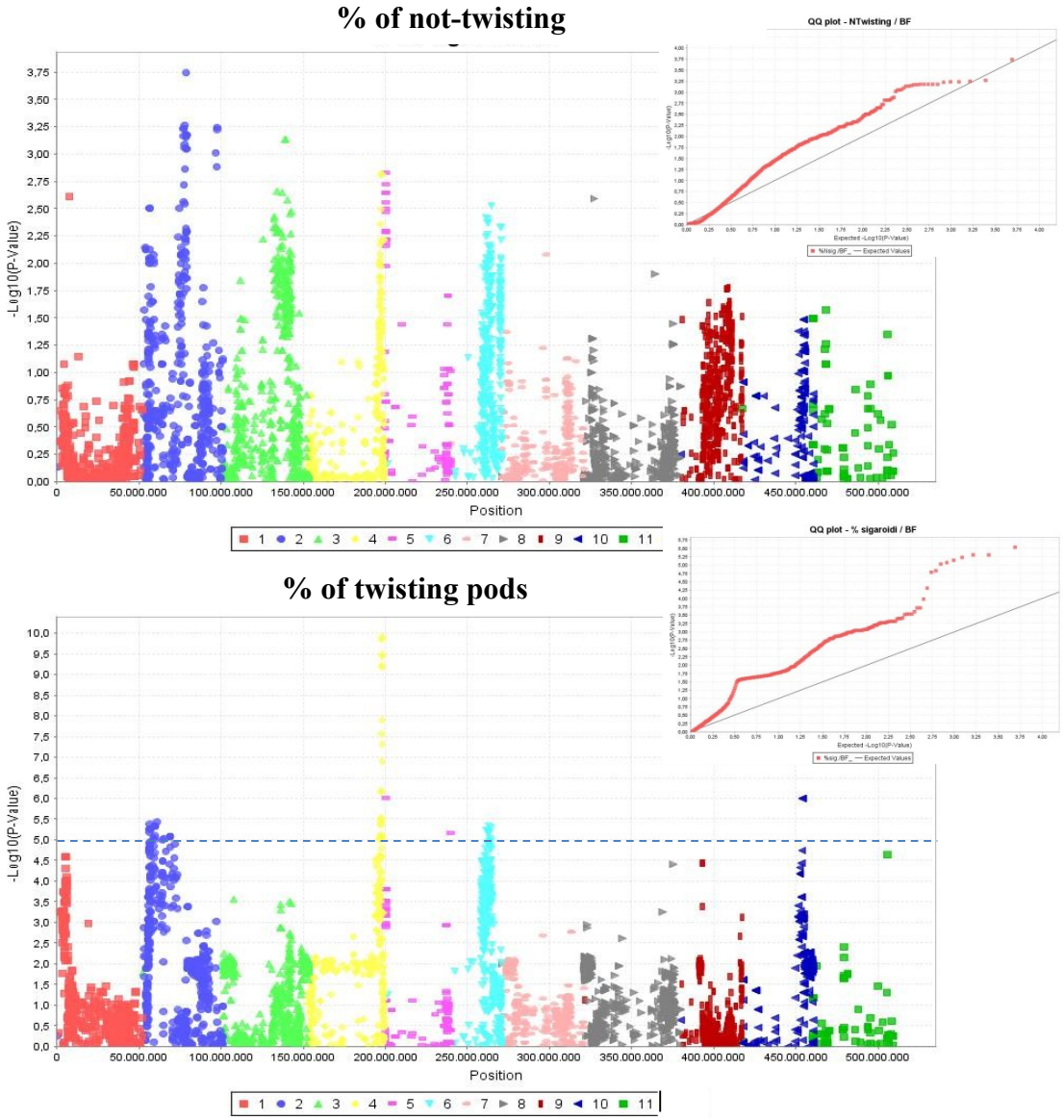
Figure 2.28 – Manhattan plot for the trait frequency of shattered pods per plant using 115 bean lines. The negative log_{10}-transformed p-values of each test are plotted against the marker position in the genome. Embedded: quantile-quantile plot for MLM (including correction for kinship).



Moreover, in regard to the modality of shattering, none of the associations observed for the frequency of not-twisting pods reached the significance threshold ($P>10^{-5}$), albeit a locus "stand up" in pericentromeric position on Chr.2. Differently, for the percentage of shattering pods six regions across four different chromosomes were found (Figure 2.29). The most impacting locus for the frequency of twisting pod was mapped on chr.4 (R^2 = 0.28). Other QTLs ($P<10^{-5}$) were identified on chr.10 (R^2 = 0.28), two on chr.5 (R^2 = 0.20 and R^2 = 0.19, respectively), one on chr.6 (R^2 = 0.16) and one on chr.2 (R^2 = 0.20).

Figure 2.29– Manhattan plots for the percentage (%) of not-twisting and twisting pods. using 115 bean lines (143-28 with <7.14% of shattering). The negative log_{10}-transformed p-values of each test are plotted against the marker position in the genome.

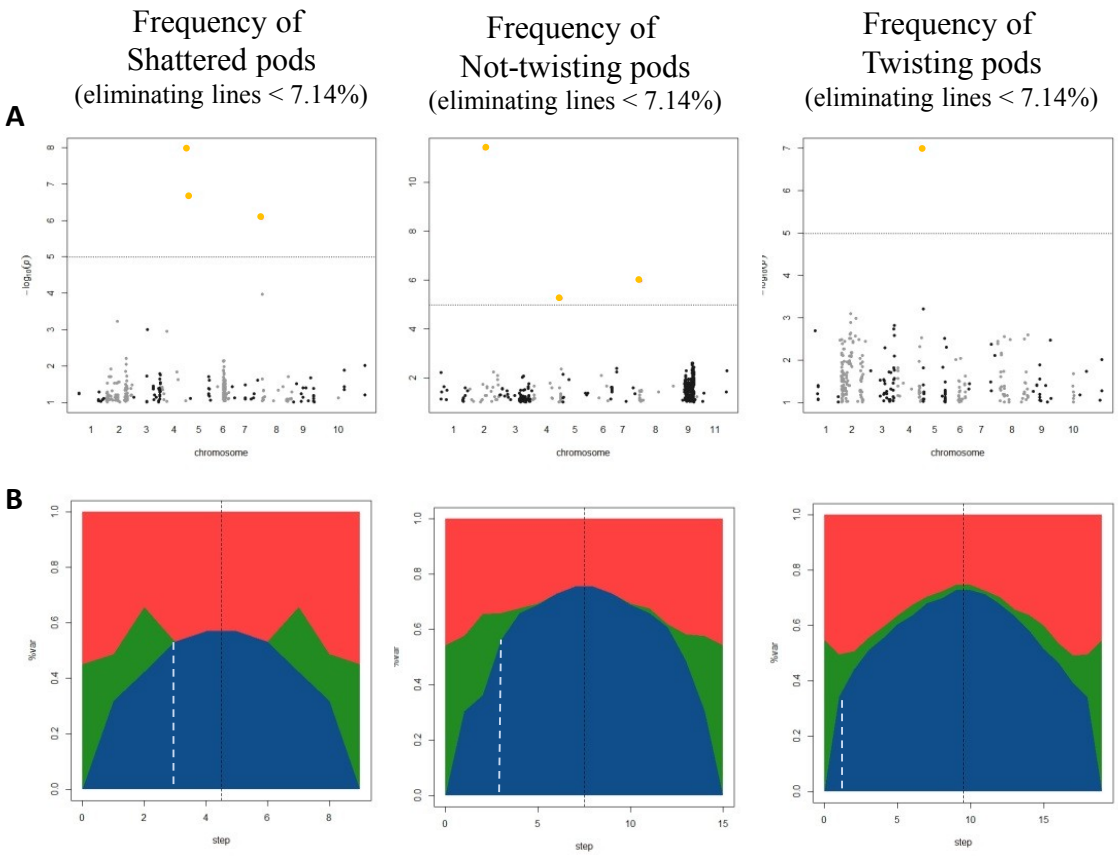
Dotted line: P= 10^{-5}. Embedded: quantile-quantile plot for MLM (including correction for kinship).



