UNIVERSIDAD POLITÉCNICA DE CARTAGENA

DEPARTAMENTO DE CIENCIA Y TECNOLOGÍA AGRARIA

INSTITUTO DE BIOTECNOLOGÍA VEGETAL

Ph.D Thesis

Genotyping, phenotyping and transcriptomic analysis of accessions of *Vicia faba, Pisum sativum* and *Vigna unguiculata*

Marina Marta Martos Fuentes

Supervised by

Dr. Catalina Egea Gilabert

Dr. Julia Weiss

Cartagena, 2017

Genotyping, phenotyping and transcriptomic analysis of accessions of *Vicia faba, Pisum sativum* and *Vigna unguiculata*

Marina Marta Martos Fuentes

Cartagena, 2017

Supervised by:

Dr. Catalina Egea Gilabert & Dr. Julia Weiss

AGRADECIMIENTOS

En primer lugar tengo que agradecer a mis directoras de la tesis Caty y Julia todo su esfuerzo para que esto haya salido adelante y haberme dado la oportunidad de ampliar mis conocimientos de genética vegetal.

A Juan Fernández por haber confiado en mí, ya que sin él no hubiera podido realizar la tesis.

A Marcos Egea por todo el apoyo que me ha dado a lo largo de todos estos años y todo lo que me ha enseñado no solo a nivel académico.

A Jesús Ochoa por su ayuda en los últimos ensayos de campo.

A los técnicos del IBV Perla y Mariano por haber estado ahí siempre que los he necesitado.

A los nuevos técnicos, Álvaro, María y Sara, por hacer más divertida la vida del laboratorio.

A Francisco, Bernardo Martín, Eva Armero y toda la gente de la Finca Tomás Ferro (Sergio, Isa, Jose, Fina, Carlos y Conchi) que me han ayudado en todos mis ensayos.

A todos l@s amig@s, compañer@s y profesores que a lo largo de este camino han estado dándome palabras de ánimo y apoyo y con los que he podido compartir muchas historias para recordar. Pero en especial a Maricarmen López, César Petri, Marta Terry, Claudio Brandoli y Victoria Ruiz.

A todos los becarios y estudiantes que me han ayudado en mis tareas porque sin ellos la vida del laboratorio sería muy aburrida.

A Nerik por haber sido tanto en los días malos como en los buenos la persona que me ha iluminado las mañanas y ha velado mis sueños.

A mis abuelos y a mis padres por toda su paciencia, apoyo y cariño para poder llegar hasta aquí.

A mi hermana porque a pesar de la distancia, siempre me motiva en todos mis proyectos e inquietudes.

PUBLICATIONS RELATED TO THIS PhD

- **Scientific journals**

Martos-Fuentes M, Fernández J A, Ochoa J, Carvalho M, Carnide V, Rosa E, Pereira G, Barcelos C, Bebeli P J & Egea-Gilabert C (2017). Genotype by environment interactions in cowpea (*Vigna unguiculata* L. Walp.) grown in the Iberian Peninsula. *Crop and Pasture Science*. https://doi.org/10.1071/CP17071 CHAPTER 1. Journal with H-index 72, Q1.

- **Conference Proceedings**

Martos-Fuentes M, Sánchez-Navarro V, Ruiz-Hernández V, Weiss J, Egea-Gilabert C, Zornoza R, Faz A, Fernández J A y Egea-Cortines M. Genetic and growth conditions determine the protein content in cowpea (*Vigna unguiculata*). EUCARPIA International Symposium on Protein Crops: V Meeting AEL [V JORNADAS DE LA AEL], Pontevedra, Spain. 4-7 May, 2015. Plant Proteins for the Future. ACTAS AEL 6 2nd ed, pp. 143-144.

Martos-Fuentes M, Ruiz-Hernández V, Esteva J, Weiss J, Egea-Gilabert C, Fernández J A y Egea-Cortines M. Caracterización de variedades autóctonas de caupí (*Vigna unguiculata* (L.) Walp.). XLV Seminario de Técnicos y Especialistas en Horticultura. Aragón, 15-19 Junio, 2015.

Fernández J A, Martos-Fuentes M, Weiss J, Egea-Gilabert C & Egea-Cortines M. Morphological and physiological characterization of 15 accessions of cowpea. Second International Legume Society (ILS2) Conference. Troia (Portugal), 11-14 October, 2016.

Martos-Fuentes M, Appeltans J, Maes I, Mezaka I, Egea-Gilabert C, Fernández J A, Egea-Cortines M & Weiss J. Genotyping of *Pisum sativum* L. and *Vicia faba* L. by SNP

analysis using next generation sequencing. Second International Legume Society (ILS2) Conference. Troia (Portugal), 11-14 October, 2016.

Martos-Fuentes M, Egea-Gilabert C, Fernández J A, Weiss J y Egea-Cortines M. Caracterización morfológica de accesiones de caupí del sureste europeo XLVII Seminario de Técnicos y Especialistas en Horticultura. Murcia, 24-27 Abril, 2017.

- **In progress**

Martos-Fuentes M, Egea-Gilabert C, Mezaka I, Kronberga A, Kokare A, Rabante L, Fernández J A, Egea Cortines M & Weiss J. A barcoding procedure for parallel genotyping of multiple accession using Next-Generation Sequencing. CHAPTER 2.

Weiss J, Martos-Fuentes M, Letourneux L, Terry M I, Ruiz-Hernández V, Fernández J A & Egea Cortines M. Diel pattern of circadian clock and storage protein gene expression during seed filling in cowpea (*Vigna unguiculata*). CHAPTER 3.

INDEX

FIGURES INDEX

TABLES INDEX

ABREVIATIONS

ACT: *B-ACTIN*

ACT27: *ACTIN 2/7*

AtTOC1: *Arabidopsis thaliana TIMING OF CAB EXPRESSION1*

cDNA: complementary DNA

CCA1: *CIRCADIAN CLOCK-ASSOCIATED 1*

CYP*: CYCLOPHYLIN*

CV: Coefficient of variation model.

CVC*: CONVICILIN*

DD: Day-degrees

DNA: Deoxyribonucleic acid

E: Environment

EDD: Effective day-degrees

EF1A: *ELONGATION FACTOR 1-A*

EF1B: *ELONGATION FACTOR 1-B*

ELF3*: EARLY FLOWERING3*

ELF4: *EARLY FLOWERING4*

G: Genotype

GI: *GIGANTEA*

GmCCA1: *Glycine max CIRCADIAN CLOCK ASSOCIATED1*

GmELF4*: Glycine max EARLY FLOWERING4*

GmG1: *Glycine max GIGANTEA*

GmLHY*: Glycine max LATE ELONGATED HYPOCOTYL*

GmPRR: *Glycine max PSEUDO RESPONSE REGULATOR*

GmTOC1: *Glycine max TIMING OF CAB EXPRESSION1*

ha: Hectare

kg: Kilogram

LEG: Legumin

LHY*: LATE ELONGATED HYPOCOTYL*

LUX*: LUX ARRHYTHMO*

***PRR**: PSEUDO RESPONSE REGULATOR*

***SKIP16:** ASK-INTERACTING PROTEIN 16*

***TOC1**: TIMING OF CAB EXPRESSION1*

***TUA4**: ALPHA TUBULIN*

***TUB4**: BETA TUBULIN*

***UKN2**:* HYPOTHETICAL UNKNOWN PROTEIN

***VIC**: VICILIN*

***VuCVC**: Vigna unguiculata CONVICILIN*

***VuELF3**: Vigna unguiculata EARLY FLOWERING3*

***VuGI**: Vigna unguiculata GIGANTEA*

***VuLEG**: Vigna unguiculata LEGUMIN*

***VuLEGJ:** Vigna unguiculata LEGUMIN J*

***VuLHY**: Vigna unguiculata LATE ELONGATED HYPOCOTYL*

***VuTOC**: Vigna unguiculata TIMING OF CAB EXPRESSION1*

***VuVIC**: Vigna unguiculata VICILIN*

sec: Seconds

m^2: Square metre

mg: Milligram

MJ: Mega joule

mm: Millimetre

ng: Nanogram

Mt: Metric tons

NGS: Next-Generation Sequencing

NPK: Nitrogen, phosphorus and potassium

PAR: Photosynthetically active radiation

PCR: Polymerase chain reaction

RNA: Ribonucleic acid

RT-qPCR: Real time quantitative polymerase chain reaction

R: Rainfall

SNP: Single Nucleotide Polymorphism

T: Temperature

T_m: Mean temperature

T_b: Base temperature

T_M: Maximum temperature

Y: Year

µg: Microgram

RESUMEN

Las legumbres son el segundo cultivo con mayor producción mundial después de los cereales, por lo que su importancia para consumo animal y humano es crucial. El haba (*Vicia faba*), el guisante (*Pisum sativum*) y el caupí (*Vigna unguiculata*) son especies de leguminosas con interés agronómico en la Unión Europea. El objetivo de esta tesis fue la selección de variedades de estas especies, mediante técnicas de fenotipado, genotipado y análisis transcriptómico, que permitirá la obtención de líneas de mejora con cualidades nutricionales y agronómicas óptimas para una producción sostenible y competitiva de proteínas en Europa. Dentro de las cualidades nutricionales nos centramos sobre todo en la cantidad de proteína en semilla. Y en cuanto a las características agronómicas nos centramos en la resistencia a estrés hídrico y en la productividad.

Dentro de los estudios realizados con el caupí, en el primero de ellos se evaluaron 12 genotipos en tres localizaciones distintas de la Península Ibérica para determinar los componentes de la varianza y su estabilidad genética y ambiental. Los resultados mostraron variaciones interesantes entre las accesiones estudiadas las cuales podrían incorporarse en un programa de mejora.

En cuanto al haba y el guisante, se hizo un análisis de genotipado de diferentes accesiones europeas. En este estudio se usó un protocolo simple codificando las muestras permitiendo el genotipado de material vegetal basado en un número mínimo de lecturas de secuencias mediante Next-Generation Sequencing (NGS). Los resultados demostraron la viabilidad de NGS para genotipar múltiples muestras usando la codificación de estas.

Adicionalmente se estudió en caupí si la síntesis de proteínas de esta leguminosa está controlada por el reloj circadiano de la planta, donde se usó una línea de referencia. Para ello se analizaron la expresión de genes de referencia (como *ELONGATION FACTOR 1-A*), de ritmo circadiano (*VuLHY, VuTOC1, VuGI* y *VuELF3,*) y de almacenamiento de proteínas (*VuLEG, VuLEGJ,* y *VuCVC*) en distintos tejidos de la planta (hojas, vainas y semillas) con diferentes estadios de desarrollo. Con este estudio se confirmó que la síntesis de proteínas de almacenamiento puede estar regulada por el ritmo circadiano

de la planta. Hay que tener en cuenta los cambios en la expresión génica afectados por el ritmo circadiano, para poder analizar los patrones de la expresión de estos genes, por ejemplo para analizar las accesiones que difieren en el contenido de proteínas.

ABSTRACT

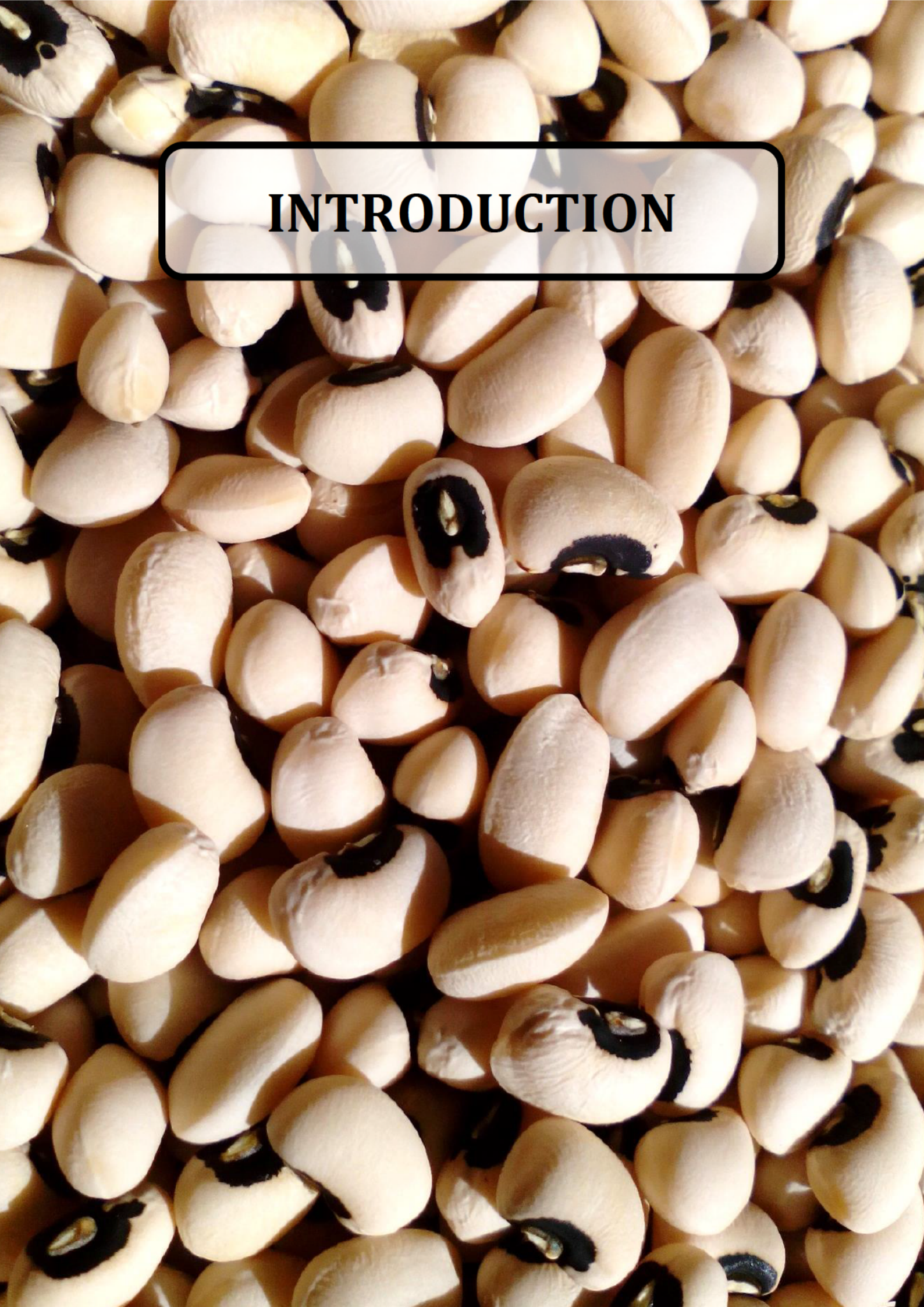
Legumes are the second crop with the highest world production after cereals, so its importance for animal and human nutrition is crucial. Faba bean (*Vicia faba*), pea (*Pisum sativum*) and cowpea (*Vigna unguiculata*) are legume species with agronomic interest in the European Union. The main objective of this project was the selection of varieties of the above mentioned species, using techniques of phenotyping, genotyping and transcriptomic analysis. This techniques allow us the improvement of varieties with optimum nutritional and agronomic qualities, for a sustainable and competitive production of proteins in Europe. Within the nutritional qualities we focus mainly on the amount of seed protein. And as for the agronomic characteristics, we focus on resistance to drought.

Among the studies carried out with cowpea, 12 genotypes were evaluated at three different locations in the Iberian Peninsula. Components of variance and their genetic and environmental stability were determined for these genotypes. The results demonstrate interesting variations among the studied accessions with which a breeding program could be performed.

As for bean and pea, a genotyping analysis of different European accessions was perfomed with the aim to analyse the genetic relation among these accessions. Genotyping of accessions is a prerequisite for the exploitation of natural genetic variation. In this study we developed a simple protocol based on sample barcoding allowing the genotyping of plant material based on a minimum number of Next-Generation Sequencing (NGS) reads. The results demonstrated the feasibility of NGS-based genotyping using multiple barcoded samples.

A further study was conducted on cowpea with the aim to investigate whether the protein synthesis is controlled by the circadian clock of the plant. A reference line of cowpea was used for this experiment. The expression of reference genes (such as *ELONGATION FACTOR 1-A*), circadian rhythm genes (*VuLHY*, *VuTOC1*, *VuGI* and *VuELF3*) and protein storage genes (*VuLEG*, *VuLEGJ*, and *VuCVC*) was measured in different tissues (leaves, pods and seeds) at different stages of development. This study confirmed that storage protein deposition may be circadian regulated by the circadian

clock. Diel changes in expression need to be taken into account when analysing expression patterns of these genes i.e. in accessions differing in protein content.



INTRODUCTION

INTRODUCTION

1. General history and taxonomy of legumes

Legumes were the earliest food crops cultivated. Indeed, the first legumes crop go back to Neolithic times. In some archaeological researches, remains of peas, lentils, and a legume called "blue vetchling" have been found at Halicar (Turkey) and in Jarno (Turkistan) dated about 5500 B.C. In addition, remains of lentils (*Lens culinaris*) have been found in Egyptian tombs of the 12th dynasty (2400-2200 B.C.), and the preparation of lentil soup is depicted in a fresco of the time of Ramses II (1200 B.C.). For the ancient Egyptians lentils were a very important crop, they esteemed and cultivated them carefully and extensively (Aykroyd & Doughty, 1982).

Legumes also appear early in the development of agriculture in the New World. In Mexico, remains of kidney beans (*Phaseolus vulgaris*) and the Lima bean (*Phaseolus lanatus*) were found, dated at 4000 B.C. and about 1500 B.C. respectably (Aykroyd & Doughty, 1982).

Moreover there are some legumes depictions in the earliest printed botanical illustrations, in the famous herbals of Fuch (1542), Bock (1556), Dalechamps (1587), and Malpighi (1679) (Allen & Allen, 1981).

The taxonomy of grain legumes is relatively uncomplicated compared to that of cereals, brassicas and some other groups of plants because, in general, only limited gene pools have been available for selection and subsequent plant breeding (Polhill & Van der Maesen, 1985). This enormous family, with a worldwide distribution, has a currently estimated 16,000 to 19,000 species in about 750 genera. Taxonomists conventionally have divided the family into three clearly distinct subfamilies, *Mimosoideae*, *Caesalpinioideae* and *Papilonoideae* (Allen & Allen, 1981).

2. History and morphology of legumes under study

Pea, faba bean and cowpea are the species studied in the present thesis, which following are described.

- **Pea:**

Pea (*Pisum sativum* L.) has been found growing wild from the Mediterranean to India. Probably peas are originated in Abyssinia and Afghanistan (Cousin, 1997). Remains of cultivated peas have been discovered in the excavations at Aggetelek in Hungary, in a Neolithic lake-dwellings in Switzerland (ca. 4500 B.C.), in the Predynastic Egyptian tombs, and in the ruins of Troy (Fourmont, 1956; Aykroyd & Doughty, 1982). Some authors like Smartt (1990) and Erskine et al. (1994) indicate that peas date back to 7000-6000 B.C. Until relatively recent times peas were eaten mainly in the dried form; fresh green peas became popular only in the 17th century, in France at the court of King Louis XIV, as is showed in the Figure 1. It may perhaps be recalled that in the mid-19th century Mendel used this legume to found the modern science of genetics (Aykroyd & Doughty, 1982).

Figure 1. Still Life with Strawberries of a French painter of 17th century (http://www.metmuseum.org/art/collection/search/436345).

In botanical terms, pea has indeterminate growth habit. Generally pea has two flower per node. The flower is typical of the *Papilionaceae* family, with five petals, the standard, two wings and a keel formed as the result of fusion of two petals (Figure 2) (Cousin, 1997). The flower can be white, cream white, rose or purple. The fruit is a pod with multiple curvatures and lengths depending on the variety. The curvature may be pronounced, very weak or absent (Red Andaluza de Semillas, 2017). There is a large genetic diversity in other traits like the number of seeds per pod, width and pod length, number of branches, thousand-seed weight, leaf area and height of the plant. All these traits can change the yield depending on the environmental conditions (Cousin, 1997).

Figure 2. Morphology of *Pisusm sativum* L. (Thomé, 1885)

- **Faba bean:**

Faba bean (*Vicia faba* L.) is a one of the most widespread and cultivate legume crops in the world, and like peas have a very long history. Beans have been growing wild from immemorial times. A variety of *V. faba* has been found in Switzerland in deposits ascribed to the Bronze Age, but probably is originating in North Africa.

In terms of botany, faba bean is an annual plant that requires cool environments, for that is grown in the spring in northern latitudes and in the winter in warm-temperate and subtropical areas. There are different kinds of cultivars adapted to these specific environments, which have differences in several common traits like number of stems. Stems have an indeterminate growth and develop flowers since 5th to 10th node. In each reproductive node there is a raceme of 2 to 12 flowers axillary to the leaf. Flowers, as well as in peas, have a typically papilionaceous structure (Figure 3). Flower colour can vary between white, brown or violet. Pods can be short or long and erected or hanging. Seeds per pod can vary between 3 to 12 seeds (Duc, 1997).

Figure 3. Morphology of *Vicia faba* L. (Thomé, 1885)

- **Cowpea:**

Cowpea (*Vigna unguiculata* L. Walp.) is a crop that come from Africa. The precise place of the origin is difficult to determine, but according to Padulosi & Ng (1997), the greatest genetic diversity and most primitive of the wild cowpea occur in southern Africa. The speciation of cowpea probably was in the Transvaal region of the Republic of South Africa (Padulosi & Ng, 1997).

Cowpea is an annual plant that requires warm environments and is one of the most widely adapted, versatile, and nutritious grain legumes (Ehlers & Hall, 1997). As the others species described before, cowpea has flowers with a typically papilionaceous structure (Figure 4). Flower colour can be white, yellow, red and purple. Cowpea growth is determinate or indeterminate and climbing or not. Shape of seeds is kidney, globose, ovoid, rhomboid or crowder (IBPGR, 1983). In cowpea there are four cultivar-groups, *Unguiculata*, *Biflora*, *Sesquipedalis* and *Textilis* (Ng & Marechal, 1985). Length pod is a trait that can vary depending on the cultivar-group, the genotypes that belongs to *Sesquipedalis* have long pod characters (Figure 4) (Ogunkanmi et al., 2013).

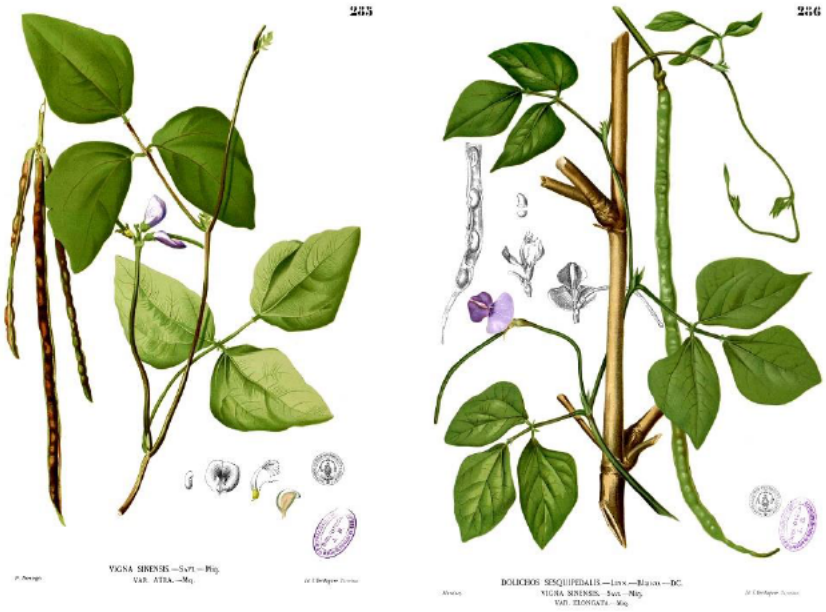


Figure 4. Morphology of *Vigna unguiculata* L. Walp (Blanco, 1877)

3. Importance of legumes

Legumes are second only to the *Gramineae* in their importance to humans (Graham & Vance, 2003), mainly by its nutritional characteristics and its benefits for the environment. The role of legumes in the human diet and nutrition of animals and farming systems is increasingly important as they have a low environmental impact compared to other crops or protein rich sources (Tilman & Clark, 2014).

Grain and forage legumes are grown on about 180 million hectares, or 12 - 15% of the Earth's arable land and it accounts for 27% of the world's primary crop production (Graham & Vance, 2003). Worldwide, legumes provide at least 33% of the dietary protein nitrogen (N) needs of humans, however, in the tropics and subtropics plant sources provide much more (up to 80%) of the dietary N requirements (Vance et al., 2000).

In respect of the species studied, cultivated pea since centuries has been an important source of animal feed and human food. There are many diverse types and systems of cultivation of pea widely dispersed around the world (Redden, 2005), and peas have been selected for many uses, as for fresh peas, canning or freezing peas or for dried and forage peas, but they are mostly grown for dry peas (Cousin, 1997).

Same goes for the faba bean, that is used as human food but also for animal feed depending on the country of production (Duc, 1997). As far as the cowpea is traditional crop in the semi-arid regions of the tropics and an important food legume. Worldwide there are about 14 million ha of cowpea, which only West Africa accounts for about 9.3 million ha with annual production of about 2.9 Mt. Seeds are consumed as human food and the haulms are a nutritious fodder for livestock. (Tarawali et al., 1997, Singh et al., 2003). Generally for human food dry seeds of cowpea are the most important part, but leaves and immature pods are also consumed.

- **Nutritional importance:**

Legumes have been an important source of protein, starch, oil, minerals, vitamins and health protecting compounds from the beginning of human history. Their seeds, which contain up to 250 g protein/kg, play an important role in the traditional diet of many peoples of the world and are a valuable basic material for the food and animal feed industries (Schuster-Gajzágó, 2004).

In addition to traditional food and forage uses, legumes can be milled into flour, used to make bread, chips, doughnuts, tortillas, spreads, and extruded snacks or used in liquid form of milks, yogurt, and infant formula (Gupta et al., 2014).

Seed composition is one of the most important trait in legumes since the last researches are focus in nutrition values. In most faba bean varieties vicine and convicine

are present, which cause favism in humans and their elimination is the objective of several breeding programs. On the other hand in the feed industry are more focus in the seed protein content and in tannins. Especially for poultry high protein content is very important. Seed protein content in faba beans ranges between 27 to 34% of dry matter, where 80% are vicilin and legumin. On the other hand, tannins are antinutritional and reduce the protein digestibility in faba bean (Duc, 1997).

According to Iqbal (2006), cowpea and pea seeds have approximately 25% of protein content. Moreover, apart from of a significant amount of dietary protein (18-35%), cowpea grains provide a source of calories, vitamins, minerals and essential amino acids (Singh, 2002).

- **Environmental importance:**

  - **Nitrogen fixation:**

As it is known, Nitrogen is required by all living organisms for the synthesis of proteins, nucleic acids and other nitrogen containing compounds. Legumes play a critical role in natural ecosystems, agriculture, and agroforestry, since they have the ability to fix N in symbiosis. Legumes are an economic and an environmentally friendly crop because they can growth without deficiencies in low-N environments. Thanks to *Rhizobium* bacteria legumes can fix N through symbiosis. *Rhizobium* is the most well-known species of a group of rhizobacteria that acts as the primary symbiotic fixer of atmospheric nitrogen. These rhizobacteria can infect the roots of leguminous plants, leading to the formation of lumps or nodules where the nitrogen fixation takes place. About 40-60 million metric tons (Mt) of atmospheric nitrogen are fixed by agriculturally important legumes annually, with another 4-5 million Mt fixed by legumes in natural ecosystems (Graham & Vance, 2003).

  - **Tolerance to abiotic stresses:**

Abiotic stresses reduce productivity in crops in such a way that they end up leading to a great economic loss. Among all abiotic stresses, drought is a major agronomic problem resulting a reduction in crop yield (Somerville & Briscoe, 2000;

Chaves et al., 2003). It is known that water is a scarce commodity, and is becoming limited for agriculture in many areas of the world, above all in rapid expansion of water-stressed areas (Barrera-Figueroa et. al, 2011).

According to FAO (2017), nowadays, droughts are more disastrous and more frequent. This problem, combined with underlying economic, social and environmental vulnerabilities, makes droughts a destructive impact on at-risk populations. Figure 5 shows the Agricultural Stress Index (ASI) of 2016 in the World. ASI represents the percentage of cropland area affected by drought.

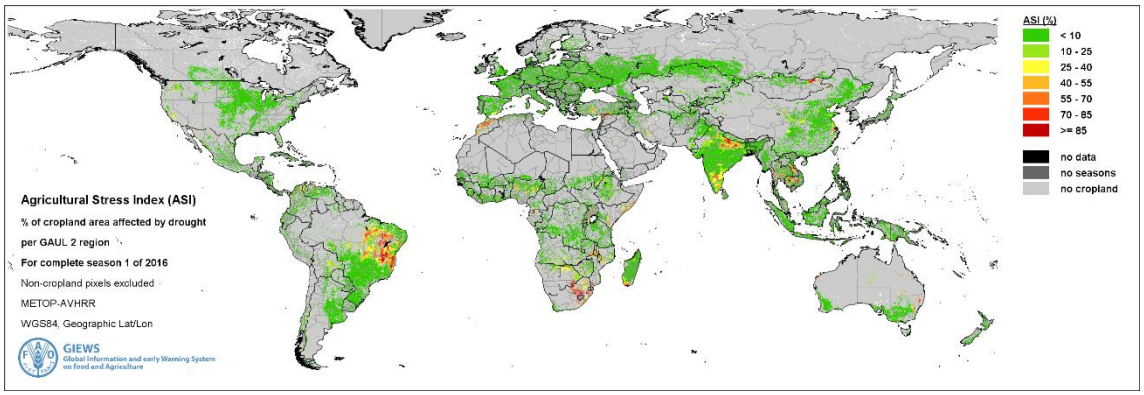


Figure 5. Agricultural Stress Index (ASI) of 2016 in the World. (http://www.fao.org/giews/earthobservation/)

Drought is one of the prime abiotic stresses in the world (Ashraf, 2010). It is a complex syndrome involving several climatic, edaphic and agronomic factors (Serraj et al., 2003) and it cause reduction or total loss of crop production (Barrera-Figueroa et. al, 2011). There is a crucial need to increase drought tolerance in legumes in many areas (Graham & Vance, 2003), therefore there are many studies of agronomic and genetic management strategies focused on drought tolerance breeding (Serraj et al., 2003). But despite the demand for drought-resistant cultivars, there is still work to be done in order to achieve the identification of traits that reflect true drought resistance (Amede et al., 2003).

The more drought-tolerant legumes, such as cowpea, are deeply rooted and may have reduced leaf size with thickened cuticles to reduce water loss. (Graham & Vance, 2003). Moreover, cowpea is well adapted to semi-arid and arid tropical regions in Africa, Asia, and Central and South America because of its ability to fix nitrogen in poor soil and due to its greater drought tolerance (Barrera-Figueroa et al., 2011).

4. The soybean issue:

Nowadays, soybean is globally the main protein source used (FAO, 2004). The European livestock and feed industry is becoming increasingly reliant on imported soya bean meal whose higher price is prompting investigation into home-grown protein alternatives. Domestic legumes like peas and faba beans offer a possibility to increase the protein self-sufficiency in livestock feeding, thus promoting reliance on home-grown feedstuffs (O'Neill et al., 2012; Palander et al., 2006).

Due to this problem, the last studies about livestock feeding are related to feeding values, above all to amino acid ideal digestibility in peas and faba beans for different poultries (Koivunen et al., 2016; O'Neill et al., 2012; Palander et al., 2006).

In addition, not only alternatives to animal feed are emerging, but also for humans. Indeed, within the Eurolegume project, some new food products based on pea and faba bean purees are being developed (Klug et al., 2017). As well as, new ways of storage and packaging for a better conservation of these fresh products (Collado et al., 2017). According to Oton (2017), faba bean is a food appreciated by the consumer, as much for its flavor, forms of preparation as for its nutritional values. Proof lies in the fact that faba bean has entered into a food category known as minimally processed products in fresh, offering the consumer the possibility of consumption with all its properties without the need to shrink and extract the seeds allowing its use in a fast, healthy and safe way.

5. Genomics in legumes

The technologies such as transcriptomics, proteomics and metabolomics are being developed in major legume species with the aim to analyse molecular data on a genome wide scale (Wienkoop et al., 2010). These developments are now becoming major landmarks in understanding legume biology in a precise manner (Gupta et al., 2014).

Genetic modification approaches can be used to enhance legume productivity and reduce crop loss by making plants tolerant to pests, weeds, herbicides, insects,

viruses, salinity, temperature, pH, drought, frost, and weather (Paoletti & Pimental, 2006).

On the other hand, the control of host-specificity in the *Rhizobium*–legume symbiosis has been a topic of long-standing interest to plant biologists, thus there are recent studies have begun to reveal the genetic and molecular basis of Nod-factor perception in legumes, a signalling system that also controls plant interactions with mycorrhizal fungi (Riely, 2004).

- **Gene expression in legumes**

  - **Circadian clock genes**

The circadian clock is a cellular mechanism which generates rhythms in biological processes in all living organisms. Nowadays there are many genetic analyses that have uncovered a lot of components and molecular mechanisms comprising the clock. Due to the circadian clock, plants can synchronise their metabolism, physiology, biological processes, reproduction and development to anticipate diurnal and seasonal environmental changes (Young & Kay, 2001; Bendix et al., 2015; Takahashi et al., 2015).

The circadian clock regulates a number of central plant activities through controlling a substantial proportion of transcriptional activity and protein function. All the genes subject to circadian clock regulation are central to many important physiological processes. For that reason, it is so important to study the circadian clock genes, since they regulate flower development, flowering time, light harvesting, yield, biomass, carbon fixation, nitrogen assimilation and growth habit. In fact, there are emerging insights into the responses to abiotic and biotic stresses related to circadian clock regulation that can help for further crop improvement (Kreps & Kay, 1997; Bendix et al., 2015). Due to this, the expression of circadian clock genes during the seed filling in cowpea were evaluated in the current thesis.

Circadian clock has been deeply analyzed above all in *Arabidopsis thaliana* in which the roles of circadian clock genes like *PSEUDO-RESPONSE REGULATORs*, *GIGANTEA*, *CIRCADIAN CLOCK ASSOCIATED 1, LATE ELONGATED HYPOCOTYL*, *TIMING OF CAB EXPRESSION 1* and the evening complex genes *EARLY FLOWERING 3*, *EARLY*

FLOWERING 4 and *LUX ARRHYTHMO* have been analyzed (Alabadi et al., 2001; Kinmonth-Schultz et al., 2013; Bendix et al., 2015; Takahashi et al., 2015).

There are some studies in pea concerning identified orthologues to the circadian clock genes. Their roles in the changes of flowering phenology, photoperiod responsiveness, deetiolation and growth habit have been studied (Hecht et al., 2007; Liew et al., 2009; Weller et al., 2012). But in legumes there is yet a lot to be investigated about circadian clock.

- **Seed protein storage genes**

The seed storage proteins are classified depending on their extraction and solubility in different solutions. Albumins are soluble in water, globulins in dilute saline, prolamins in alcohol-water mixtures and glutelins in dilute acid or alkali. Albumins, globulins and prolamins are the major seed storage proteins (Osborne, 1924). These storage proteins determine not only the total protein content of the seed but also its quality for various end uses. Knowing the proportion of these proteins it is possible to determine the nutritional quality of the seeds (Shewry et al., 1995).

Legumin and vicilin are the major seed globulins in pea and faba bean. These storage proteins are inherently heterogeneous (Scholz et al., 1974; Derbyshire et al., 1976; Thompson et al., 1991). In *Vicia sativa* L., vicilin represents the initial source of amino acids for early growth and differentiation processes and legumin is a bulk amino acid source for subsequent seedling growth during post-germinated globulin degradation (Tiedemann et al., 2000). Convicilins are also storage proteins described only in species of the legume tribe Vicieae. These differ from vicilins mainly because of the presence of the N-terminal extension, an additional sequence of amino acids in the sequence corresponding to the first exon (Sáenz de Miera et al., 2008).