The genetic bases of the tuning of shattering were also studied using a multilocus mixed-model approach (MLMM). This method indicated that the level of shattering depended by three loci that were on chr.4 (P=1.0x10^{-08}), chr.5 (P=2.2x10^{-7}) and chr.8 (P=8.5x10^{-7}) that cumulatively explained the 55% of the total variance. The frequency of not-twisting pods was also explained by three loci that were on chr.2 P=4.0x10^{-12}), chr. 8 (P=1.1x10^{-06}) and chr. 4 (P=5.71x10^{-6}) that cumulatively explained 56% of the total variance for the trait. The frequency of sigaroids is best explained by a

single locus on chr.4 ($P=9.9x10^{-08}$) that accounted for 35% of the phenotypic variance. The locus on chr.4 was the same for the three traits.

Figure 2.30 – (**A**) The posterior probability of association scan after the Bayesian MLMM has included from one to three loci in the model (orange circles). (**B**) Partition of phenotypic variance for each forward inclusion (nine, 15 and 20 steps, respectively) and backward elimination (nine, 15 and 20 steps, respectively after the dashed black line). The vertical dashed white line marks the number of loci entered in the model (three, three and one, respectively) with the Bonferroni correction.



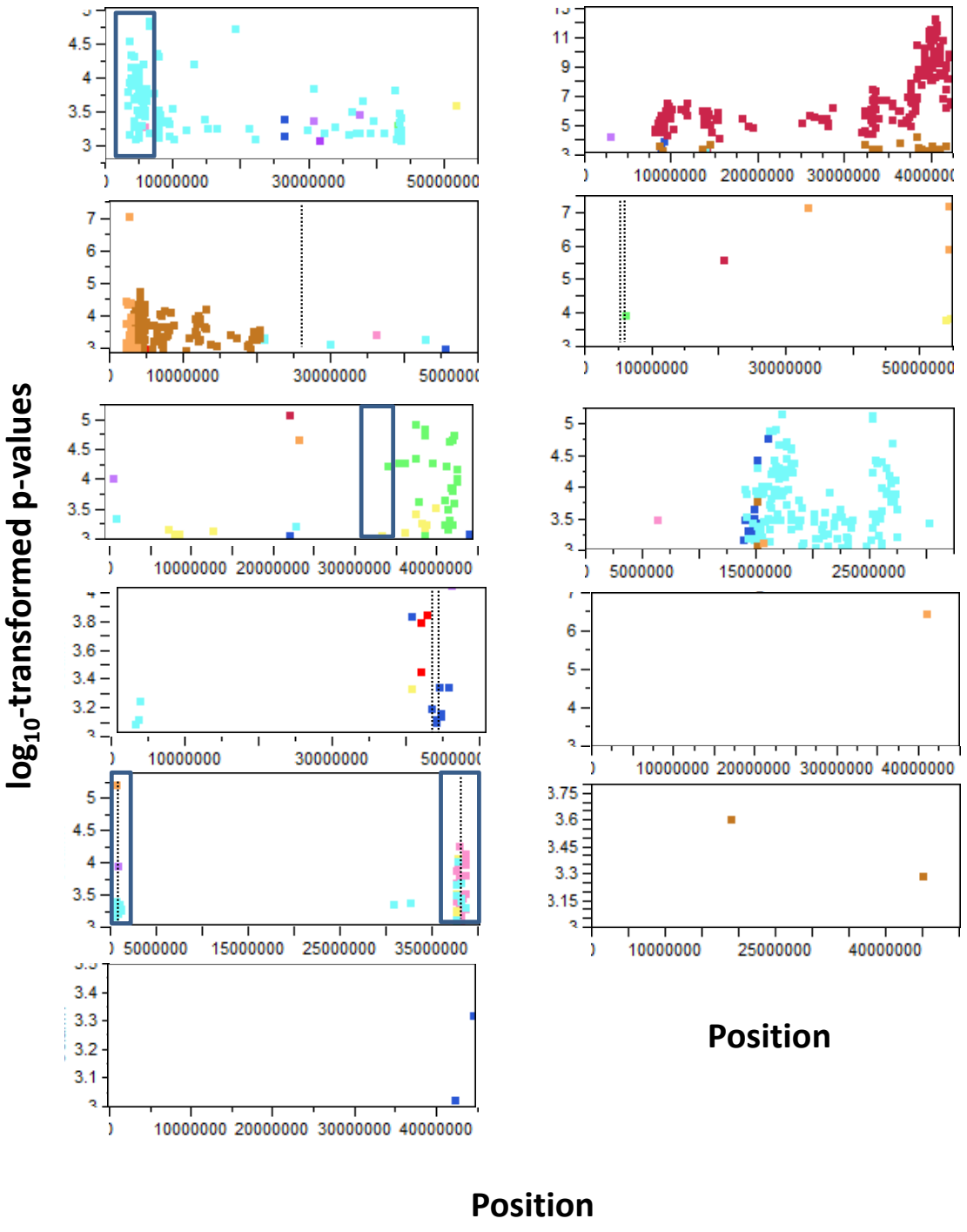
Finally, for comparative purposes with the pod shattering QTLs, we have mapped six pod traits together with plant vigor, flowering time and growth type traits. Results are presented in Table 2.20 and Figure 2.31. Overall, it appeared that pod characteristics likely depends on several genes distributed all across the bean chromosomes, with a lower relevance for chr.10 and chr.11 (Table 2.20).