In pea, it was determined whether the gene expression of pea legumin and pea vicilin is regulated transcriptionally or post-transcriptionally in response to changing sulfur status in the developing seed (Beach et al., 1985). And pea vicilin and convicilin gene expression have been evaluated in transgenic tobacco plants (Higgins et al., 1988; Newbigin et al., 1990). But despite the existence of some researches about storage protein gene expression, there is little information on the expression patterns of these

genes in different legume tissues. This thesis describes the diel gene expression patterns of *LEGUMIN, LEGUMINJ* and *COVICILIN* during pod and seed development.

- **Next-Generation Sequencing**

Nowadays Next-Generation Sequencing (NGS) is a technique that has caused a revolution in biology and has been a great advance in genomics (Koboldt et al., 2013; van Dijk et al., 2014). In NGS requires the preparation of libraries. The libraries are fragments of DNA or RNA fused with adapters followed by PCR amplification and sequencing (van Dijk et al., 2014). In sequencing, the major challenge is the methods for data storages and data analysis, because the classical statistical methods are inadequate for analysing a big amount of genomic sequence data (Xiong et al., 2010).

NGS has been used recently in *Vicia faba* and in *Pisum sativum*. In case of *Vicia faba*, a large scale of SSR marker has been developed and analysed by NGS. These markers can be useful for constructing genetic linkage maps, QTL mapping, assessing genetic diversity and marker-assisted trait selection facilitating bean breeding programs (Yang et al., 2012; Chung et al., 2013; Lavania et al., 2015).

In *Pisum sativum*, comprehensive genomic resources are recently developed. This specie has a large and complex genome, but fortunately through NGS, repetitive regions of the pea genome as well as transcriptome of diverse pea tissues have been efficiently analysed (Novák et al., 2010; Franssen et al., 2011; Alves-Carvalho et al., 2015). These data can be used to assess the variability and evolutionary divergence of repeat families, discover and characterize novel elements, and aid in subsequent assembly of their consensus sequences (Novák et al., 2010; Franssen et al., 2011). Indeed, recently an Unigene set of pea is available online (http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi), which represents most of the expressed genes of pea. To create the Unigene around one billion reads and 100 Gb of sequence were *de novo* assembled (Alves-Carvalho et al., 2015).

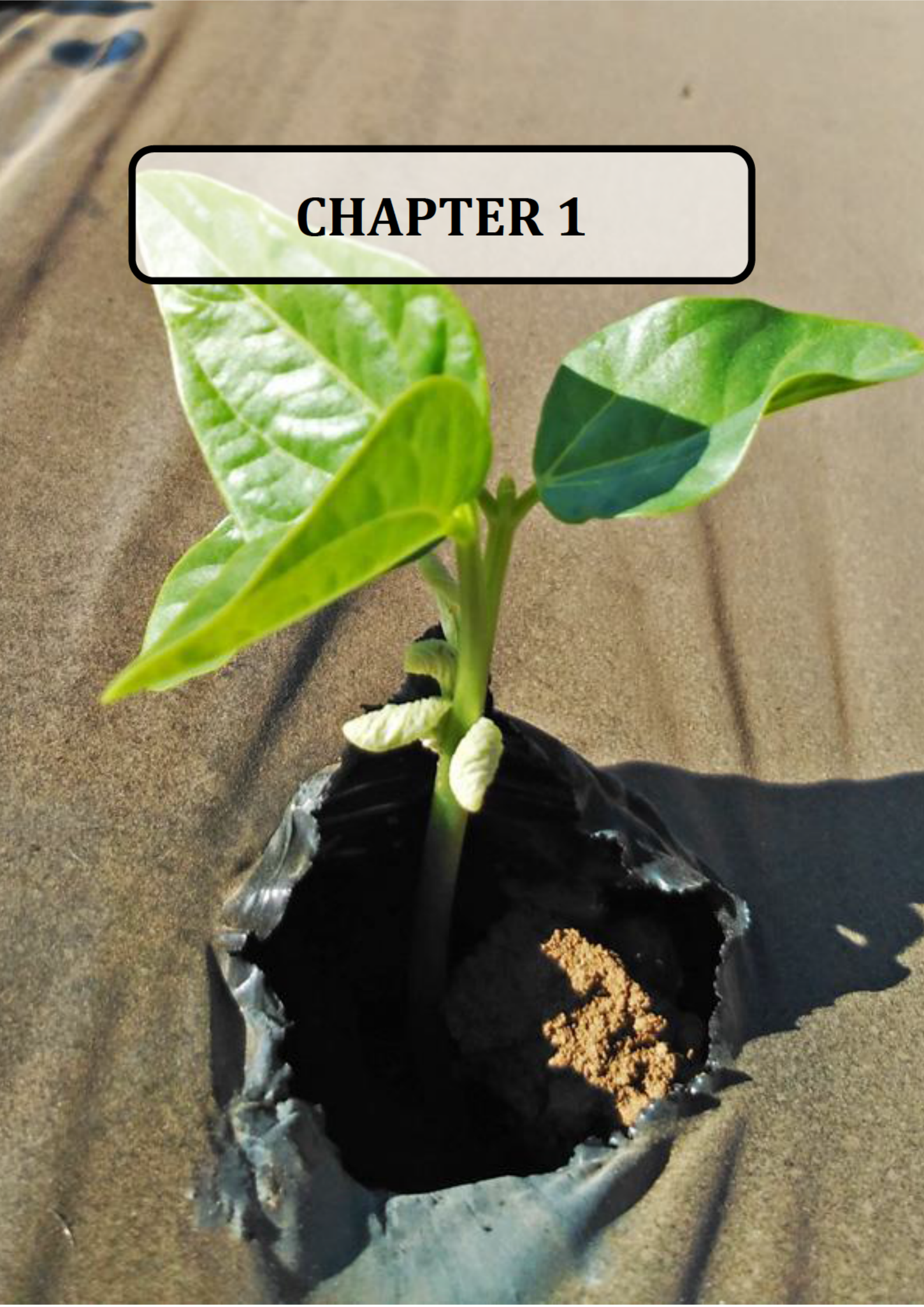
In the current thesis NGS was used to identify genetic diversity between different accessions of peas and faba beans. For this analysis we selected SNPs marker based on sequence information available for high-density genetic maps of *P. sativum* and *V. faba* (Kaur et al., 2014; Leonforte et al., 2013). SNPs are an important source of polymorphism

among genotypes. So sequencing with the use of SNP markers helps us to identify species and to make linkage mapping (Ganal et al., 2009)

6. Objectives

Due to the importance of legumes and the problem of soybean in Europe, this project was developed with a selection of varieties of faba bean (*Vicia faba*), pea (*Pisum sativum*) and cowpea (*Vigna unguiculata*). The main objectives of the research are described below:

- Selection and phenotyping of faba bean (*Vicia faba*), pea (*Pisum sa*tivum) and cowpea (*Vigna unguiculata*) accessions adapted to the Mediterranean climate.
- Genotyping of the Mediterranean accessions and compare them with accessions adapted to cold areas.
- Identify the expression of genes related to protein quality and quantity.
- Obtain an in-depth knowledge on gene expression and regulation of traits such as seed development and composition. To know the role of the circadian rhythm genes in the filling of the grains.



CHAPTER 1

CHAPTER 1 - Genotype by environment interactions in cowpea (*Vigna unguiculata* L. Walp.) grown in the Iberian Peninsula

1.1. Introduction

Cowpea (*Vigna unguiculata* L. Walp.) is originated in Southern Africa, belongs to the family *Fabaceae*, tribe *Phaseoleae* and genus *Vigna*, which comprises several species, subspecies and varieties depending on morphology and domestication (Padulosi & Ng, 1997). Cultivated cowpea belongs to *V. unguiculata*, spp. *unguiculata*, which contains the cultigroups Unguiculata, Biflora, Sesquipedalis and Textilis (Ng & Marechal, 1985). This annual warm-season legume is one of the most widely adapted, versatile, and nutritious grain legumes (Ehlers & Hall, 1997). During 2010-2014 period, the world cowpea planting area was 58.1 million hectares and the production was 33.5 million tons. Africa has been responsible for 95.8% of worldwide cowpea production (FAOStat, 2017). Nigeria and Niger are the largest producers with 3.4 and 1.6 million tons, respectively. In contrast, Europe is only responsible for 0.4% of worldwide cowpea production (FAOStat, 2017). Nowadays, cowpea is mainly grown by subsistence farmers in west and central sub-Saharan Africa, but also is an important food source in the rest of Africa, Central and South America, Southeast Asia and in Southern United States (Davis et al., 1991; Timko & Singh, 2008). In addition, cowpea is being cultivated at small scale in many parts of Southern Europe and countries around the Mediterranean Basin (Domínguez-Perles et al., 2015), providing these countries a considerable income through exports to Northern European and non-European countries (EC, 2016). Like other grain legumes, cowpea has the capacity to establish association with nitrogen-fixing bacteria (like rhizobia) and vesicular-arbuscular mycorrhizal fungi that makes this crop interesting for predicted climatic changes. Cowpea can be used for human food and for fodder livestock (Tarawali et al., 1997). For human food, dry grain is the most important part, but leaves and immature pods are also consumed. Dry grains provide a significant amount of dietary protein (18-35%), as well as, a source of calories, vitamins, minerals and essential amino acids as lysine and tryptophan (Singh, 2002). For all this

proprieties, this is an attractive crop with which many research is being done to promote it and include it in diets, not only because of its protein content but also because other functional properties, such as chlorophylls, carotenoids and phenolic contents, and high antioxidant activity (Khalid et al., 2012; Campbell et al., 2016; Karapanos et al., 2017). However, the value of grain legumes as a source of nutrients depends on a plethora of factors, including genetic characteristics, agro-climatic conditions, and postharvest management (Gonçalves et al., 2016).

The environment plays a very important role in the development and growth of plants. The major driving force that pushes crop growth and development is temperature although there are other environmental factors that can modify the effect of temperature such as PAR radiation or photoperiod. Locations, growing seasons, rainfall, etc. may have positive or negative impacts on several plant species as well as in cowpea genotypes. The thermal time concept or the accumulation of temperature for a life cycle or a particular phase of plant development, in contrast to the chronological time, has been used frequently to study the cowpea development, with the advantage to be independent of location and time of sowing. Craufurd et al. (1997) have described the effects of photoperiod and temperature on several development stages. Thus, the base temperature for development of seed germination, seedling emergence, leaf appearance, and days from sowing to first flowering is 8-11°C and the optimum temperature for most rapid reproductive development is close to 28°C. In addition, inclusion of radiation will allow describing development when temperature is not the only environmental variable affecting the process (Jones, 2014). Thus, the 'effective day-degrees' can be used to combine both temperature and radiation effects on plant development (Scaife et al., 1987). To our knowledge, no previous information exists on the effects of both temperature and radiation on cowpea development.

The association between the environment and the phenotypic expression of a genotype constitute the GxE interaction which determines if a genotype is widely adapted for an entire range of environmental conditions or separate genotypes must be selected for different subenvironments. Presence of the GxE interaction indicates that the phenotypic expression of one genotype might be superior to another genotype in one environment but inferior in a different environment (Falconer & Mackay, 1996). Most of the studies in cowpea have been carried out on the genotypic variability and

stability of some grain yield components (e.g. Akande, 2007; Adewale et al., 2010; Shiringani & Shimelis 2011), showing generally significant GxE interactions. In addition, the protein content in seeds is also influenced by environmental and genotypic factors, being negatively correlated to yield (Oluwatosin, 1997). Therefore, G x E should be taken into account in any breeding program.

Thus, the aim of this work was to determine the variance components and genetic and environmental stability of 12 selected cowpea genotypes at three locations of the Iberian Peninsula in two consecutive years. The results of this study may assist cowpea breeders in the manipulation of interested traits.

1.2. Material and methods

1.2.1 Plant material

Ten cowpea landraces (five from Portugal, three from Spain and two from Greece), one commercial variety from Portugal and one advanced line from Nigeria (Table 1) were used in three field experiments in 2015 and in 2016.

Table 1. Cowpea accessions, origin and breeding status.

Accession	Origin	Status of accession
IT 97K-499-35	Nigeria	Advanced line
AUA1	Greece	Landrace
AUA2	Greece	Landrace
Cp 4877	Portugal	Landrace
Cp 5051	Portugal	Variety
Cp 5553	Portugal	Landrace
Vg 59	Portugal	Landrace
Vg 60	Portugal	Landrace
Vg 73	Portugal	Landrace
BGE038479	Spain	Landrace
BGE038474	Spain	Landrace
BGE038478	Spain	Landrace

The accessions were selected based on previously studies that were developed in the three locations where morphological and agronomical characteristics were evaluated. The agronomic characterization of the twelve genotypes was done at: Universidad Politécnica de Cartagena (UPCT), Cartagena, Spain (N 37°36'; W 00°58'; 40 m) - field experiment 1; National Institute for Agrarian and Veterinarian Research

(INIAV), Elvas, Portugal (N 38°53', W 07°09', 208 m) - field experiment 2; University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal (N 41°17'51", W 07°44'12", 465 m) - field experiment 3.

1.2.2. Field experiment 1

Cultivars were planted on 29 May 2015 and 15 June 2016 in a randomized complete block design with four replications. One row per plot with 8 m length, 0.9 m row spacing and 7 m^2 were used. Seeds were sown by hand and seed rate was 10 seeds/m^2. The topsoil (0-20 cm) was classified as clay loam with a medium texture in both growing seasons, and presented 1.97 organic matter, 78 mg/kg of P_2O_5, 354 mg/kg of K_2O_2 and pH (KCl) 8.4 in 2015 growing season, and 2.18 organic matter, 80.13 mg/kg of P_2O_5, 415.82 mg/kg of K_2O_2 and a pH (KCl) 8.3 in 2016. Before sowing, in both growing seasons, the experimental field was ploughed with a rotary tiller and fertilized with 30 kg/ha of ammonium nitrate, 170 kg/ha of potassium nitrate and 250 kg/ha of monoammonium phosphate. The trails were drip irrigated from beginning June until end of September.

1.2.3. Field experiment 2

Cultivars were planted on 28 April 2015 and 24 May 2016 in a randomized complete block design with four replications. Two row plots with 3 m length, 0.6 m row spacing and 3.6 m^2 were used. Seeds were sown by hand and seed rate was 11 seeds/m^2. The topsoil (0-20 cm) was classified as sandy clay loam with a medium texture in both growing seasons, and presented 1.3 g/kg organic matter, >200 mg/kg of P_2O_5, 153 mg/kg of K_2O_2 and pH (KCl) 6.9 in 2015 growing season, and 0.80 g/kg organic matter, >200 mg/kg of P_2O_5, >200 mg/kg of K_2O_2 and pH (KCl) 6.4 in 2016. Before sowing, the experimental fields were ploughed with a rotary tiller and fertilized with 200 kg/ha of 15:15:15. The trials were drip irrigated from beginning May until end of August.

1.2.4. Field experiment 3

Cultivars were planted on 11 May 2015 and 3 June 2016 in a randomized complete block design with four replications. Three row plots with 3 m length, 0.75 m row spacing and 6.7 m^2 were used. Seeds were sown by hand and seed rate was 11 seeds/m^2. The topsoil (0-20 cm) was classified as gleyic fluvisol with a medium texture in both growing seasons, and presented in 2015 1.29 g/kg organic matter, 36 mg/kg of P_2O_5, 103 mg/kg of K_2O_2 and a pH (KCl) 4.2, while in 2016 1.61 g/kg humus content, 44 mg/kg of P_2O_5, 110 mg/kg of K_2O_2 and a pH (KCl) 5.2. Before sowing in both growing seasons, the experimental field was ploughed with a rotary tiller and fertilized with 250 kg/ha of nitromagnesium 27 and 200 kg/ha of NPK (Ca-Mg-S) 8-12-12 (2-2-14). The trails were drip irrigated from beginning July until end of August.

1.2.5. Climatic data and calculation of accumulated day-degrees and effective day-degrees

The mean daily air temperature, total rainfall (mm) and accumulated global radiation (MJ/m^2) from April to September for each experiment are presented in Table 2.

Table 2. Mean temperatures, precipitation and global radiation from April to September 2015 and 2016 in each location.

Environment/ Month	Year	Cartagena			Elvas			Vila Real		
		T (°C)	R (mm)	Solar radiation (MJ/m^2)	T (°C)	R (mm)	Solar radiation (MJ/m^2)	T (°C)	R (mm)	Solar radiation (MJ/m^2)
April	2015	16.0	10.2	596.02	16.6	110.3	630.57	13.5	48.8	453.97
	2016	16.1	14.6	627.72	14.3	80.6	648.39	11.0	193.0	456.62
May	2015	20.2	0.0	825.38	22.0	2.8	813.55	17.5	69.6	713.86
	2016	18.6	3.0	791.06	17.2	119.7	717.96	14.2	124.4	536.18
June	2015	23.1	1.6	876.83	25.6	37.9	820.66	20.9	2.2	729.61
	2016	22.8	0.0	853.27	23.7	0.0	953.87	19.1	25.2	759.73
July	2015	27.2	0.6	852.9	26.5	0.0	847.19	22.5	0.4	781.33
	2016	25.4	0.0	825.15	28.5	0.1	956.20	23.8	0.2	821.62
August	2015	27.2	1.0	693.22	25.4	0.9	809.93	20.9	0.6	647.58
	2016	25.5	1.2	759.95	27.2	0.1	858.85	23.3	0.2	657.18
September	2015	22.8	72.6	519.29	22.1	32.8	604.26	17.4	1.2	489.99
	2016	23.7	25.0	587.09	24.0	0.0	671.23	19.6	28.4	515.56

Summations of heat units were determined based on base temperature using the coefficient of variation model (CV) to identify the accurate base temperature to adjust the method, according to Ochoa et al. (2011). The base temperatures tested ranged from 0°C to 16°C. The following methods were used:

Method 1. Standard day-degrees method: $DD = \Sigma (T_m - T_b)$ where, T_m and T_b are the daily mean and base temperatures respectively.

Method 2. Use of maximum instead mean temperature: $DD = \Sigma (T_M - T_b)$ where, T_M and T_b are the daily maximum and base temperatures respectively.

Method 3. The day-degrees method modified by the effect of the daily photosynthetic radiation input or effective day-degrees (EDD), calculated according to following equation: $EDD^{-1} = DD^{-1} + f\ PAR^{-1}$, where PAR is photosynthetically active radiation (MJ/m^2 day) and f is a constant that defines the relative importance of radiation and temperature (m^2/MJ).

DD and EDD were calculated considering the climatic conditions from sowing to flowering and sowing to maturity.

1.2.6. Morphological and agronomical traits

Phenotypic data for days to flowering and maturation were collected when 50% of the plants begin to flower and have mature pods, respectively. Plant height, first pod height, pod length and width and number of seeds per pod were measured in ten plants per plot randomly selected. Yield, adjusted to 12% moisture, and 100-seed weight were evaluated per plot. Protein content (AOAC, 1990) was derived from the estimated Nitrogen (N) content, which was determined by the Kjeldahl method (Bremner, 1960), by the following formula: Protein content (%) = N content (%) × 6.25.

1.2.7. Data analysis

Analysis of variance (ANOVA) of the three factors (genotype, location and year) followed by the Tukey's test was performed for each parameter in each environment and in the assembly of the three environments using the IBM SPSS Statistics 20 software.

A complete linear mixed model was used to estimate variance components of parameters in the analysis of all the quantitative parameters within and across the accessions and locations using Restricted Maximum Likelihood (REML) algorithm of SPSS program version 8.0. The heritability of each quantitative parameter was calculated for each environment using the following equation: $h^2 = V_g^2 / [V_g^2 + (V^2/r)]$, where V_g^2 and V^2 represent genotypic and error variance for each parameter and r the number of replications. For the three environments, the heritabilities were calculated using the equation $h^2 = V_g^2 / [V_g^2 + (V_{ge}^2/e) + (V^2/re)]$, where V_{ge}^2 is the GxE interaction variance and e is the number of environments (Mendes-Moreira et al., 2015).

Pearson correlation coefficients between the different quantitative parameters and environments were determined through SPSS program version 8.0.

The principal components analysis (PCA) was performed using the MVSP program version 3.22.

1.3. Results and discussion

The plant genetic resources collections provide genetic variants, genes or genotypes that allow breeders to respond to new challenges based on systems of high production, high nutritional quality and disease and environmental resistance/tolerance. In the present study, we evaluated 12 cowpea accessions growing in three locations in the Iberian Peninsula (Southeast of Spain: Cartagena, South of Portugal: Elvas and North of Portugal: Vila Real) during 2 years (2015, 2016) to identify morphological and agronomical parameters and the interactions among genotypes, environment and year.

In general, Vila Real registered the lowest temperatures and the lowest solar radiation. It is worth to highlights that the rainfall in Vila Real was four-fold in 2016 than in 2015, while Cartagena had the driest conditions during the studied period (Table 2).

Analysis of variance to determine the effects of genotype, environment, year(Y) and their reciprocal interactions (GxE; GxY; ExY; GxExY) on 10 morphological and agronomical parameters showed that all the factors had a high influence on the majority of the parameters (Table 3). These findings are according to those obtained by Shimelis & Shiringani (2010), who showed significant interactions among genotypes, locations

and planting dates in cowpea. The genotype and the environment significantly influenced all the parameters evaluated. Year effect was also an important factor affecting all parameter except first pod height, pod length and number of seed per pod. The GxE interaction was significant for all parameter, but GxY interaction was only significant for days to flowering and to maturity, first pod height, seed yield and number of seeds per pod. The ExY interaction affected all parameters, except first pod height and pod width. Finally, the GxExY interaction was significant for all parameters, except 100-seed weight, pod length and protein content (Table 3). This high variability among the cowpea accessions indicates their utility in breeding programs.

Table 3. Analysis of variance for the 10 morphological and agronomical parameters evaluated in 12 cowpea accessions at 3 environments (UPCT, INIAV, UTAD) during 2 years (2015, 2016).

Paramaters	ANOVA						
	G	E	Y	GxE	GxY	ExY	GxExY
Days to flowering	***	***	***	***	***	***	***
Days to maturity	***	***	***	***	***	***	***
Plant height	***	***	***	***	n.s.	***	*
First pod height	***	*	n.s.	***	**	n.s.	*
Seed yield	***	***	***	***	**	***	***
100-Seed Weight	***	***	***	***	n.s.	**	n.s.
Pod length	***	**	n.s.	***	n.s.	**	n.s.
Pod width	***	***	***	***	n.s.	n.s.	*
Number of seeds / pod	***	***	n.s.	***	*	**	***
Protein content	***	*	***	***	n.s.	***	n.s.

n.s. – no significant; * - significant at P<0.05; ** - significant at P<0.01; *** - significant at P<0.001

The duration of the periods sowing to flowering and sowing to maturation were affected by the three factors and their interactions. In Cartagena, the days from sowing to maturity were the shortest, whereas in Elvas were the longest in both years (Table 4). Also in Cartagena the time from flowering to maturity was the shortest in both years, probably due to the effects of high temperature and radiation in this period (Table 2and Table 4).

The analysis of the three methods showed the least CV was obtained with the Method 3 (Table 5), demonstrating that PAR radiation had an important effect on the duration of crop cycles in all accessions. The best fit for *f* ranged from 0.11 to 0.12. The accurate base temperature for all methods and accessions ranged from 2°C to 14°C, varying in some accessions for each calculation method and period. This temperature range differed to that proposed by Craufurd et al. (1997), who fixed 8-11°C for development of cowpea cultivated in Nigeria. An explanation of our different findings

could be due to the base temperature drops with the increase of the daily thermal amplitude (Bonhomme, 2000), higher in our conditions than in Nigeria.

Table 4. Means and standard deviation of the 10 morphological and agronomical parameters evaluated in 12 cowpea accessions at 3 environments (Cartagena, Elvas, Vila Real) during 2 years (2015, 2016).

Parameters	2015			2016		
	Cartagena	Elvas	Vila Real	Cartagena	Elvas	Vila Real
Days to flowering	75.58±10.34 a	66.44±6.28 c	70.00±9.21 b	59.21±9.13 b	67.94±13.54 a	69.98±3.56 a
Days to maturity	86.42±10.08 c	101.29±4.60 a	89.04±9.01 b	69.58±8.88 b	89.27±11.53 a	89.25±3.9 a
Plant height (cm)	212.47±63.26 a	123.87±61.36 b	57.07±32.58 c	212.72±62.33 a	146.38±60.67 b	80.05±37.39 c
First pod height (cm)	37.48±7.54 a	38.70±9.15 a	39.22±10.55 a	38.88±7.24 ab	36.97±4.78 b	41.84±8.59 a
Seed yield (g/m^2)	89.84±33.24 b	197.69±103.21 a	95.28±43.39 b	102.43±44.18 b	312.06±122.89 a	65.68±26.56 b
100-Seed Weight (g)	15.91±2.26 c	17.27±4.52 b	18.63±5.39 a	15.90±2.32 c	17.92±5.10 b	19.93±5.43 a
Pod length (cm)	17.16±3.75 a	16.94±2.82 a	16.63±1.81 a	17.10±3.89 a	16.38±3.21 b	17.14±1.94 a
Pod width (cm)	0.88±0.09 a	0.78±0.08 b	0.46±0.09 c	0.89±0.08 a	0.80±0.10 b	0.47±0.09 c
Number of seeds/pod	11.60±1.26 a	11.63±1.03 a	11.19±1.13 a	11.88±0.92 a	10.87±1.02 b	11.30±1.17 b
Protein content (%)	21.71±2.52 a	21.69±1.78 a	22.41±0.98 a	23.88±2.26 a	22.44±1.14 b	22.39±1.15 b

For each year, means followed by the same letter in the row are not significantly different at the 0.05 level using Tukey test, n = 4

Table 5. The base temperature for each cowpea accession and over growing periods incorporating PAR radiation.

Accession	Tbase (°C)		*f*		CV (%)	
	S-F	S-M	S-F	S-M	S-F	S-M
IT-97K-499-35	5	5	0.12	0.11	10.61	12.80
AUA1	2	2	0.12	0.11	11.18	07.80
AUA2	2	9	0.12	0.11	06.27	10.24
Vg 59	2	2	0.12	0.11	20.39	13.07
Vg 60	10	10	0.12	0.11	17.83	12.85
Vg 70	2	2	0.12	0.11	10.87	10.28
Cp 4487	2	2	0.12	0.11	10.33	09.80
Cp 5051	14	14	0.12	0.11	09.73	12.20
Cp 5553	2	2	0.12	0.12	10.83	09.49
BGE038479	7	11	0.12	0.12	04.97	04.83
BGE038474	12	12	0.12	0.11	10.59	02.25
BGE038478	2	2	0.12	0.11	10.94	08.63

Tbase is the base temperature. *f* is a constant that defines the relative importance of radiation and temperature in the Method 3 (EDD calculation) as described before. CV is the coefficient of variation expressed as a percentage. S-F is the growing period from sowing to 50% of flowering. S-M is the growing period from sowing to maturity of pods.

The seed yield was also affected by the three factors and their interactions (Table 3). For both years, the highest yields were observed in the trials located in Elvas (Table 4). In the second year (2016), the seed yield average increased in Cartagena and Elvas, while decreased in Vila Real (data not shown). At Cartagena, the most productive accessions were BGE038474 and IT97K-499-35 in 2015 and BGE038474 in 2016 (Table 6). In this location, the yield ranged from 52 to 165.4 g/m^2 and from 63.8 to 226.5 g/m^2 in 2015 and in 2016, respectively. At Elvas, Cp 5051 and Vg73 were the most productive in 2015, whereas in 2016 the most productive was Vg73. The yield ranged from 18.49 to 329.6 g/m^2 in 2015 and from 120.8 to 514.4 g/m^2 in 2016. The commercial variety Cp

5051 revealed to be one of the well-adapted accessions to this environment, this result could be expected due to this variety was selected at the INIAV Breeding Station in Elvas. Finally, in the first year, the most promising accessions in Vila Real were Cp 5553, Vg 60 and Vg 73, whereas in 2016 the most productive was AUA1. The seed yield varied from 34.3 to 167.2 and from 39.4 to 123.0 g/m^2 in 2015 and 2016, respectively. In general, the most productive accessions in each location were those that originally came from their own country, due to they are better adapted to their environmental conditions.

Table 6. Seed yield (g/m^2) for the 12 cowpea accessions evaluated at three environments (Cartagena, Elvas, Vila Real) during 2 years (2015, 2016).

Acessions	2015			2016		
	Cartagena	Elvas	Vila Real	Cartagena	Elvas	Vila Real
IT 97K-499-35	142.43±26.51 a	247.17±93.29 abc	49.25±20.91 b	130.02±45.87 b	152.32±20.68 d	40.05±24.21 b
AUA1	52.00±13.89 c	84.14±17.41 de	101.05±55.20 ab	86.65±28.95 b	263.75±99.86 cd	123.03±47.77 a
AUA2	94.40±13.22 b	258.20±56.64 abc	76.73±28.21 ab	86.77±26.71 b	292.35±55.32 bcd	45.90±27.41 ab
Cp 4877	64.98±11.41 bc	169.89±37.26 bcd	89.50±31.05 ab	88.98±18.69 b	237.05±46.19 cd	39.35±16.62 b
Cp 5051	77.15±18.93 bc	329.56±45.06 a	79.45±17.04 ab	82.65±39.32 b	310.90±80.93 bcd	46.53±33.97 ab
Cp 5553	97.53±26.03 b	279.93±81.56 ab	167.15±71.96 a	122.23±40.51 b	506.82±91.66 a	82.58±51.01 ab
Vg 59	67.85±12.81 bc	231.72±41.56 abc	57.10±14.26 b	70.20±13.59 b	465.75±137.68 ab	57.05±22.51 ab
Vg 60	64.53±11.09 bc	257.20±58.07 abc	153.58±58.14 a	83.58±9.88 b	352.35±53.57 abc	103.80±9.51 ab
Vg 73	75.20±4.38 bc	310.67±95.24 a	157.78±38.44 a	63.83±17.34 b	514.37±85.65 a	61.80±13.58 ab
BGE038479	83.55±9.77 bc	40.66±2.36 de	34.33±15.44 b	73.43±20.33 b	197.82±41.13 cd	44.40±15.05 ab
BGE038474	165.35±1.11 a	35.98±9.42 e	78.68±27.45 ab	226.45±60.19 a	217.12±43.74 cd	71.65±31.79 ab
BGE038478	91.23±2.66 b	127.12±18.31 cde	98.75±12.80 ab	114.38±15.77 b	234.11±4.71 cd c	72.18±51.72 ab

Means followed by the same letter in the column for each year are not significantly different at the 0.05 level using Tukey test, n=4

The highest plant height and pod width were observed in Cartagena in both years (Table 4). In 2015, the first pod height, pod length and number of seeds per pod did not differ among the trial places. Vila Real was the location in which the seeds reached the highest 100-seed weight in both years. The seed size, measured as 100-seed weight, is one of the most important parameter for the consumer's preference.

As regards protein content, it was influenced by genotype, environment and their interaction (Table 3), in agreement with the results obtained by Oluwatosin (1997) with 15 cowpea cultivars grown in three locations in Nigeria and Ravelombola et al. (2016) who grown 11 cowpea breeding lines in three locations in Arkansas. The highest percentage was found in Cartagena in 2016 (ca 24% in average) (Table 4). The values of protein content obtained in this study are in agreement with the results found in literature (Singh, 2002; Timko et al., 2007).

In general, correlation coefficients between the 10 parameters in the three environments and two years were not too high (Table 7). The highest correlation coefficient was between days to flowering and days to maturity (r=0.737, *p*=0.01) and between plant height and pod width (r=0.488, *p*=0.01). The correlation between days to flowering and days to maturity was expected because they are closer in the plant development. Plant height and 100-seed weight showed the highest negative correlation (r=-0.360, *p*=0.01) (Table 7), which shows that selection for the increase of plant height can induce a reduction in the 100-seed weight. There was also a negative correlation between the beginning of flowering and seed production in 2016 as it was reported by Silva et al. (2014). The seed yield and protein content exhibited negative correlations, which is agreement to the results obtained by Oluwatosin (1997) in cowpea and by Simmonds (1995) in cereals, and consequently indicates some restrictions in breeding alongside for high-yielding and high-protein genotypes. For the first pod height and seed yield, a positive correlation was registered. And for number of seeds per pod and 100-seed weight it was negative in agreement with the result obtained by Silva et al. (2014). The correlation between pod length and 100-seed weight was positive as the results obtained by Peksen & Artik (2004).

Table 7. Pearson correlation coefficients for 10 morphological and agronomical parameters for 12 cowpea accessions in the three environments (Cartagena, Elvas, Vila Real) and two years (2015, 2016).

	Days to flowering	Days to maturity	Plant height	First pod height	Seed yield	100 Seed Weight	Pod length	Pod width	Number Seeds per Pod	Protein content
Days to flowering	1									
Days to maturity	0.737**	1								
Plant height	0.089	-0.178**	1							
First pod height	0.118*	0.084	0.210**	1						
Seed yield	-0.277**	0.018	-0.006	0.029	1					
100 Seed Weight	-0.142*	-0.031	-0.360**	-0.092	0.092	1				
Pod length	-0.220**	-0.231**	0.167**	0.186**	-0.033	0.176**	1			
Pod width	-0.150*	-0.234**	0.488**	-0.149*	0.240**	0.077	0.055	1		
Number of seeds/pod	-0.108	-0.116*	0.160**	0.142*	0.076	-0.216**	0.227**	0.047	1	
Protein content	-0.148*	-0.255**	0.115	0.105	-0.172**	-0.106	0.196**	0.025	0.081	1

* - Correlation significant at *p*<0.05; ** - Correlation significant at *p*<0.01

Heritability reflects the genetic variability that is transmitted from parents to their offspring (Robinson et al., 1949). Heritability, in broad-sense, estimates across

environments, ranged from 0.29 for seed yield to 0.91 for pod width (Table 8). In general, it was higher at the Cartagena than at the other environments, with the exception of days to flowering, days to maturity, 100-seed weight and number of seeds per pod. The parameters plant height and pod length had the highest heritability at UPCT (0.99), while at Elvas and Vila Real was 100-seed weight that had the highest heritability (0.99) (Table 8). A hundred had high values of heritability in the three environments (0.94 at UPCT, 0.99 at Elvas and 0.99 at Vila Real) and across the three environments (0.89). These values are very close to the ones obtained in other studies with cowpea, which were always higher than 0.83 (Drabo et al., 1984; Omoigui et al., 2006; Manggoel et al., 2012; Egbadzor et al., 2013). These parameters with high heritabilities can be used in future breeding programs and for further quantitative genetic studies. However, it is important to refer that a high heritability alone is not enough to perform an efficient selection in advanced generations unless that it is accompanied by substantial genetic gains (Johnson et al., 1955; Mishra and Singh, 2014). In the three environments (Cartagena*Elvas*Vila Real), the lowest values of heritability were estimated in seed yield (0.29) and protein content (0.53) (Table 8). The days to maturity (0.79), first pod height (0.52) and protein content (0.36) were the parameters with lowest values of heritability in the Cartagena, Elvasand Vila Real, respectively. The low seed yield heritability was also reported by Omoigui et al. (2006) in cowpea. The value obtained in protein content is in agreement with the value reported by Ravelombola et al. (2016), who estimated a protein content of 0.58, and pointed out that this parameter can be inherited and can be selected for in the progeny.

Table 8. Heritability for the 10 morphological and agronomical parameters evaluated for UPCT, UTAD and INIAV and across the three environments in 12 cowpea accessions.