Table 2.20 – Distribution of the significant associations ($P>10^{-3}$) obtained by applying the MLM model on six pod traits and three others plant life history traits (vigor, habitus and flowering time). Color key is the same used for figure 2.31.

	Chromosomes											
Trait	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
Pod color	0	0	1	0	0	0	242	1	0	0	0	244
Average pod weight	0	128	0	0	0	0	20	0	2	0	2	154
N. of seeds per pod	0	1	0	3	0	0	0	0	0	0	0	4
100-seeds weight	121	5	4	4	53	0	2	0	175	0	0	366
Harvest index (pod level)	3	1	0	0	42	0	0	0	1	0	0	47
Pod area	3	1	2	14	0	3	6	0	16	0	0	45
Maximum/curved length	1	0	16	1	8	0	0	2	0	0	0	28
Plant vigor	0	28	1	0	1	0	0	3	1	1	0	35
Habitus	1	0	35	0	0	0	0	1	0	0	0	37
Flowering	3	0	1	1	1	0	1	0	0	0	0	7
Total	132	164	60	23	105	3	271	7	195	1	2	967

In several cases, it has been observed an overlapping of the QTLs for pod shattering with regions associated to other traits (Figure 2.31). The qPD5.2-Pv overlapped with loci associated to 100-seeds weight, harvest index at pod level and the maximum/curved length. However the significance levels observed for these traits ($P<10^{-5}$) in this region were much below those observed for pod shattering ($P=10^{-19}$; Figure 2.31). The qPD1-Pv and qPD5.2-Pv also overlapped with loci for 100-seeds weight other than with flowering time (qPD1-Pv) and plant vigor and harvest index at pod level (qPD5.1-Pv). The level and the modality of shattering on chr.4 were associated with pod area and n. of seeds per pod. Minor overlaps were observed also in other cases (such as with the plant habitus) except for pod color.

Figure 2.31 – map of the loci controlling six pod related traits and plant vigor, habitus and flowering time. Empty boxes represent the genomic region detected by pool-seq analysis. Dashed lines represent the loci for shattering: occurrence (yes/no), level (percentage of shattered pods), and modality (sigaroids not sigaroids) as identified by MMLM approach. For color key refer to Table 2.20.



DISCUSSION

The first objective of the present study was to conduct the phenotypic and molecular characterization of a collection of introgression lines of common bean (*Phaseolus vulgaris* L.) that was developed *ad hoc* to study the domestication syndrome in particular for pod and pod-related traits. This objective was fulfilled. The collection of introgression lines was analyzed for 27 phenotypic traits including quantitative, qualitative and chemical traits; most of them, particularly those related to pod characteristics, showed good level of variation. Moreover, a Genotype by Sequencing approach (Elshire et al.,2011; Poland et al.,2012; Sonah et al.,2013). was pursued to genetically characterize 143 of the phenotypically studied lines. This allowed building a phenotypic and molecular dataset that in the next future will serve marker-traits association studies that may ultimately lead to the identification and the cloning of genes underlying very important agronomic traits.

To investigate the potentiality of the analyzed plant materials and the built dataset, we specifically focused on the study of pod shattering, one of the most relevant domestication traits in common bean, with the aim to decipher its genetic bases. With this purpose we completed the GBS analysis with a bulked segregant analysis using a pool sequencing approach (Kofler et al.,2011; Rellstab et al., 2013).

Deciphering the genetic bases of pod shattering is important for both evolutionary studies (Lin et al., 2012; Dong and Wang, 2015) and breeding (Sing, 2001; Santalla et al.,2004) purposes. We have demonstrated that it is indeed possible to dissect the genetic bases of this trait determining the number and the genomic locations of the QTLs, either conditioning the indehiscent/dehiscent phenotype as also the tuning of the level and the modality of the trait. Moreover, we drafted the genomic map of the domestication syndrome at pod level. This allowed to discuss the identified QTLs in relation to QTLs for other plant traits and therefore investigate possible pleiotropic and linkage drag effects.

Phenotyping of pod shattering

Overall, the analyzed collection displayed a good level of variation for both the intensity and the mechanism of pod shattering. Within our set of bean lines we distinguished all the different shattering modes described in previous studies (Lamprecht et al., 1932; Prakken, 1934), varying from completely indehiscent to "twisting". This latter is characterized by very strong distortion of pod valves, passing through two supposed (Lamprecht, 1932) "intermediate" states as "fissured" and shattered but not-twisting pods. We characterized each line by counting and classifying its pods into the above cited four categories as also, independently measuring the degree of resistance to manual shattering.

A first outcome, as showed by partition analysis, indicated that to model the resistance to manual shattering it is better to consider the frequency of shattered pods, and only secondarily the modality of shattering (twisting/ not-twisting). Moreover, a low threshold (10%) of shattered pods is sufficient to distinguish low and medium-high resistance. All this indicates that the mechanism of shattering might be first determined by a major factor controlling the separation/non separation of the valves along the sutures, while the modality of shattering (twisting/not twisting) can be further modulated. This result also suggests that for association studies other than to consider all the lines simultaneously it could be more informative 1) to contrast indehiscent vs. dehiscent lines (regardless of the degree of shattering); 2) to deepen the analysis into the dehiscent lines only (regardless of the modality); 3) among dehiscent, to map separately the traits twisting and non-twisting. These steps would help dissecting the shattering trait into its components. As a corollary, it should also be noted that the variable "fissured" was not considered useful to predict manual resistance to shattering; this suggests that fissured pods should be investigated separately from dehiscent and indehiscent pods in association studies.

Shattering level was very poorly correlated with the other morpho-phenological traits and productive characteristics of the plants. This suggested that a (relatively) limited "overlap" was to be observed between QTLs for pod shattering and other QTLs for pod characteristics (i.e. relatively limited linkage drag and / or pleiotropic confounding effects). However, from the correlation study another interesting consideration stemmed from the observation that shattering (albeit poorly) is significantly associated with low 100-seeds weight, low harvest index at pod level and

small pod size. Indeed, this suggests that pod shattering might have an "energy cost" (Mcginley et al., 1988, Chiariello et al.,2000), i.e. the synthesis of biomolecules needed for shattering might reduce the resources available for seed and pod development. Alternatively, the same results might be explained by pleiotropic effects (genes for shattering might exert minor effects on other pod traits); it would be also possible that shattering gene(s) are physically closely linked to genes involved in the partition of photosynthetsis at pod level (linkage drag).

Concordant with the hypothesis of an "energy cost" for dehiscence, we observed that carbon content of the pod valves is strongly and positively correlated with the level of shattering. Interestingly, an abrupt increase in carbon content was observed for a level of shattering of about 7.14%; this value is similar to the threshold of 10% of shattered pods needed to explain the largest portion of the variance for resistance to manual shattering. This suggests that in association studies it would be useful to compare lines with less than 7.14% of shattering against those with more than 7.14% of shattering. This is also corroborated by the fact that carbon content is strongly correlated to lignin content (see below).

The content in C, H and N is expected to be stoichiometrically correlated to the amount of organic matter of the tissues (Chiariello et al., 2000) and thus to the cumulative content of, carbohydrates, protein, lipids and all others organic compounds. However, in plant differences in carbon content were frequently correlated to differences in lignin content (Loader et al.,2003). Coherently, when we compared two groups of genotypes strongly contrasting for shattering level and carbon content we observed a clear-cut difference in the content of cell wall fibers, higher in shattering types than in not-shattering type. More in particular, the passage from indehiscent to dehiscent types was characterized by an abrupt change in lignin content followed by those in hemicellulose and cellulose content. Moreover, carbon content was best correlated with lignin content than with hemicellulose and cellulose content. Interestingly, we also observed that lines characterized by a shattering level higher than MG38, consistently have lower carbon content, lower hemicellulose and cellulose but higher lignin content. This indicates that not only the overall content of lignin but also its relative incidence compared to other cell wall components might play a role in modulating the level of shattering in bean. All this pointed towards the hypothesis that some histological and/or anatomical differences attributable to differential patterns of

fiber deposition and/or lignification might underlay the shattering/ non shattering phenotypes.

The results of the anatomical and histological investigation fitted these expectations. Indeed, clear-cut differences between MG38 and MIDAS were found either in the ventral and dorsal sheaths as also in the layer between the inner and the outer parenchyma of the pod wall.

Specifically, the ventral sheath of MG38 was characterized by a very high ratio thickened fiber cells/wood cells, while the opposite was observed for MIDAS where almost all cells were wood cell types. This observation is consistent with that made by Prakken (1934) (see fig 1 of the introduction of this chapter) comparing "stringy" and "stringless" varieties. Indeed, he considered this anatomical difference at the basis of the presence/absence of pod strings. Moreover, it has been suggested that the indehiscent fruit result from the loss of fibers in the sutures ("stringless"), which is under the control of a major QTL, *St* locus (Koinange et al., 1996). Contrary to the observations obtained in common bean in the present, in soybean no differences between shattering and non-shattering types were observed for cell types or degree of lignification in the lateral part of the ventral sheath (Dong et al. 2014). Indeed, in soybean the difference between wild shattering and cultivated non-shattering types was limited to the dehiscent zone (Dong et al. 2014); specifically in soybean it has been observed an excessive lignification of the fiber cap cells in non-shattering genotypes compared to shattering genotypes. Furthermore, a major gene, *SHA1-5*, has been identified as responsible of this phenotype (Dong et al., 2014). Our survey did not evidence clear differences in the dehiscence zone where only wood cells with relatively low wall thickness are present both in MG38 and MIDAS . Thus, albeit we cannot exclude that in bean differences in the dehiscence zone are present that we did not detect in our anatomical-histological survey, all these data suggest that bean and soybean are at least partially different in the mechanism of pod dehiscence.

It has been observed that MG38 has a strong fibrous and lignified layer between the inner and the outer parenchyma of the pod wall, while these was completely absent in MIDAS. This difference was also noted by Prakken (1934) comparing "stringy" and "stringless" types. Interestingly, Funatsuki et al. (2014) also noted that a differential lignification in the lignin-rich inner sclerenchyma of pod walls is at the basis of the pod shattering in cultivated germplasm of soybean. Furthermore, they show that the twisting

ability of the pod is under the control of a major gene, (*PDH1*) involved in lignin deposition in this layer. We point out here that in Funatsuki et al. (2014) the difference between shattering and non-shattering types is in the *degree* of lignification of the layer while in the present study in bean is in the presence/absence of the lignified layer. This strongly suggests that in common bean the role of this lignification might be more relevant or have a different role when compared to soybean.

However, not all anatomical or histological differences between the wild-like parent (MG38) and the cultivated varieties (MIDAS) are necessarily correlated with shattering traits. However, encouragingly, the difference found between MG38 (dehiscent) and MIDAS (not dehiscent) were also confirmed when considering two lines with very contrasting level of shattering, indicating that the differences found in the ventral sheath and in the lignin-rich inner sclerenchyma of pod walls might be indeed responsible of the shattering/non-shattering phenotype.

Interestingly, Prakken (1934) suggested that in bean the control of the traits "stringlesness" (that depends on the characteristics of the ventral sheaths) and "parchment" (that depends on the layer between the inner and the outer parenchyma of the pod wall) was independent and in both cases under a simple monogenic control. Other authors suggested the existence of an oligogenic control for the stringless trait with the contribution of either environmental effects or epistatic interactions (Dong and Wang,. 2015). Thus, it is apparent that in bean, artificial selection might have targeted multiple cellular mechanisms and the controlling genes to minimize seed loss during soybean domestication.

This somewhat evokes the scenario resulting from jointly considering the results obtained by Dong et al. (2014) and Funatsuki et al (2014).

A further and relevant conclusion of the phenotyping characterization is that given the observed parallelism with other systems (in particular with soybean) it might be useful to map on the bean genome all of the gene homologs known to be involved in the shattering process (see table 2.1 of Dong and Wang, 2015). Indeed, it might be possible that polymorphisms in other homologs of the shattering genes are also associated with pod indehiscence in the common bean genome. Additionally, given the results of our chemical analyses and the anatomical-histological survey as also considering that the fibers are mainly composed of sclerenchyma cells with well-developed secondary cell walls, it is likely that genes involved in the regulation of

secondary cell wall deposition or fiber cell differentiation may underlay *St* locus (Koinange et al. 1996) or to other QTLs controlling pod dehiscence detected from the present association mapping study.

Gene analisys

As a “proof of concept”, the mapping study of pod shattering trait conducted in this work showed that this population can be a valid tool to assist gene mapping. Indeed, we showed that it was possible to map with an adequate resolution a major QTL underlying the pod shattering in *Phaseolus vulgaris*, in a region of ~200.000 bp on chr.5. We also dissect the trait into its different components to decipher the genetic bases of the “tuning” of the trait (as proportion of shattered pods per plant and proportion of the different mode of shattering). Complementarily, we also drafted in common bean the map of the genomic region underlying the genetic control of the domestication syndrome at pod level and we showed that several regions scattered across the genome were involved with any evident genomic “hot spot” of domestication.