	Cartagena	Elvas	Vila Real	Cartagena*Elvas*Vila Real
Days to flowering	0.80	0.94	0.80	0.59
Days to maturity	0.79	0.84	0.77	0.60
Plant height (cm)	0.99	0.96	0.90	0.78
First pod height (cm)	0.98	0.52	0.92	0.77
Seed yield (g/m^2)	0.93	0.82	0.78	0.29
100 Seed Weight (g)	0.94	0.99	0.99	0.89
Pod length (cm)	0.99	0.97	0.93	0.81
Pod width (cm)	0.98	0.97	0.97	0.91
Number of seeds/pod	0.84	0.64	0.88	0.65
Protein content (%)	0.89	0.65	0.36	0.53

Principal component analysis (PCA) of the 12 cowpea accessions in three different environments in two seasons is presented in the Figure 6. The first two principal components explained 98.59% (PC1=59.75 and PC2=38.84) of total variation. In PC1, the main contributive parameter was yield (0.98) and in PC2 plant height (0.98) (Table 9). Manggoel (2011) and Doumbia (2013) also obtained in their studies that yield and plant height were parameters that contribute to the divergence between accessions. In addition, they found another parameter such as number of peduncles and flowers per plant, the days to flowering and days to maturity, which also contributed to the divergence, although some of them were not analyzed in the present study. The accessions characterized at Vila Real were grouped principally in the third quadrant, those characterized in Cartagena were mainly distributed in the second quadrant and the accessions at Elvas were dispersed for the four quadrants, although the majority were in first and fourth ones (Figure 6).

Table 9 Eigen value, factor scores and contribution of the first two principal component axis to variation in the 10 morphological and agronomical parameters of 12 cowpea accessions.

	Axis 1	Axis 2
Eigenvalues	9456.529	6147.711
Percentage	59.747	38.842
Cumulative percentage	59.747	98.589
Days to flowering	-0.034	0.001
Days to maturity	0.027	-0.022
Plant height	-0.201	0.979
First pod height	-0.009	0.018
Seed yield	0.978	0.202
100 Seed weight	0.008	-0.021
Pod length	-0.003	0.006
Pod width	0	0
Number of seeds/pod	0	0.003
Protein content	-0.005	0.002

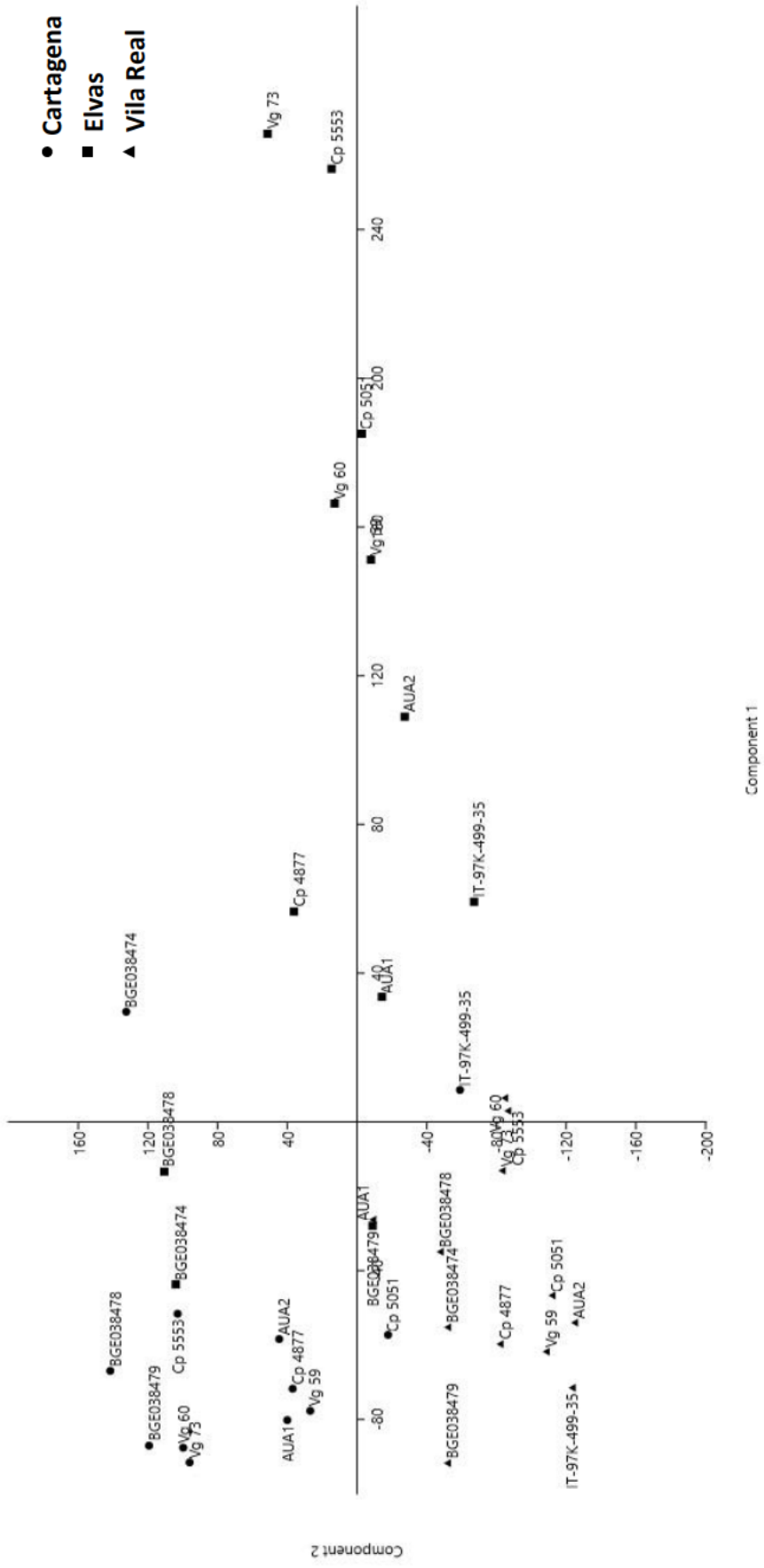
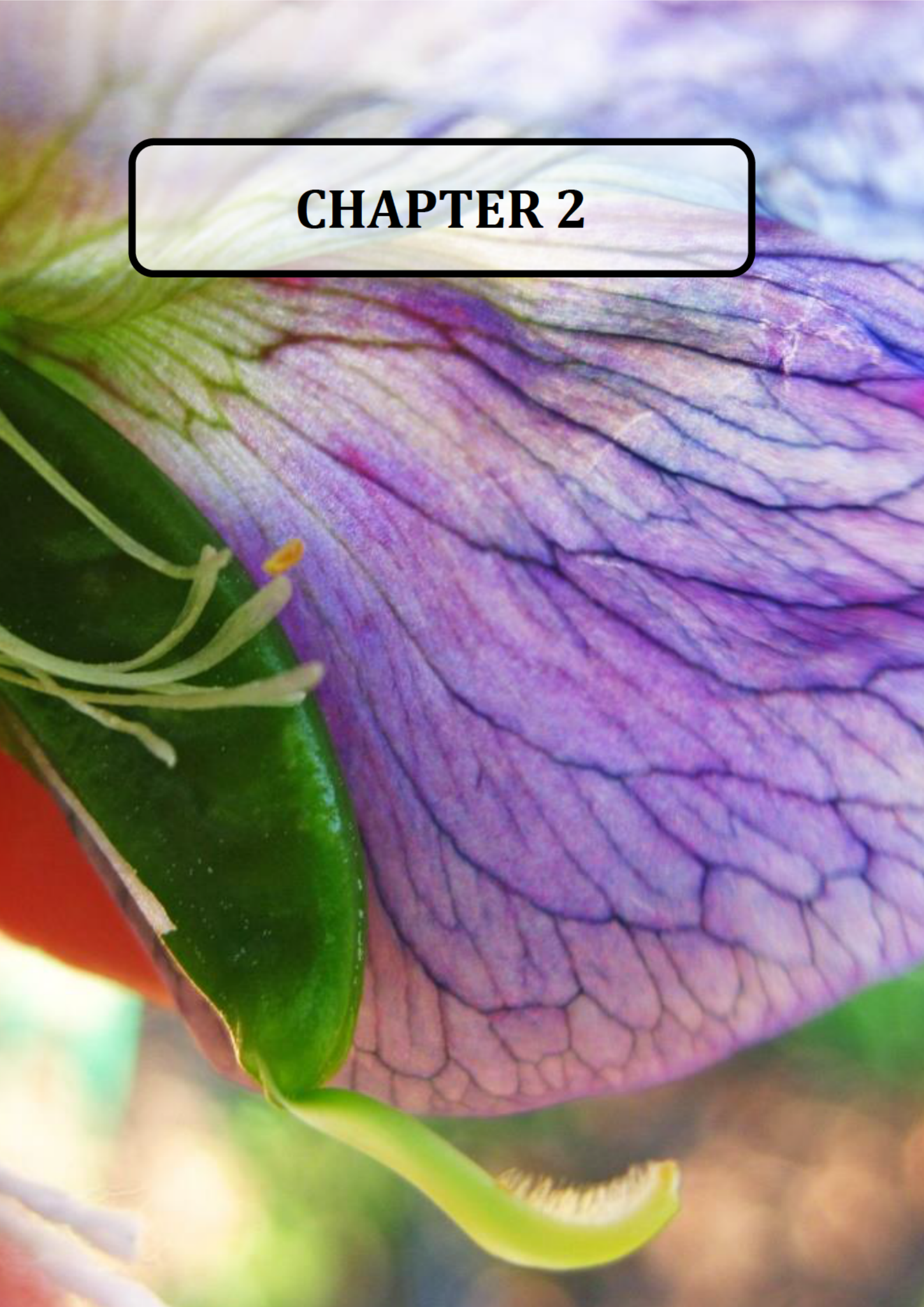


Figure 6. Principal component analysis of 12 cowpea accessions (average of 2 years) and the 3 environments based on 10 quantitative traits. The data are the mean of 2 years.

1.4. Acknowledgments

This study was supported by EUROLEGUME project. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement nº 613781. We would like to thank Timothy J. Close for seeds supply.



CHAPTER 2

CHAPTER 2 - A barcoding procedure for parallel genotyping of multiple accession using Next-Generation Sequencing

2.1. Introduction

Understanding the distribution of genetic variation and accession structure are important to manage genetic resources, but also to utilize proper germplasm in association mapping and breeding programs. Advances in molecular techniques have enabled the study of genetic variability at the DNA level, which has significantly increased the accuracy in assessing the genetic diversity and identifying cultivars. Germplasm analysis takes advantage of the advent of Next-Generation Sequencing (NGS), which has caused a major decrease in the price of sequencing and changed the way molecular markers can be obtained. The highest end is the one found in model organisms such as Arabidopsis or rice, where sequenced genomes allow the resequencing of thousands of accessions, creating a framework of Single Nucleotide Polymorphisms (SNPs), indels and copy number variants linked to traits (Li et al., 2014; Kawakatsu et al., 2016). This is the so-called Genotyping by Sequencing or GBS (Deschamps et al., 2012; He et al., 2014). SNP identification by sequencing plays an important role both for species identification and modern linkage mapping (Ganal et al., 2009) as SNPs present an important source of polymorphism among genotypes. The number of SNPs necessary for genotyping depends on a ranking of informative characters which include the minor allele frequency and lack of missing data across genotypes and may be as low as 7 SNPs as reported for genotyping hop varieties (Henning et al., 2015). Advances in sequencing technologies have led to the development of large collections of EST-derived SNPs for several legume species including faba bean (Kaur et al., 2014) and pea (Leonforte et al., 2013).

The economic aspect of NGS can become an issue when large germplasm sets have to be analysed. The current trend is to perform a very high number of reads in NGS projects. This is obviously fostered by the companies behind NGS technologies. However read coverage of as little as 2x-30x have been shown to be valid for heterozygosity

detection or rare alleles in large populations (Lynch, 2009; Li et al., 2011). Barcoding of sub-populations of samples and a multiplex sequencing application, the so-called "Bar-seq", may be a solution for efficient and affordable GBS (Smith et al., 2010). However, the performance of barcoded PCR amplicon sequencing may be hampered by biases like those introduced during the preparation of DNA, a failure of amplification depending on marker (SNP)/sample combination and a low efficiency during amplification which depends on the sequence composition (Mallona et al., 2011; Pawluczyk et al., 2015). As a result, the number of sequences representing specific SNP/sample combinations in the different amplicon libraries may vary with under- and over-representations of certain sequences. The novel approach of genotyping by sequencing still requires the development of routine protocols specially as they become increasingly used in non-model species (Nybom et al., 2014).

The role of legumes in the human diet and nutrition of animals and farming systems is increasingly important as they have a low environmental impact compared to other crops or protein rich sources (Tilman & Clark, 2014). They can be considered as multipurpose crops because they are used as vegetables, grains and forage. Furthermore, integration of legumes in crop rotation contributes to an improvement in soil fertility through nitrogen fixation and sustainability. Nowadays, soybean is globally the main protein source used (FAO, 2004). The volatile prices of soybean in Europe and the supply of non-genetically modified soybeans is diminishing and thus the related premiums are increasing. As a result, there is a strong interest in maximizing the use of locally produced grain legumes, like peas (*Pisum sativum* L.) or faba beans (*Vicia faba* L.) (Koivunen et al., 2016).

Constraints for the production of faba bean and the reason for its decrease during the last century in European countries are low yields, susceptibility to biotic and abiotic stresses, contents of non-nutritional factors as well as mechanization and therefore a reduced demand of bean as fodder plant together with a lack of breeding programs (Lazányi, 2002; Cubero, 2011). Concerning pea production, Europe accounts for 50-75% of world pea production. Dry pea ranks third to common bean and chickpea as the most widely grown pulse in the world with more than 11 million tons produced in 2013 (Tayeh, 2015). In the coming decades, a reduced expansion in dry pea production is expected due to competition with other crops that offer better returns

and due to a decrease in the demand of grain legumes as food protein supplier and its replacement by cereals and a reduced demand as animal feed.

The present study presents a novel protocol for the sequencing of a barcoded amplicon library, designed for the simultaneous identification of SNPs in independent samples on a single sequencing run. We were able to develop a minimal set of 40 SNPs for *P. sativum* and 45 for *V. faba* and a total of 24 accessions, 12 per species, that could be resolved for on single sequencing reactions.

2.2. Material and methods

2.2.1. Plant material and phenotyping

The accessions of *V. faba* and *P. sativum* selected for genotyping originated from Latvia, Estonia, Portugal and Spain. In total, 72 accessions were selected for each species based on local availability and usage (Supplementary Table 1). The accessions included cultivars and local varieties as well as genetic stocks. Spanish accessions were supplied by Universidad Politécnica de Cartagena (UPCT), Red Murciana de Semillas (RMS) and National Plant Genetic Resources Centre of the Spanish National Institute for Agriculture and Food Research and Technology (CRF-INIA). Portuguese accessions were provided by the National Institute for Agricultural and Veterinary Research (INIAV). Latvian accessions were provided by the State Priekuli Plant Breeding Institute (SPPBI) and Pure Horticultural Research Centre (PHRC) and by Estonian Crop Research Institute (ECRI) for the Estonian accessions. Seeds were germinated in a mixture of substrate/perlite/vermiculite (1:1:1) and leaves were harvested shortly after germination when sufficient plant material was available. Plants were grown in the field in random blocks 10 m2 of 15-20 plants, with four repetitions per accession.

A phenotyping of independent selected accessions was performed in Spain and Latvia (Supplementary Table 1). For *P. sativum*, 9 accessions were phenotyped in Spain and 19 in Latvia. For. *V. faba*, 8 accessions were phenotyped in Spain and 6 in Latvia. In Spain, accessions were grown under conventional field conditions at the ‘‘Tomás Ferro’’ Experimental Agro-Food Station of Universidad Politécnica de Cartagena, located in southeast of Murcia region, Campo de Cartagena. In Latvia, accessions were grown in

the organic field of Institute of Agricultural Resources and Economics at Priekuli Research Centre. Samples were taken from 10 randomly chosen plants from the different blocks. Pod length and pod width were evaluated in 10 typical dry pods to get an average. Plant height for each accession was determined at the end of the culture period. It was measured in cm from the ground to the top of the plant. Days to flowering were monitored from sowing date to the stage when 50% of plants had begun to flower in all the blocks. Average seed weight and pod yield were determined using the average pod yield in grams of all plants evaluated. Seed protein content was derived from the estimated Nitrogen (N) content, which was determined by the Kjeldahl method (Bremner, 1960) using the following formula: Protein content (%) = N content (%) × 6.25.

2.2.2. DNA extraction

Fresh leaf material was pooled from three young plants/variety. DNA samples were extracted using the commercial kit 'Plant NucleoSpin' (Machery and Nagel, Düren, Germany). All extracted samples were quantified with a Nanodrop 2000 and were diluted to 50 ng/µl in order to have identical concentrations, if necessary in combination with isopropanol-ethanol precipitation. PCR reactions were performed using 100 ng of genomic DNA.

2.2.3. SNP selection and primer design

SNP sequences were selected based on sequence information available for high-density genetic maps of *P. sativum* with nine linkage groups corresponding to fourteen chromosomes and *V. faba* with twelve linkage groups corresponding to twelve chromosomes (Kaur et al., 2014; Leonforte et al., 2013). SNP sequences representing the described linkage groups were chosen and 43 primer sets for *P. sativum* and 59 primer sets for *V. faba* were designed using the Custom Primers-Oligo Perfect™ Designer tool from Invitrogen for optimal primer size of 15 bases, a primer melting temperature (Tm) of 50°C and a PCR product size around 100 bp (Supplementary Table 2). The forward SNP primers were designed with a common starting sequence "P1-CCTCTCTATGGGCAGTCGGTGAT" (Ion Amplicon Library Preparation (Fusion Method) Life

Technologies) and the reverse SNP primers started with a common sequence "B – CAGTAACAGCGGACGTGCTCTGAT" (Figure 7 PCRI). These primers were applied in a first round of PCR. For the second round of PCR, we used forward primer "P1" and, as reverse primer, 12 special barcoded primers consisting of adaptor "A-CCATCTCATCCCTGCGTGTCTCCGAC" sequence plus a specific barcode of 5 bases and partial sequence "B- CAGTAACAGCGGACGTGCTC" (Figure 7 PCRII). These 12 primers were used for the amplification of all varieties in pools of 12. The sequences "A" and "P1" are universal primer sequences for the Ion Amplicon Library Preparation.

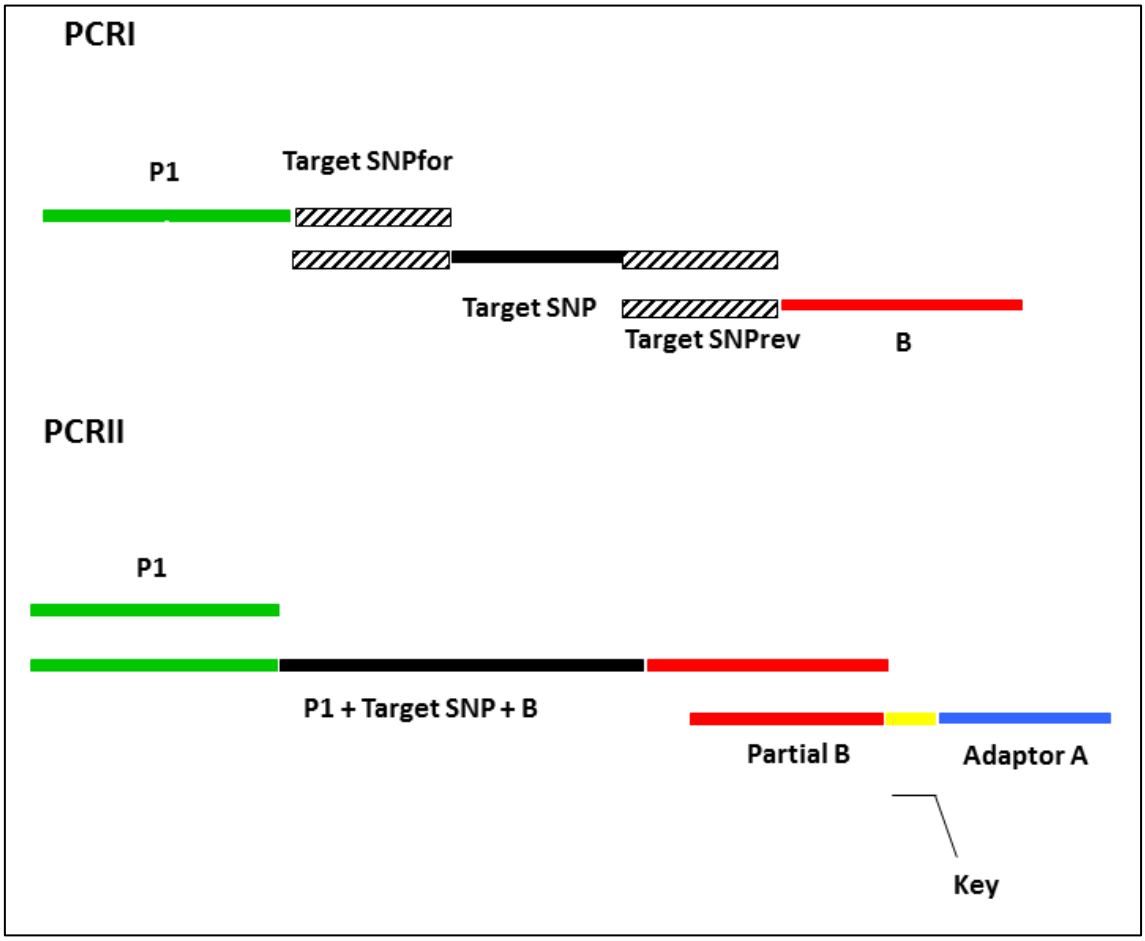


Figure 7. Schematic presentation of conventional PCR steps in amplicon library preparation. Primers P1 and Adaptor A are adapted from Ion Amplicon Library Preparation (Fusion Method), Life Technologies.

2.2.4. Amplicon library generation

PCR I and II were performed using the OneTaq ® DNA polymerase Kit (New England Biolabs) according to the manufacturer´s instruction. For PCR I, 100 ng (2 µl) of

genomic DNA were amplified separately with each SNP primer pair. PCR I conditions were as follows: 10 cycles of denaturation at 94°C for 10 sec., annealing at 50°C for 15 sec., extension at 68°C for 20 sec. and 30 cycles of denaturation at 94°C for 10 sec., annealing at 60°C for 15 sec., extension at 68°C for 20 sec. The PCR I products of all SNPs were pooled for each accession and 1 µl aliquots were used for PCR II, always 12 varieties at a time applying the 12 different barcode primers. PCR II conditions were as followed: 10 cycles of denaturation at 94°C for 30 sec., annealing at 60°C for 30 sec., extension at 68°C for 30 seconds and 30 cycles of denaturation at 94°C for 30 sec., annealing at 65°C for 30 sec., extension at 68°C for 30 sec.

2.2.5. Emulsion PCR and sequencing

Concentration of PCR II products of all 12 pools from *P. sativum* and *V. faba* was determined with Nanodrop 2000 after PCR cleanup kit (MinElute PCR purification kit, Qiagen). Aliquots of equal concentrations of these pools were joined. Concentration and average size of the amplicon library were used to calculate the amount of DNA fragments per micro litre and libraries were diluted to 26 pM (Ion Amplicon Library Preparation (Fusion Method, Life Technologies, CA, USA) (Jünemann et al., 2012). The adjusted library was amplified by emulsion PCR prior to sequencing, using the OneTouch system with the Ion XPressTemplate kit V2.0 (Life Technologies) as described by the manufacturer in the user Guide (Part No. 4469004 Rev. B 07/2011). Qubit 2.0 instrument using the Qubit Test kit assay (Ion Sphere TM Quality Control Kit Part Number 44686569 (Life Technologies) was used to analyse the quality and quantity of the enriched spheres as described in the appendix of the Ion Xpress Template Kit User Guide (Part Number 4467389 Rev. B, 05/2011, Life Technologies). Sequencing of the amplicon libraries was performed using the Ion Sequencing 200 kit V2 (Part Number 4482006, Life Technologies) following the corresponding protocol according to the manufacturer's instructions and sequencing was performed using Ion 314TM Chip Kit V2 (Part Number 4488144, Life Technologies). A total of 6 sequencing reactions were performed, corresponding to 12 varieties each from *P. sativum* and *V. faba* in each sequencing reaction.

2.2.6. Sequence data analysis and phylogenetic tree construction

Sequences belonging to specific SNPs and varieties were identified from the amplicon libraries based on the SNP and the 5 bp variety barcode sequence. Within these sequences, the allelic SNP variants described by Leonforte et al. (2013) and Kaur et al. (2014) were determined and used for phylogenetic tree construction based on a matrix Using 0 for no amplification, 1 and 2 for the allelic variant and 4 in cases where the two allelic variants were detected for a variety either due to heterozygote plants or mixture of homozygotes for the two alleles (heterogeneous population). Distances were calculated using the UPGMAdendro programme (http://genomes.urv.es/UPGMA/index.php) (Garcia-Vallvé, Palau, & Romeu, 1999). Distances were calculated with Pearson correlation distance using 100 bootstraps. The corresponding Newick formatted dendrogram was rendered with MEGA7 (Kumar, Stecher, & Tamura, 2016).

2.2.7. Association mapping

The software program TASSEL 5.0 was applied to evaluate associations between SNPs and phenotypic traits (Bradbury et al., 2007). A general linear model (GLM) was used for calculating the associations between the marker and the trait. Population structure was scored using PCA analysis and included in GLM analysis. A significance threshold for association was set at p=0.05.

2.3. Results and discussion

2.3.1. SNP detection in barcoded amplicon library sequencing

We applied a genotyping protocol based on an internal barcoding procedure for the simultaneous analysis of multiple varieties of *P. sativum* and *V. faba* in a single NGS run on one chip. Species, variety and SNP were then identified based on the specific combination of SNP sequence and barcode sequence. In total, we used 115 primers which included 43 SNP specific primers for *P. sativum* and 59 for *V. faba*, as well as one

generic P1 adaptor primer and 12 barcoded A adaptor primers (Figure 7), specified by the Ion Torrent One touch sequencing protocol. Although aliquots of equal concentration of *P. sativum* and *V. faba* were used in all pools, the total number of sequence reads differed among sequencing reactions (Table 10). The number of SNPs without reads was 14 for *V. faba* and 3 for *P. sativum*. At the other side of the spectrum, 5 SNPS in *V. faba* and 11 in *P. sativum* accumulated on over 500 reads on average of all pools (Supplementary Figure 1). The total number of valuable reads did not correlate significantly with the total number of reads representing the amplified library (r^2=0.27434). The percentage of reads from all libraries containing the complete reverse primer sequence from PCR I was similar for *P. sativum* and *V. faba* with 16.7 and 16.1%, respectively. However, this percentage varied among pools. From all reads containing the complete reverse primer of PCR I, 22.6% contained the complete sequence up to the SNP in case of *P. sativum* and 17.2% in case of *V. faba.* The remaining reads, although specific for the reverse primer, mostly consisted of non-SNP specific sequences or they were too short and did not reach the particular allelic SNP site. The mean read length of the sequenced libraries was 68 bp while the average amplicon size till SNP site was 90 bp with a range of 53 to 129 bp in case of *P. sativum* and 98 bp with a range of 49 to 130 in case of *V. faba* (Table 10). However, we did not find a significant correlation between the amplicon size till SNP site and the read number of particular SNPs (r^2=0.1102 for *P. sativum* and 0.1443 for *V. faba)*, indicating that read size was not severely biased.

Table 10. Information on sequence quality and read number.

Pool	(A) Total reads /pool	(B) Reads with common sequence B plus reverse primer		(C) Reads including SNP		Mean read length (bp)
		Pisum sativum	*Vicia faba*	*Pisum sativum*	*Vicia faba*	
1-12	253,269	21,002 (8.3% *)	1,7548 (6.9%*)	2,977 (14.2% **)	1,981 (11.3%**)	52
13-24	457,362	55,542 (12.1%*)	83,672 (18.3%*)	10,695 (19.3%**)	13,905 (16.6%**)	67
25-36	391,746	48,940 (12.5%*)	42,846 (10.94%*)	8,495 (17.4%**)	10,189 (23.8%**)	59
37-48	398,961	69,206 (17.35%*)	66,725 (16.72%*)	19,815 (28.6%**)	11,079 (16.6%**)	69
49-60	222,800	73,909 (33.17%*)	54,249 (24.35%*)	23,718 (32.1%**)	7,545 (13.9%**)	84
61-72	241,144	60,050 (24.9%*)	52,217 (21.65%*)	14,361 (23.9%**)	12,905 (24.7%**)	78
	Total: 19,65,282	Total: 328,649 (16.7%*)	Total: 317,217 (16.1%*)	Total: 80,061 (22.6%**)	Total: 57,603 (17.2%**)	Mean: 68

*=percentage of (A); ** = percentage of (B)

The aim of the current work was to develop a simple, yet robust protocol for genotyping of multiple varieties and/or species in single NGS runs. Two aspects were

taken into account. First the number of primers required to do the libraries and sequencing procedures, and second the number of sequencing reactions that would give enough data to be able to obtain meaningful results. In order to do so we adopted an internal barcoding procedure. Our results show that our primer design allows simultaneous sequencing of all SNPs for 12 varieties of two species, *P. sativum* and *V. faba*, while using a minimum number of primers. Species, variety and SNP are then identified based on the specific combination of SNP sequence and barcode sequence and this analysis is used to gather information on sequence quality and read number of the bar-seq amplicon libraries for the six pools. In principle, we wanted to obtain a minimum number of reads per SNP that would be over 30. The theoretical minimum number of reads of 30x would require 92880 reads for *P. sativum* and 127440 for *V. faba,* even so the reported valid range for the detection of rare alleles or heterozygosity lies between 2x-30x (Y. Li et al., 2011; Lynch, 2009). The actual number of primer specific reads containing the informative SNP site was 26x (80061 reads) for *P. sativum* and 14x (57603 reads) for *V. faba*, falling within the acceptable range of reads/SNP. The detailed analysis of sequence quality and read number showed that from all sequence reads, a similar percentage for *P. sativum* and *V. faba* contained the complete reverse primer sequence from PCR I, indicating that amplicon library qualities were similar for both species. On the other hand, two types of biases during pool sequencing were seen. Firstly, the percentage of reads containing the complete reverse primer sequence from PCR I varied among pools, indicating a bias during amplicon library preparation. Secondly, we found an important bias in the number of reads per SNP (Supplementary Figure 1). While 14 SNPs for *V. faba* and 3 for *P. sativum* were not represented among reads, 5 SNPS in *V. faba* and 11 in *P. sativum* accumulated over 500 reads. As the number of reads does not follow a distribution with a peak in the middle range, but a valley, we can conclude that some SNPs are amplified better than others, despite having the same adaptors. It was shown previously that failure of amplification may depend on specific marker (SNP)/sample combination, which may result in no amplification or a low efficiency during amplification (Mallona et al., 2011; Pawluczyk et al., 2015), suggesting that the actual amplicon structure may play a key role in its success as a marker.

2.3.2. Characterisation of SNP markers in populations of V. faba and P. sativum accessions

Results on SNP identification for 72 genotypes are listed in Table 11. Forty out of 43 SNPs successfully amplified in case of *P. sativum* and 45 out of 59 in case of *V. faba*, even so not across all genotypes. One genotype in case of *V. faba* showed more than 80% missing data for all SNP amplifications, probably due to low DNA quality of this particular genotype. Among the amplified SNPs, two resulted monomorphic within the analysed population of accessions both for *P. sativum* and *V. faba*. Ten out of 72 genotypes in case of *P. sativum* and 14 out of 72 genotypes in case of *V. faba* had more that 20% of SNPs in heterozygosis. The percentage of polymorphic SNPs (Minor allele frequency (MAF) > 0.01 = 1%) was 93% for *P. sativum* and 92% for *V. faba* (Table 12 and Table 13). In both species, the MAF ranged from 0.8 to 50% with a mean MAF of 20% for *V. faba* accessions and 17% for *P. sativum* accessions. The selected SNPs did not include rare variants (MAF < 0.5%) for the given populations, while 26% (*V. faba*) and 18% (*P. sativum*) were low-frequency (0.5% ≤ MAF < 5%) variants.

Table 11. Information on SNP detection for *Pisum sativum* and *Vicia faba*.

	Pisum sativum	*Vicia faba*
Number of amplified SNPs	40/43	45/59
Genotypes with >20% heterozygous SNPs	10	14
SNPs with >80% missing data	8	16
Genotypes with >80% missing data	0	1

The SNPs selected from the literature for genotyping faba bean and pea, were polymorphic and bi-allelic in RIL–mapping populations between ascochyta blight resistant and susceptible cultivars of *V. faba* (Kaur et al., 2014) and salinity sensitive and moderately tolerant cultivars of *P. sativum* (Leonforte et al., 2013). In case of the plant material used here, 93% of all SNPs amplified in case of *P. sativum* and 76% in case of *V. faba*, even so not across all genotypes. On the other hand, only one genotype in case of *V. faba* showed more than 80% missing data for all SNP amplifications, indicating that failure of amplification cannot be explained by DNA quality but rather by specific primer–DNA combinations. Sequence polymorphisms in the selected primer regions

among our genotypes and those originally used for SNP identification may be the cause of amplification failure.

Table 12. Information of 42 SNPs amplified in *Vicia faba*.

72 accessions of *Vicia faba*					
Amplified SNP number	**SNP id[a]**	**Chr[a]**	**SNP type**	**Type of change**	**MAF[b]**
2	SNP_50001182	I.B/VI	A/T	Transversion	0.04
4	SNP_50002306	I.B/VI	A/G	Transition	0.18
5	SNP_50002318	I.B/VI	A/T	Transversion	0.13
6	SNP_50000197	V.1	C/T	Transition	0.50
7	SNP_50000487	V.1	C/G	Transversion	0.14
8	SNP_50000557	V.1	C/T	Transition	0.05
10	SNP_50000760	V.1	C/G	Transversion	0.11
12	SNP_50000432	VI	C/T	Transition	0.50
13	SNP_50001365	VI	C/T	Transition	0.14
14	SNP_50001987	VI	A/T	Transversion	0.16
17	SNP_50001872	I.A/V	A/T	Transversion	0.15
18	SNP_50001826	I.A/V	G/T	Transversion	0.07
19	SNP_50001916	I.A/V	C/T	Transition	0.23
21	SNP_50000057	V.2	A/C	Transversion	0.05
24	SNP_50000911	V.2	A/T	Transversion	0.02
25	SNP_50001679	V.2	C/T	Transition	0.21
26	SNP_50002062	I.A	A/G	Transition	0.03
27	SNP_50000965	I.A	A/C	Transversion	0.01
28	SNP_50001040	I.A	A/T	Transversion	0.05
29	SNP_50001146	I.A	C/T	Transition	0.22
30	SNP_50000089	I.A	A/T	Transversion	0.25
31	SNP_50000217	II	C/T	Transition	0.17
32	SNP_50000310	II	A/G	Transition	0.11
33	SNP_50000356	II	A/G	Transition	0.01
34	SNP_50000332	II	C/T	Transition	0.18
35	SNP_50000787	II	C/T	Transition	0.21
36	SNP_50000084	I.A/III/V	T/C	Transition	0.02
37	SNP_50000308	I.A/III/V	C/T	Transition	0.11
38	SNP_50000440	I.A/III/V	C/T	Transition	0.14
39	SNP_50001023	I.A/III/V	C/T	Transition	0.50
40	SNP_50002450	I.A/III/V	C/T	Transition	0.15
41	SNP_50000764	III.2	C/G	Transversion	0.17
42	SNP_50001532	III.2	C/T	Transition	0.01
43	SNP_50001531	III.2	A/T	Transversion	0.13
44	SNP_50001647	III.2	A/G	Transition	0.23
47	SNP_50000203	III.3	A/T	Transversion	0.16
49	SNP_50000824	III.3	A/G	Transition	0.13
50	SNP_50001586	III.3	A/G	Transition	0.06
51	SNP_50000208	I.B.2	C/T	Transition	0.09
53	SNP_50002378	I.B.2	A/G	Transition	0.50
54	SNP_50001498	I.B.2	C/T	Transition	0.01
55	SNP_50002155	I.B.2	A/G	Transition	0.50

[a]SNP id and chromosome position (Chr) according to Kaur et al., 2014. [b]MAF, Minor allele frequency.