The main result of the present study is that consistently for pool-seq and GBS analyses, a major QTL for pod dehiscence on the distal part /sub-telomeric region; mary guarda su phytozom) of the long arm of chromosome 5. This is in contrast to what has been previously suggested by Koinange et al. (1996) which mapped “pod string” on chromosome 2. Two hypotheses can be made to explain the lack of congruence between the present study and that of Koinange et al. (1996). First, it is conceivable that *St* locus was erroneously mapped on chromosome 2 because the phenotyping of the pod string trait might be not reliable. Indeed, while Koinage et al. (1996) recognized that the loss of dispersal ability is conditioned by the presence of fibers in the pods, both in their sutures ("string") and their walls, they simply determined “the presence of fibers in pod sutures and pod walls […] by breaking the pod beak or pod wall, respectively, and examining the break surface for the presence of fibers”.

A second hypothesis is that this phenotyping address a component of pod shattering that is not necessarily the “main trait”. Based on our genetic data, this hypothesis seems realistic because when we compared the lines with < 7.14% vs. those with ≥ 7.14% a secondary QTL was detected on chr. 2 in position 44603427 (figure xx or results section). Using the bean genome as a reference, we observed that this marker is ~10^{-6} bp from the ortholog of the *AtIND* gene of *Arabidopsis* mapped in bean by Gioa et al. (2012), that is also quite close to the *St* locus (Gioa et al. 2012). Thus, the *St* locus might represent a gene with a certain (slight) impact on bean shattering and not consistent with the major role of *qPD5.2-Pv* as observed in the present study.

In common bean, Prakken (1934) after crossing stringless and stringy bean varieties concluded that stringness (the absence of string along the valves suture) is under the control of a single dominant locus. This author also hypothesized that the character 'parchment' (the lack the membraneous coating on the inside of the pod (parchment layer) is under the control of a second (recessive) gene independent form the first. Albeit with some contradictory results, a simple genetic control of this trait was also proposed by Tshemark (1901, 1902), Emerson (1904), Wellensiek (1922) and Tjebbes and Kooiman (1922), and in some cases more complex models including epistatic interaction and gene-environmental interactions were hypothesized (Currence 1930). The data presented in the present work, however, also indicate that while the determination of the shattering/not-shattering phenotype is conditioned by a single QTL, the control of its mechanism could instead be more complex with several QTLs involved and with some of them specifically underlying twisting and not twisting types. This seems to evoke Lamprect (1932) who hypothesized that a major factor influences the shattering trait while three other genes additively act to tune the expression of the trait. More specifically, the multilocus model used in the present study indicates that the same locus on chr.4 is associated to both the level and the mode of shattering. Two other loci act specifically on the level of shattering and two on the frequency of not-twisting pods. Thus, the model arising from this study can be hereafter summarized:

- a major QTL seems to play a "switching" role and is almost completely associated with the *occurrence* of the shattering;
- a second important locus on chr.4 correlates with the *level* and mode of the shattering;
- four other loci have some additional effect and act only on the *mode* of shattering.

Based on the histological and the genetic analyses of this PhD thesis, it can be hypothesized that the first two genes have different roles (that might interactor not): one determine the characteristics of the ventral sheath and the other the pattern of the inner layer of lignified fiber in the pod wall. This would be similar to the two-genes system described for soybean (Dong et al., 2014;Funazuki et al., 2014; Dong et al. 2015).

However, overall it is apparent that artificial selection might have targeted multiple cellular mechanisms and different controlling genes, to minimize seed loss during bean domestication. Meanwhile, these findings also raise an intriguing question on the way these genes genetically interact to fine-tune the indehiscence degree. Future analysis in particular devoted to the elucidation of the role of epistatic interaction among genes will probably shed some light on these aspects.

Major QTLs controlling the seed shattering were also found in soybean, the closest bean crop relative (Dong et al., 2014; Funatzuki et al. 2014). However none of the two genes cloned in soybean co-map with the QTL identified on chr5. This suggests that in bean the control of shattering might be different, at least partially, from that of soybean. This hypothesis is reinforced by the observation that the QTL found on bean chromosome 5 did not contain any genes with a NAC domain (which characterizes the soybean *SHAT1-5*; Dong et al., 2014) nor genes for protein belonging to the dirigent-like family (which is the family of the protein encoded by the soybean *PDH1*; Funatsuki et al., 2014). Thus, it could be inferred that between these two Leguminosae species parallel or convergent responses to selection under domestication might be limited to the phenotypic level while the underlying molecular mechanisms might be different for relevant aspects. However, it must be acknowledged that this assertion needs to be validated by the cloning of the bean shattering gene and the elucidation of the relative molecular mechanism.

The clarification of this aspect will be of great interest as this will allow the comparison between the Legume and cereal crops where convergent evolution has been described at molecular level in particular for *Sh1* under parallel selection during sorghum, rice and maize domestication. (Lin et al. 2012; Dong and Wang, 2015). Finally, with the limited intent to contextualize the observation made for pod shattering we also preliminary mapped the genomic region underlying the genetic control of this domestication trait. It emerged: the first is.: the first is that several regions scattered across the genome are involved without any very evident genomic “hot spot” for pod domestication, albeit some regions, such as the one on chromosome 1, seems to have a preeminent role. This was also previously noted by Koinange (1996). The second is that this analysis evidenced a certain degree of overlap among the QTLs for the different traits as also between shattering and other traits. However, pleiotropy and linkage drag with the examined traits cannot be invoked to explain the results obtained for qPD5.2-

Pv. Indeed, the magnitude of this QTL is much bigger than that for all of the others examined traits. The QTLs for the level and mode of shattering on chr.4 and chr8 are also stronger than their overlapping counterparts, while the QTL on chr2 did not overlap with other QTLs.

In perspective the present work represents a key step in the cloning of the genes involved in pod shattering of common bean. To this regard, the co-mapping of the major QTL for carbon content, considering that the fibers are mainly composed of sclerenchyma cells with well-developed secondary cell walls, suggested that genes involved in the regulation of the secondary cell wall deposition or in fiber cell differentiation may underly the major QTL found in this work on chromosome 5.

Bibliography

Abbo, S., Saranga, Y., Peleg, Z., Kerem, Z., Lev-Yadun, S., & Gopher, A. (2009). Reconsidering domestication of legumes versus cereals in the ancient Near East. *The Quarterly review of biology*, *84*(1), 29-50.

Abbo, S., Lev-Yadun, S., & Gopher, A. (2011). Origin of Near Eastern plant domestication: homage to Claude Levi-Strauss and "La Pensée Sauvage".*Genetic Resources and Crop Evolution*, *58*(2), 175-179.