Table 13. Information about 33 SNPs amplified in *Pisum sativum*.

72 accessions of *Pisum sativum*					
Amplified SNP number	**SNP id[a]**	**Chr[a]**	**SNP type**	**Type of change**	**MAF[b]**
1	SNP_100000443	Ps I (Mt 5)	T/G	Transversion	0.09
2	SNP_100000341	Ps I (Mt 5)	T/G	Transversion	0.12
3	SNP_100000597	Ps I (Mt 5)	A/C	Transversion	0.17
4	SNP_100000346	Ps I (Mt 5)	C/T	Transition	0.10
5	SNP_100000414	Ps I (Mt 5)	A/G	Transition	0.15
7	SNP_100000648	Ps II (Mt 1)	T/C	Transition	0.02
8	SNP_100000616	Ps II (Mt 1)	A/C	Transversion	0.01
9	SNP_100000543	Ps II (Mt 1)	A/C	Transversion	0.15
10	SNP_100000664	Ps II (Mt 1)	A/G	Transition	0.24
11	SNP_100000360	Ps III (Mt 2/3)	A/G	Transition	0.16
13	SNP_100000091	Ps III (Mt 2/3)	A/G	Transition	0.22
14	SNP_100000095	Ps III (Mt 2/3)	C/T	Transition	0.13
15	SNP_100000288	Ps III (Mt 2/3)	A/G	Transition	0.15
16	SNP_100000395	Ps IV (Mt 4/8)	A/G	Transition	0.20
17	SNP_100000035	Ps IV (Mt 4/8)	A/T	Transversion	0.11
18	SNP_100000786	Ps IV (Mt 4/8)	A/G	Transition	0.22
19	SNP_100000282	Ps IV (Mt 4/8)	C/T	Transition	0.14
20	SNP_100000071	Ps IV (Mt 4/8)	C/G	Transversion	0.02
21	SNP_100000171	Ps V (Mt 7)	A/G	Transition	0.10
22	SNP_100000408	Ps V (Mt 7)	C/T	Transition	0.14
23	SNP_100000264	Ps V (Mt 7)	G/A	Transition	0.01
24	SNP_100000822	Ps V (Mt 7)	A/T	Transversion	0.10
25	SNP_100000042	Ps V (Mt 7)	C/T	Transition	0.50
26	SNP_100000485	Ps V.2 (Mt 7)	A/T	Transversion	0.23
27	SNP_100000746	Ps V.2 (Mt 7)	C/T	Transition	0.06
29	SNP_100000736	Ps V.2 (Mt 7)	C/G	Transversion	0.23
32	SNP_100000350	Ps VI (Mt 2/6)	C/T	Transition	0.06
35	SNP_100000702	Ps VI (Mt 2/6)	C/T	Transition	0.08
37	SNP_100000655	Ps VI.2 (Mt 2)	C/T	Transition	0.22
38	SNP_100000058	Ps VI.2 (Mt 2)	A/G	Transition	0.06
39	SNP_100000328	Ps VII (Mt 4/8)	A/T	Transversion	0.08
41	SNP_100000255	Ps VII (Mt 4/8)	C/G	Transversion	0.17
43	SNP_100000045	Ps VII (Mt 4/8)	C/T	Transition	0.11

[a]SNP id and chromosome position (Chr) according to Leonforte et al., 2013. [b]MAF, Minor allele frequency.

Fourteen and 19% of all genotypes in case of *P. sativum* and *V. faba*, respectively, had more that 20% of SNPs in heterozygosis. *V. faba* is a facultative allogamous species, that means primarily outcrossing (Kaur et al., 2014) while *P. sativum* is usually regarded as a self-pollinator with occasional cross-pollination (Bogdanova & Berdnikov, 2000). Despite the different pollination biology, there was no significant difference in terms of degree of heterozygosity (Chi-test *p*-Value=0.2295). This is consistent with a lack of population structure observed in a survey of *V*. *faba* accessions from the Western Mediterranean Basin, indicating a high degree of gene flow (Oliveira et al., 2016). These

results indicate that at least part of the accessions are not pure lines but rather heterozygous or mixed landraces.

Genetic association studies typically apply a minor allele frequency (MAF) threshold of 1-2% for genetic association studies (Anderson et al., 2010) and as a convention, ideal SNPs are those with MAFs>15-20%, which applies to 43% (*P. sativum*) and 52% of all SNPs (*V. faba*). The percentage of polymorphic SNPs (Minor allele frequency (MAF) > 0.01=1%) was 93% for *P. sativum* and 92% for *V. faba.* The selected SNPs did not include rare variants (MAF < 0.5%) for the given populations and even so 26% (*V. faba*) and 18% (*P. sativum*) were low-frequency (0.5% ≤ MAF < 5%) variants (Lee, et al., 2014), our data indicate that the selected SNPs can be applied in association studies.

2.3.3. Resolution of SNP based genotyping

A final number of 40 SNP markers of *P. sativum* were used for UPGMA tree construction. Each accession had a unique SNP fingerprint. Accessions for North-Eastern Europe (Latvia, Estonia) and South-Western Europe (Spain and Portugal) were interspersed with the exception of a cluster of 11 and a smaller sub-cluster of 7 accessions exclusively from Latvia. Accessions from Estonia did not cluster preferentially with those from Latvia, as might be expected from their geographic vicinity (Figure 8). Among the 30 Spanish accessions included in the dendrogram, 7 genotypes formed a two-leaved clade with other Spanish accessions, indicating a high level of similarity, while three and two Spanish accession formed a two-leaved clade with genotypes from Latvia and Estonia, respectively, and none from Portugal. Among the 33 Latvian accessions, 9 genotypes formed two-leaved clades with another accession from Latvia, one from Portugal and two from Estonia. A pairwise genetic similarity (GS) matrix was calculated using Poisson´s coefficient. The highest GS of 0.854 was found between two Spanish accessions. The average GS was 0.369 among Spanish accessions, 0.401 among Latvian accessions, 0.491 among Portuguese accessions and 0.395 among Estonian accessions. The similarity between accessions from Spain was significantly lower compared to Latvia (p=0.0036).

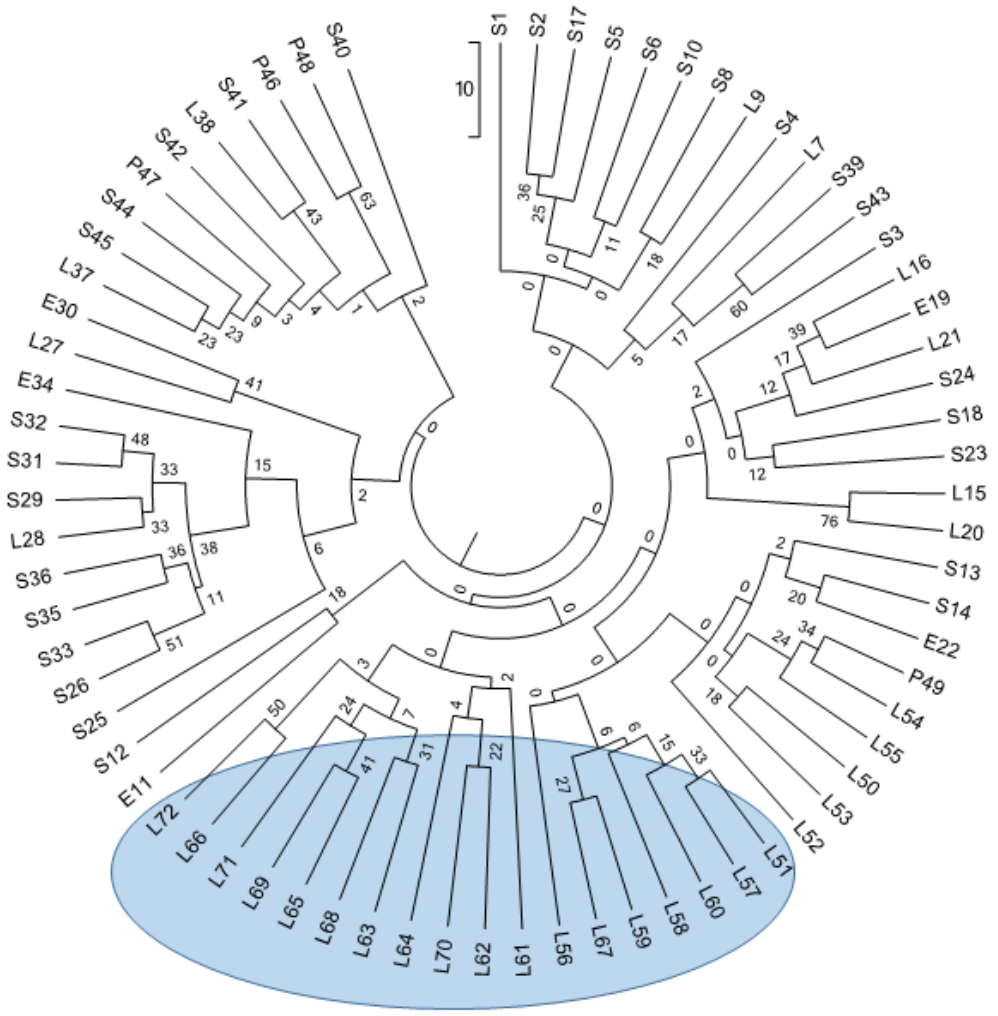


Figure 8. Neighbour-joining dendrogram (100 botstraps) for 72 *P. sativum* accession. Capital letters indicate place of origin (E=Estonia; L= Letonia; P=Portugal; S=Spain).

In case of *V. faba,* we used a final number of 44 SNP markers for UPGMA tree construction. 34 out of the 42 analysed Spanish genotypes formed two separate clusters comprising 23 and 11 Spanish accession, while accessions from other countries were interspersed (Figure 9). Forty two percent of Spanish accessions formed two-leaved clades of highest similarity with other Spanish accessions and only one Spanish variety clustered in a clade with a variety from Latvia. Surprisingly, 3 of the 15 Latvian accessions showed highest similarity with Portuguese accessions. Calculation of pairwise GS using Poisson´s coefficient showed that the highest GS of 0.902 was found between two Spanish accessions. The average GS was 0.321 among Spanish accessions, 0.369 among Latvian accessions, 0.486 among Portuguese accessions and 0.449 among Estonian accessions. The similarity between accessions from Spain was significantly lower

compared to Latvia (p=0.025).

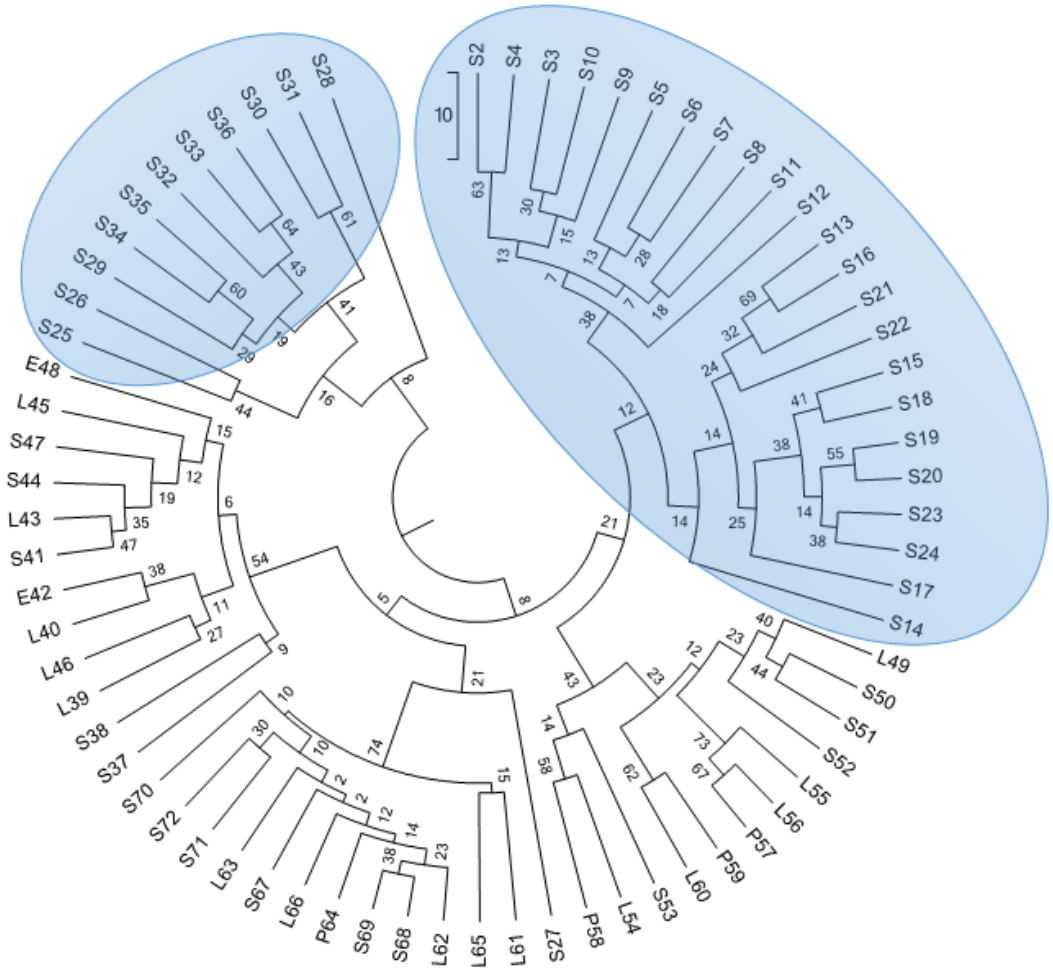


Figure 9. Neighbour-joining dendrogram (100 botstraps) for 72 *V. faba* accession. Capital letters indicate place of origin (E=Estonia; L= Letonia; P=Portugal; S=Spain).

UPGMA tree construction and cluster analysis for 72 *P. sativum* accessions from North-Eastern Europe (Latvia, Estonia) and South-Western Europe (Spain and Portugal) based on 40 SNP markers showed, that that even so pairs of most similar genotypes often refer to the same country, our pea genotypes classify into groups that are not consistent with their geographical origin, except for the two clusters of Latvian accessions, while accessions from south-western Europe did not form separate clusters. An analysis of the worldwide *P. sativum* population structure using microsatellite markers showed, with some exceptions, consistent and separate clusters of accessions from Canada, USA, and Europe and West Asian Origin. An additional cluster was more cosmopolitan (Ahmad et al., 2015) and authors state that this is possibly due to free

exchange and wide trading of pea germplasm for several decades, which may also explain our observation of interspersed clustering among southern and eastern European accession.

Similar to *P. sativum*, the 72 *V. faba* accessions formed pairs of most similar genotypes that referred to the same country, but clustering was not always consistent with the geographical origin. However, and opposite to *P. sativum,* we observed two separate clusters for Spanish accession, while those from Latvia did not form a cluster. An analysis of genetic diversity among 53 *V. faba* accessions from the Western Mediterranean basin showed a lack of population structure (Oliveira et al., 2016). The authors proposed that this may be partly due to frequent seed exchange and introduction, which may also explain the clustering across geographical barriers observed here.

The analysis of pairwise GS values indicated that the highest degree of similarity was found between two Spanish accessions of *V. faba*. However, on average, *P. sativum* accessions of specific countries showed higher similarities than *V. faba* accessions, even so this difference in similarity was significant only in case of the Spanish accessions (p=0.0002). This can be possible explained by a higher level of seed transfer in case of *V. faba.*

2.3.4. Association mapping

A threshold of p=0.05 was utilized to identify candidate SNP markers associated to phenotypic traits using a generalized linear model. As two extremely different environments were used for phenotypic data acquisition, the mapping strategy was applied isolated for Spain and Latvia (Table 14). No significant associations were observed for *P. sativum* in Spain and *V. faba* in Latvia, probably due to a low number of accessions included in the analysis. We found SNP association to the traits: days to flowering, pod yield and protein content for *P. sativum* grown in Latvia and to the traits; days to flowering, pod yield and protein content as well as pod length and plant height for *V. faba* grown in Spain.

Association mapping allows the identification of genetic markers linked to phenotypic traits, which may be polygenic. In case of the Latvian accessions of *P.*

sativum, phenotyped under the specific geographical conditions of Latvia, we found SNP association to the polygenic traits: days to flowering, pod yield and protein content. For the Spanish accession of *V. faba* grown in Spain, we could add the traits: pod length and plant height. For both populations and in case of some traits we identified more than one trait associated to one SNP (e.g. SNP_100000543 and SNP_100000255 to protein content). On the other hand, one of the identified SNPs of *V. faba*, SNP_50001040, was associated simultaneously to two traits, protein content and days to flowering. Our findings are an agreement with the observation that multiple loci confer to one quantitative trait and that QTLs conferring multiple agronomic traits with pleiotropic effects may be organized in form of clusters around SNP marker regions (X. Hu et al., 2015). The identified associations may be valuable as molecular markers for future research programmes on the improvement of the aforementioned traits.