Abbo, S., Lev-Yadun, S., & Gopher, A. (2012). Plant domestication and crop evolution in the Near East: on events and processes. *Critical Reviews in Plant Sciences*, *31*(3), 241-257.

Abbo, S., van-Oss, R. P., Gopher, A., Saranga, Y., Ofner, I., & Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in plant science*, *19*(6), 351-360.

Acosta-Gallegos, J. A., Kelly, J. D., & Gepts, P. (2007). Prebreeding in common bean and use of genetic diversity from wild germplasm. *Crop Science*, *47*(Supplement_3), S-44.

Akibode, S., & Maredia, M. (2011). Global and regional trends in production, trade and consumption of food legume crops. *Department of Agricultural, Food and Resource Economics, Michigan State University*, 87.

Angioi, S. A., Desiderio, F., Rau, D., Bitocchi, E., Attene, G., & Papa, R. (2009). Development and use of chloroplast microsatellites in Phaseolus spp. and other legumes. *Plant Biology*, *11*(4), 598-612.

Arnaud, N., Girin, T., Sorefan, K., Fuentes, S., Wood, T. A., Lawrenson, T., ... & Østergaard, L. (2010). Gibberellins control fruit patterning in Arabidopsis thaliana. *Genes & development*, *24*(19), 2127-2132.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., ... & Papa, R. (2012). Mesoamerican origin of the common bean (Phaseolus vulgaris L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences*, *109*(14), E788-E796.

Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., ... & Papa, R. (2013). Molecular analysis of the parallel domestication of the common bean (Phaseolus vulgaris) in Mesoamerica and the Andes. *New Phytologist*, *197*(1), 300-313.

Brown, T. A., Jones, M. K., Powell, W., & Allaby, R. G. (2009). The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology & Evolution*, *24*(2), 103-109.

Broughton, W. J., Hernandez, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans (Phaseolus spp.)–model food legumes.*Plant and soil*, *252*(1), 55-128.

Chiariello, N. R., Mooney, H. A., & Williams, K. (2000). Growth, carbon allocation and cost of plant tissues. In *Plant physiological ecology* (pp. 327-365). Springer Netherlands.

Cohen, M. N. (2009). Introduction: rethinking the origins of agriculture.*Current Anthropology*, *50*(5), 591-595.

Currence, T. M. (1930). Inheritance studies in Phaseolus vulgaris. *Tech-Bull. Minn. Agric. Expt. Stn.*, *68*.

Darwin, C. (1859). On the origin of the species by natural selection.

Deegan, R. D. (2012). Finessing the fracture energy barrier in ballistic seed dispersal. *Proceedings of the National Academy of Sciences*, 2012 .109(14), 5166-5169

Delgado-Salinas, A. Bonet, and P. Gepts. "The wild relative of Phaseolus vulgaris in Middle America." *Genetic resources of Phaseolus beans*. Springer Netherlands, 1988. 163-184.

Delgado-Salinas, A., Bibler, R., & Lavin, M. (2006). Phylogeny of the genus Phaseolus (Leguminosae): a recent diversification in an ancient landscape.*Systematic Botany*, *31*(4), 779-791.

Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, *127*(7), 1309-1321.

Doebley, J. (1989). Isozymic evidence and the evolution of crop plants. In*Isozymes in plant biology* (pp. 165-191). Springer Netherlands.

Dong, Y., Yang, X., Liu, J., Wang, B. H., Liu, B. L., & Wang, Y. Z. (2014). Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature communications*, *5*.

Dong, Y., & Wang, Y. Z. (2015). Seed shattering: from models to crops.*Frontiers in plant science*, *6*.

Drijfhout, E. (1978). Genetic interaction between Phaseolus vulgaris and bean common mosaic virus with implications for strain identification and breeding for resistance. *Verslagen van Landbouwkundige Onderzoekingen*, (872), 1-89.

Dwivedi, S. L., Upadhyaya, H. D., Stalker, H. T., Blair, M. W., Bertioli, D. J., Nielen, S., & Ortiz, R. (2008). Enhancing crop gene pools with beneficial traits using wild relatives. *Plant Breeding Reviews*, *30*, 179.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

Emerson, R. A. (1904). Heredity in bean hybrids. *Ann Rep Nebr Agric Exp St*,*17*, 33-78.

Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L., & Gaut, B. S. (1998). Investigation of the bottleneck leading to the domestication of maize.*Proceedings of the National Academy of Sciences*, *95*(8), 4441-4446.

Eshed, Y., & Zamir, D. (1994). Introgressions fromLycopersicon pennellii can improve the soluble-solids yield of tomato hybrids. *Theoretical and Applied Genetics*, *88*(6-7), 891-897.

Evans, L. T. (1996). *Crop evolution, adaptation and yield*. Cambridge University Press.

Frankel, O. H. (1974). Genetic conservation: our evolutionary responsibility.*Genetics*, *78*(1), 53-65.

Fuller, D. Q. (2007). Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World.*Annals of Botany*, *100*(5), 903-924.

Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., ... & Fujino, K. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences*, *111*(50), 17797-17802.

Gepts, P., Osborn, T. C., Rashka, K., & Bliss, F. A. (1986). Phaseolin-protein variability in wild forms and landraces of the common bean (Phaseolus vulgaris): evidence for multiple centers of domestication. *Economic botany*,*40*(4), 451-468.

Gepts, P. (2004). Crop domestication as a long-term selection experiment.*Plant breeding reviews*, *24*(2), 1-44.

Gioia, T., Logozzo, G., Kami, J., Zeuli, P. S., & Gepts, P. (2012). Identification and characterization of a homologue to the Arabidopsis INDEHISCENT gene in common bean. *Journal of Heredity*, ess102.

Girin, T., Paicu, T., Stephenson, P., Fuentes, S., Körner, E., O'Brien, M., ... & Østergaard, L. (2011). INDEHISCENT and SPATULA interact to specify carpel and valve margin tissue and thus promote seed dispersal in Arabidopsis.*The Plant Cell*, *23*(10), 3641-3653.

Glémin, S., & Bataillon, T. (2009). A comparative view of the evolution of grasses under domestication. *New phytologist*, *183*(2), 273-290.

Grant, W. F. (1996). Seed pod shattering in the genus Lotus (Fabaceae): A synthesis of diverse evidence. *Canadian Journal of Plant Science*, *76*(3), 447-456.

Grant, O. M., Chaves, M. M., & Jones, H. G. (2006). Optimizing thermal imaging as a technique for detecting stomatal closure induced by drought stress under greenhouse conditions. *Physiologia Plantarum*, *127*(3), 507-518.

Gu, Q., Ferrándiz, C., Yanofsky, M. F., & Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development*, *125*(8), 1509-1517.

Hammer, K. (1984). Das domestikationssyndrom. *Die Kulturpflanze*, *32*(1), 11-34..

Hawkes, J. G. (1983). The diversity of crop plants. *The diversity of crop plants.*

Harlan, J. R., De Wet, J. M. J., & Price, E. G. (1973). Comparative evolution of cereals. *Evolution*, 311-325.

Harlan, J. R. (1975). Crop and man. *Amer. Soc. Agron. Madison, Wisconsin*,*169*.

Harlan, J. R. (1992). Origins and processes of domestication. *Grass evolution and domestication*, *159*, 175.

Holechek, J. L., & Vavra, M. (1982). Comparison of micro-and macro-digestion methods for fiber analysis. *Journal of Range Management*, 799-801

Hymowitz, T. (1970). On the domestication of the soybean. *Economic Botany*,*24*(4), 408-421.

Jin, J., Huang, W., Gao, J. P., Yang, J., Shi, M., Zhu, M. Z., ... & Lin, H. X. (2008). Genetic control of rice plant architecture under domestication. *Nature genetics*, *40*(11), 1365-1369.

Joosten JHL: 1927. An inquiry into the str inglessness of several bean races. Meded. Landb. Wageningen 31(3)

Kang, S. T., Kwak, M., Kim, H. K., Choung, M. G., Han, W. Y., Baek, I. Y., ... & Lee, S. H. (2009). Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [Glycine max (L.) Merr.].*Euphytica*, *166*(1), 15-24.

Kerem, Z., Lev-Yadun, S., Gopher, A., Weinberg, P., & Abbo, S. (2007). Chickpea domestication in the Neolithic Levant through the nutritional perspective. *Journal of Archaeological Science*, *34*(8), 1289-1293.

Kluyver, T. A., Charles, M., Jones, G., Rees, M., & Osborne, C. P. (2013). Did greater burial depth increase the seed size of domesticated legumes?.*Journal of experimental botany*, *64*(13), 4101-4108.

Koenig, R., & Gepts, P. (1989). Allozyme diversity in wild Phaseolus vulgaris: further evidence for two major centers of genetic diversity. *Theoretical and Applied Genetics*, *78*(6), 809-817.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435-3436.

Koinange, E. M., Singh, S. P., & Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Science*, *36*(4), 1037-1045.

Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., & Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication.*Science*, *312*(5778), 1392-1396.

Ladizinsky, G. (1987). Pulse domestication before cultivation. *Economic Botany*, *41*(1), 60-65.

Lamprecht, H. (1932). BEITRÄGE ZUR GENETIK VON PHASEOLUS VULGARIS. *Hereditas*, *16*(1-2), 169-211.

Lenser, T., & Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends in plant science*, *18*(12), 704-714

Li, C., Zhou, A., & Sang, T. (2006). Rice domestication by reducing shattering. *science*, *311*(5769), 1936-1939.

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., & Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature*, *404*(6779), 766-770.

Liljegren, S. J., Roeder, A. H., Kempin, S. A., Gremski, K., Østergaard, L., Guimil, S., ... & Yanofsky, M. F. (2004). Control of fruit patterning in Arabidopsis by INDEHISCENT. *Cell*, *116*(6), 843-853.

Lin, Z., Li, X., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G., ... & Yu, J. (2012). Parallel domestication of the Shattering1 genes in cereals. *Nature genetics*, *44*(6), 720-724.

Liu, B., Fujita, T., Yan, Z. H., Sakamoto, S., Xu, D., & Abe, J. (2007). QTL mapping of domestication-related traits in soybean (Glycine max). *Annals of botany*, *100*(5), 1027-1038.

Loader, N. J., Robertson, I., & McCarroll, D. (2003). Comparison of stable carbon isotope ratios in the whole wood, cellulose and lignin of oak tree-rings. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *196*(3), 395-407.

Ma, Y., & Bliss, F. A. (1978). Seed proteins of common bean. *Crop Science*,*18*(3), 431-437.

Mao, L., Begum, D., Chuang, H. W., Budiman, M. A., Szymkowiak, E. J., Irish, E. E., & Wing, R. A. (2000). JOINTLESS is a MADS-box gene controlling tomato flower abscission zone development. *Nature*, *406*(6798), 910-913.

McGinley, M. A., & Charnov, E. L. (1988). Multiple resources and the optimal balance between size and number of offspring. *Evolutionary Ecology*, *2*(1), 77-84.

Mitra, P. P., & Loqué, D. (2014). Histochemical Staining of Arabidopsis thaliana Secondary Cell Wall Elements. *JoVE (Journal of Visualized Experiments)*, (87), e51381-e51381.

Mitsuda, N., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2005). The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *The Plant Cell*, *17*(11), 2993-3006.

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2007). NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. *The Plant Cell*, *19*(1), 270-280.

Mitsuda, N., & Ohme-Takagi, M. (2008). NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. *The Plant Journal*, *56*(5), 768-778.

Nakano, T., Kimbara, J., Fujisawa, M., Kitagawa, M., Ihashi, N., Maeda, H., ... & Ito, Y. (2012). MACROCALYX and JOINTLESS interact in the transcriptional regulation of tomato fruit abscission zone development. *Plant physiology*, *158*(1), 439-450.

Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., ... & Papa, R. (2011). Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (Phaseolus vulgaris L.). *Theoretical and applied genetics*, *123*(8), 1341-1357.

Ogawa, M., Kay, P., Wilson, S., & Swain, S. M. (2009). ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are polygalacturonases required for cell separation during reproductive development in Arabidopsis. *The Plant Cell*, *21*(1), 216-233.

Olsen, K. M., & Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual review of plant biology*, *64*, 47-70.

Olsen, K. M., & Wendel, J. F. (2013). Crop plants as models for understanding plant adaptation and diversification. *Frontiers in plant science*, *4*.

Papa, R., Nanni, L., Sicard, D., Rau, D., & Attene, G. (2006). The evolution of genetic diversity in Phaseolus vulgaris L. *New approaches to the origins, evolution and conservation of crops. Darwin's harvest. Columbia University Press, USA*.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one*, *7*(2), e32253.

Prakken, R. (1934). Inheritance of colours and pod characters in Phaseolus vulgaris L. *Genetica*, *16*(3), 177-296.

Purugganan, M. D., & Fuller, D. Q. (2011). Archaeological data reveal slow rates of evolution during plant domestication. *Evolution*, *65*(1), 171-183.

Rajani, S., & Sundaresan, V. (2001). The Arabidopsis myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Current Biology*,*11*(24), 1914-1922.

Rellstab, C., Zoller, S., Tedder, A., Gugerli, F., & Fischer, M. C. (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species.

Rick, C. M. (1974). *High soluble-solids content in large-fruited tomato lines derived from a wild green-fruited species*. University of California, Division of Agriculture and Natural Resources.

Rodríguez, G. R., Moyseenko, J. B., Robbins, M. D., Morejón, N. H., Francis, D. M., & van der Knaap, E. (2010). Tomato Analyzer: a useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *Journal of visualized experiments: JoVE*, (37).

Roeder, A. H., Ferrándiz, C., & Yanofsky, M. F. (2003). The role of the REPLUMLESS homeodomain protein in patterning the Arabidopsis fruit.*Current Biology*, *13*(18), 1630-1635.

Roth, I. (1977). *Fruits of angiosperms*. Borntraeger.

Salinas, A. D., Bonet, A., & Gepts, P. (1988). The wild relative of Phaseolus vulgaris in Middle America. In *Genetic resources of Phaseolus beans* (pp. 163-184). Springer Netherlands.

Santalla, M., Menéndez-Sevillano, M. C., Monteagudo, A. B., & De Ron, A. M. (2004). Genetic diversity of Argentinean common bean and its evolution during domestication. *Euphytica*, *135*(1), 75-87.

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., ... & Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics*, *46*(7), 707-713.

Schumacher, K., Schmitt, T., Rossberg, M., Schmitz, G., & Theres, K. (1999). The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHIID protein family. *Proceedings of the National Academy of Sciences*,*96*(1), 290-295.

Shelmidine, J. J., & Hartmann, R. W. (1984). Evidence for two recessive genes and two dominant genes controlling string development in one bean population [Phaseolus vulgaris]. *Annual Report of the Bean Improvement Cooperative*.

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., & Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics*, *172*(1), 547-555.

Singh, S. P., Gutierrez, J. A., Molina, A., Urrea, C., & Gepts, P. (1991). Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. *Crop Science*, *31*(1), 23-29.

Singh, S. P. (2001). Broadening the genetic base of common bean cultivars.*Crop Science*, *41*(6), 1659-1675.

Smartt, J., & Simmonds, N. W. (1995). *Evolution of crop plants* (No. Ed. 2). Longman scientific & technical.

Smith, B. D. (2006). Eastern North America as an independent center of plant domestication. *Proceedings of the National Academy of Sciences*, *103*(33), 12223-12228.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., ... & Belzile, F. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PloS one*, *8*(1), e54603.

Suzuki, M., Fujino, K., & Funatsuki, H. (2009). A major soybean QTL, qPDH1, controls pod dehiscence without marked morphological change. *Plant production science*, *12*(2), 217-223.

Suzuki, M., Fujino, K., Nakamoto, Y., Ishimoto, M., & Funatsuki, H. (2010). Fine mapping and development of DNA markers for the qPDH1 locus associated with pod dehiscence in soybean. *Molecular breeding*, *25*(3), 407-418.

Tang, H., Cuevas, H. E., Das, S., Sezen, U. U., Zhou, C., Guo, H., ... & Paterson, A. H. (2013). Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proceedings of the National Academy of Sciences*, *110*(39), 15824-15829.

Tiwari, S. P., & Bhatia, V. S. (1995). Characters of pod anatomy associated with resistance to pod-shattering in soybean. *Annals of botany*, *76*(5), 483-485

Tjebbes, K., & Kooiman, H. N. (1922). Erfelijkheidsonderzoekingen bij boonen. *Genetica*, *4*(5-6), 447-456.

.

Tshemark, E. VON, 1901. Weitere Beiträge über Verschiedenwertigkeit der Merkmale bei Kreuzung yon Erbsen und Bohnen. Zeitschr. f. d. landwirtsch. Versuchsw. in Oesterreich, IV, p. 641--735.

Tshemark, E. VON, 1902.Über die Gesetzmäissige Gestaltungsweise der Mischlinge. Zeitschr. f. d. landwirtsch. Versuchsw. in Oesterreich, V, p. 781--861.

Tsuchiya, T. (1987). Physiological and genetic analysis of pod shattering in soybean. *Japan Agricultural Research Quarterly*, *21*(3), 166-175.

Vaughn, K. C., Bowling, A. J., & Ruel, K. J. (2011). The mechanism for explosive seed dispersal in Cardamine hirsuta (Brassicaceae). *American journal of botany*, 2011. 98(8), 1276-1285

Van Soest, P. U., & Wine, R. H. (1967). Use of detergents in the analysis of fibrous feeds. IV. Determination of plant cell-wall constituents. *J. Assoc. Off. Anal. Chem*, *50*(1), 50-55.

Holechek, J. L., & Vavra, M. (1982). Comparison of micro-and macro-digestion methods for fiber analysis. *Journal of Range Management*, 799-801.

Welch, R. M., House, W. A., Beebe, S., & Cheng, Z. (2000). Genetic Selection for Enhanced Bioavailable Levels of Iron in Bean (Phaseolus v ulgaris L.) Seeds. *Journal of Agricultural and Food Chemistry*, *48*(8), 3576-3580.

Wellensiek, S. J. (1922). De erfelikheid van het al of niet bezit van „draad" bij rassen van Phaseolus vulgarisl. *Genetica*, *4*(5), 443-446.

Wilke, P. J., Bettinger, R., King, T. F., & O'Connell, J. F. (1972). Harvest selection and domestication in seed plants. *Antiquity*, *46*(183), 203-209.

Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., & Gaut, B. S. (2005). The effects of artificial selection on the maize genome. *Science*, *308*(5726), 1310-1314.

Yamada, T., Funatsuki, H., Hagihara, S., Fujita, S., Tanaka, Y., Tsuji, H., ... & Hajika, M. (2009). A major QTL, qPDH1, is commonly involved in shattering resistance of soybean cultivars. *Breeding science*, *59*(4), 435-440.

Yoon, J., Cho, L. H., Kim, S. L., Choi, H., Koh, H. J., & An, G. (2014). The BEL1-type homeobox gene SH5 induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. *The Plant Journal*, *79*(5), 717-728.

Zohary, D. (1989). Pulse domestication and cereal domestication: how different are they?. *Economic Botany*, *43*(1), 31-34.

Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford University Press on Demand.

Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J., Huneau, C., ... & Chalhoub, B. (2011). Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proceedings of the National Academy of Sciences*, *108*(46), 18737-18742.

Zhou, Y., Lu, D., Li, C., Luo, J., Zhu, B. F., Zhu, J., ... & Han, B. (2012). Genetic control of seed shattering in rice by the APETALA2 transcription factor SHATTERING ABORTION1. *The Plant Cell*, *24*(3), 1034-1048.