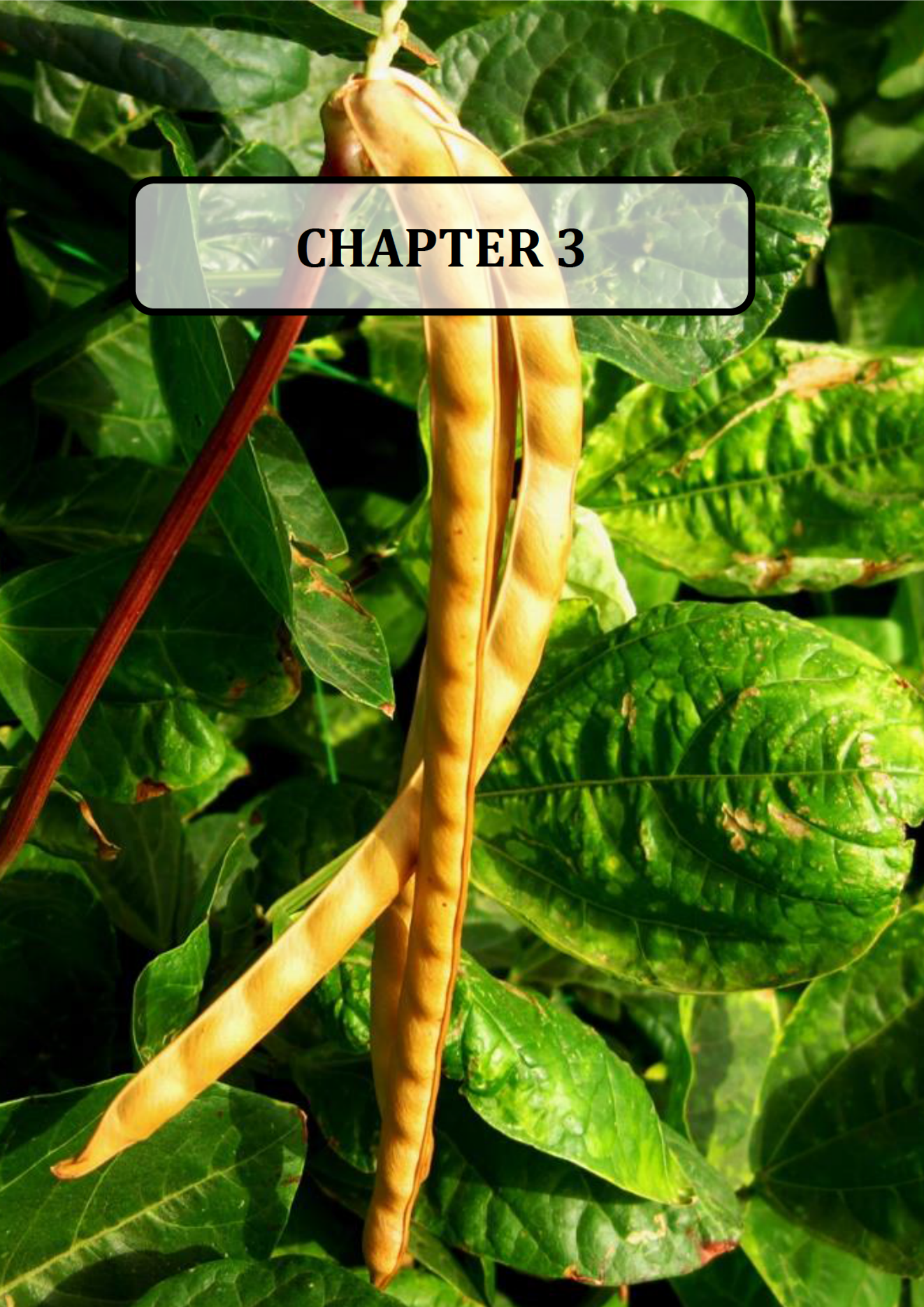
Table 14. Association mapping of phenotypic traits from *Pisum sativum* and *Vicia faba* accessions evaluated in Spain and Latvia.

Pisum sativum					
Site of evaluation	**Number and origin of accession**	**SNP**	**Chr.**	**Trait**	***p*-Value**
Spain	**9 from Spain**	No significant associations found			
Latvia	**19 from Latvia**	SNP_100000350	PsV.2	Days to flowering	0.017
		SNP_100000543	PsII	Protein content	0.027
		SNP_100000255	PsVI	Protein content	0.035
		SNP_100000736	PsV	Pod yield	0.036
Vicia faba					
Site of evaluation	**Number and origin of accession**	**SNP**	**Chr.**	**Trait**	***p*-Value**
Spain	**8 from Spain**	SNP_50001040	I.A	Protein content	0.010
		SNP_50000965	I.A	Pod yield	0.010
		SNP_50000911	V.2	Plant height	0.014
		SNP_50001365	VI	Pod length	0.028
		SNP_50001040	I.A	Days to flowering	0.030
		SNP_50000089	I.A	Plant height	0.034
		SNP_50001872	I.A/V	Pod length	0.034
		SNP_50001679	V.2	Plant height	0.039
		SNP_50000764	III.2	Pod yield	0.044
Latvia	**6 from Latvia**	No significant associations found			

2.4. Acknowledgments

This work was performed under the Eurolegume project. This project has received funding from the European Union's Seventh Framework Programme for

research, technological development and demonstration under grant agreement nº 613781.



CHAPTER 3

CHAPTER 3 - Diel pattern of circadian clock and storage protein gene expression during seed filling in cowpea (*Vigna unguiculata* L. Walp)

3.1. Introduction

Cowpea (*Vigna unguiculata*) is an important food source of African origin and was introduced to the Indian subcontinent approximately 2000-3500 years ago (Allen, 1983). It is used at all stages of growth from the green leaves, which are used like spinach, immature pods, green cowpeas or dry mature seeds (Davis et al., 1991). Like other legumes and cereals, cowpea forms an important source of protein supply for human nutrition (Gupta et al., 2010). The accumulation of seed storage proteins in legumes occurs during the seed filling phases until desiccation. On average, legume seeds contain 17-30 per cent protein (Sathe et al., 1984). In a survey of Spanish landrace accessions, protein content in cowpea varied between 14.8 to 23.6%, depending on the genotype (Martos-Fuentes et al., 2015). Globulins constitute with 55-58%, the major seed proteins in cowpea followed by albumins, basic glutelins, acid glutelins and prolamins (Gupta et al., 2010; Vasconcelos et al., 2010). This compositional characteristic was stable in a nutritional survey of seed protein types in high-yielding cowpea cultivars (Vasconcelos et al., 2010).

Among the globulins, legumes mainly accumulate 7S vicilin-type globulins and 11S legumin-type globulins. The proportions of legumin (LEG) and vicilin (VIC) are genetically and environmentally determined in pea seeds and they are synthesized from at least 40 genes and at least 10 different genetic loci (Casey et al., 2001). Legumins and vicilins share sequence identity both at the amino acid and nucleotide level, which hints to a common ancestor for these two storage proteins (Sales, 2000). The legumin genes are arranged in two clusters for *LEGA* and *LEGB* (De Pace et al., 1991). Other genes codifying for minor B-type legumin polypeptides are *LEGJ* and *LEGK* in *Pisum sativum* L. (Gatehouse et al., 1982). A time series study of transcript profiles based on a *Lotus japonicus* gene expression atlas identified genes for *VIC*, *CVC* and *LEG* among the ten

most highly expressed genes during legume seed maturation. Storage protein genes for *LEG* and *VIC* also formed part of pod-enhanced transcripts (Verdier et al., 2013).

Many plant biological activities show diurnal variation and the circadian clock acts as endogenous timer, coordinating and entraining plant activities in response to environmental cues like light and temperature (McClung CR, 2006). Biological activities under circadian control include those related to seasonal development such as flowering time, productivity, tuberization, and dormancy. Other biological processes controlled by the circadian clock are adaptation to cold or drought (Marcolino-Gomes et al., 2014), pathogen resistance, stomatal movement, and scent production (Egea-Cortines et al., 2013). Genes related to primary metabolism, including RNA, proteins and carbohydrates, are expressed cyclically (Farré & Weise, 2012). In Arabidopsis, 6% to 8% of all the open reading frames could be circadian regulated (Harmer et al., 2000; Schaffer et al., 2001) and circadian regulation under light cycling involves 23% of the annotated genes in maize (Hayes et al., 2010) and 30–40% in rice and poplar (Filichkin et al., 2011). The patterns of transcription under circadian regulation may show distinct phasing, i.e. protein synthesis and cell cycle related processes have peaks between midnight and dawn, while those related to energy metabolism peak after dawn (Michael et al., 2008).

The circadian oscillator is well understood in *Arabidopsis thaliana* and contains a central circadian oscillator complex formed by the genes *CIRCADIAN CLOCK ASSOCIATED1/LATE ELONGATED HYPOCOTYL (CCA1/LHY).* These MYB genes act together with *TIMING OF CAB EXPRESSION1 (TOC1),* a PSEUDO-RESPONSE REGULATOR in a feedback regulating system, controlling its oscillation reciprocally (de Montaigu, Toth, Coupland, & Tóth, 2010; Locke et al., 2006). Interconnected with this core midday loop are a morning loop and an evening loop. The morning loop consists of three *PSEUDO RESPONSE REGULATOR (PRR)* genes *PRR9*, *PRR7* and *PRR5*. The protein products form a complex and inhibit the midday loop genes *CCA1* and *LHY* during the day. *CCA1* and *LHY* rise during late night and inhibit the evening loop consisting of *EARLY FLOWERING 3* (*ELF3*), *EARLY FLOWERING 4* (*ELF4*) and *LUX ARRHYTHMO* (*LUX*) as well as *GIGANTEA (GI)* and *ZEITLUPE (ZTL). ELF3, LUX,* and *ELF4* in turn inhibit the morning complex *PRR* genes in the early night, thus closing the negative feedback cycle (Staiger et al., 2013).

The analysis of several other species, including CAM plants (Mallona et al., 2011), showed conservation of core clock genes, even so variation in the interconnected autoregulatory loops may occur. The genetic structure of the plant circadian clock has been analysed in *Solanaceae* where there are conserved gene duplications of some of the clock genes (Bombarely et al., 2016). Information on the molecular clock architecture in legumes exist for soybean vegetative tissue under drought stress (Marcolino-Gomes et al., 2014), as well as for soybean seeds (Hudson, 2010).

Orthologues of the known clock components *LHY/CCA1* and *TOC1* are expressed in a pattern similar to leaves in case of *GmCCA1* and *GmLHY,* but not in case of *GmTOC1* in soybean (Marcolino-Gomes et al., 2014). Microarray expression profiling in developing soybean seeds shows that 1.8% of the mRNAs detected in seeds with predicted functions in protein synthesis, fatty acid metabolism, and photosynthesis are expressed in a circadian rhythm. Thus, circadian oscillator genes are probably controlling the gating of these processes in seed tissue. No information exists for the organisation of circadian clock genes in pods and in relation to storage protein accumulation.

Although the plant circadian clock is known with detail in Arabidopsis leaves and seedlings, there is increasing data showing that there are differences in the timing and structure of the clock in different plant tissues (Endo, 2016). Studies in leaves have shown that mesophyll cells and vasculature have distinct circadian timings where the vasculature regulates the clock in other tissues (Endo et al., 2014). Further differences have been identified between leaves and roots (Bordage et al., 2016; James et al., 2008), probably as an adaptation to local requirements of the different organs. The differences between roots and leaves may be attributed to different light inputs (Bordage et al., 2016).

In the present work, we identified the core clock genes and genes codifying for storage globulins in cowpea by phylogenetic analysis. We used this information for transcriptomic analysis by qPCR in leaves, pods and seeds in six hour intervals in order to evaluate the expression pattern both of core clock components and of storage protein genes over time in the different tissues. This information will help to understand the determinants of extent and timing of protein storage. Furthermore, storage protein genes may be valuable expression markers for protein rich genotype selection and information on the particular expression pattern and the appropriate sampling time is

essential. Comparative gene expression analysis in different tissues, developmental stages and under varying environmental conditions requires essential reference points in form of genes with stable expression levels. This work therefore includes the identification of candidate reference genes, which can be reliably used to normalise the expression of genes of interest in cowpea leaves, seeds and pods.

3.2. Material and methods

3.2.1. Plant material, phenotyping and sampling

Plant material was sampled under field conditions. For the latter, the IT97K-499-35 breeding line of cowpea was grown under field conditions at the ‘‘Tomás Ferro’’ Experimental Agro-Food Station, Universidad Politécnica de Cartagena located in southeast of Murcia region, Campo de Cartagena, Spain.

Average temperatures during sampling time were 29.1°C. Leaves and pods at different stages of development were sampled at 6:45am, 12:45pm, 6:45pm, and 00:45am. These times corresponded in July to subjective time of T0, T6, T12 and T18 considering T0 as dawn. Time of sunset was at 9:31 pm. Expression of reference genes was analysed at T6 and T18, clock genes and protein storage genes at T0, T6, T12 and T18. Leaves were harvested when first pods matured. Developmental stages of pods were categorized based on the phenotyping of whole pods containing seeds and seeds only. Phenotyping included measurements of weight, length and width of 10 pods and 5 seeds per pod for three ripening stages: immature, intermediate and mature (Figure 10). These stages corresponded to 4 days, 7 days and 15 days after anthesis. Gene expression analysis was performed using four independent samples of leaves as well as immature, intermediate and mature pods and intermediate and mature seeds at each time point. The seeds were harvested from different pods.

3.2.2. Identification of genes for normalization, circadian clock genes and storage protein genes in cowpea

We identified candidate reference genes from legumes using the gene

expression atlas from *Medicago truncatula* (Benedito et al., 2008) and a set of genes found suitable for normalization in soybean (Hu et al., 2009). We used the accession numbers to identify cowpea genomic sequences by BLAST (harvest-web.org). Scaffolds were retrieved and using legume translated mRNAs, we identified putative mRNAs from cowpea using Genewise (Birney et al., 2004) (Supplementary Table 3). The genes used were *B-ACTIN (ACT), ACTIN 2/7 (ACT27), CYCLOPHYLIN (CYP), ELONGATION FACTOR 1-A (EF1A), ELONGATION FACTOR 1-B (EF1B), ALPHA TUBULIN (TUA4), BETA TUBULIN (TUB4), ASK-INTERACTING PROTEIN 16 (SKIP16)* and an hypothetical unknown protein from soybean *(UKN2).* The genes related to protein storage accumulation were *LEGUMIN (LEG), LEGUMINJ (LEGJ) and COVICILIN (CVC).* Circadian clock related genes were *GIGANTEA (GI), TIMING OF CAB EXPRESSION1 (TOC1), LATE ELONGATED HYPOCOTYL (LHY),* and *EARLY FLOWERING 3(ELF3).* Primers were designed using the software PCR efficiency as described previously (Mallona et al., 2011) (Supplementary Table 3). Primers were tested for stable, single and clear amplification products by end-point PCR with genomic DNA, visualized on 1.5% agarose gels (Supplementary Figure 2) and by quantitative PCR to assess the melting profile of the PCR products (Supplementary Figure 3).

3.2.3. Quantitative PCR

Total RNA was isolated from 100 mg homogenized plant material using an RNeasy Mini Kit for leaves and pods without seeds (Qiagen, Hilden, Germany), and a phenol-based method for seeds (Box et al., 2011). RNA concentration and purity was estimated from the ratio of absorbance readings at 260 and 280 nm. cDNA synthesis was performed with 0.5 µg of total RNA using M-MLV reverseTranscriptase (Maxima First Strand cDNA kit for RT-qPCR, with dsDNase, ThermoFisher Scientific) according to the manufacturer instruction. Genes were amplified for four biological and two technical replicates in a Stratagene Mx3000P qPCR system (www.agilent.com), with sequence-specific primers (Supplementary Table 3) synthesized by Invitrogen (www.invitrogen.com) using SYBR-Green Mastermix (ThermoFisher Sciencific) and a 25 ng RNA equivalent of cDNA. The reaction mix was subjected to the following protocol: 95°C for 30 sec followed by 45 cycles of 95°C for 10 sec, 57°C for 15 sec and 72°C for 15

sec, and a subsequent standard dissociation protocol.

3.2.4. Bioinformatics and statistical analysis

For the identification of stable reference genes during different developmental stages and tissue, PCR efficiencies and CT values were used in a web-pipeline that contains the different PCR analysis software's Bestkeeper, Normfinder, Delta CT and Genorm. PCR efficiency was calculated as described before (Mallona et al., 2011). Data from different analysis was pooled and ranked using Rank-Aggreg (Pihur & Datta, 2009). We used the software Geomean to obtain a ranking value of the candidate reference genes (Xie et al., 2012).

Statistical analysis of diurnal gene expression profiles for clock relates genes and storage protein related genes was performed using the normalized cycle threshold (Ct) values calculated as described previously (Marcolino-Gomes et al., 2014). A PCR efficiency of 2 for all primer combinations was used for the calculation of normalized expression (NE) based on efficiency calculations, which were performed as described previously with the qpcR R package (Mallona et al., 2011; Spiess et al., 2008). Average efficiencies were 1,98 for *VuEF1A*, 1,99 for *VuGI,* 1,97 for *VuELF3*, 1,95 for *VuTOC1*, 1,99 for *VuLHY*, 1,99 for *VuCVC*, 1,93 for *VuLEG* and 1,99 for *VuLEGJ*. RAIN method was applied for the determination of existence of a circadian biological rhythm represented in the transcriptome data (Thaben & Westermark, 2014). Statistical analysis for gene expression was performed using group-wise comparison with the REST program (Pfaffl et al., 2002). Phenotypic data were analysed for homogeneity of variance with the Fligner-Killeen test. The parameters showing homogeneity of variance were analysed using ANOVA and Tukey's HSD test, while the non-parametric data were analysed using Wilcoxon signed rank test with continuity correction in R version 3.2.3.

3.2.5. Phylogenenetics

For phylogenetic reconstructions, the identified *V. unguiculata* genes were used to identify orthologues and paralogs from other legumes. We used MEGA7 for evolutionary analysis (Kumar et al., 2016). Translated cDNAs were aligned with

CLUSTALW (Larkin et al., 2007). The evolutionary history was inferred using the Neighbor-Joining method (Saitou & NEI, 1987). The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the analysed taxa (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl & Pauling, 1965) and are in the units of the number of amino acid substitutions per site.

3.3. Results and discussion

3.3.1. Phenotyping of pods and seeds

In order to gain a more profound inside into the developmental patterns of cowpea seeds and pods, we characterized the growth in weight, length and width for 10 pods and 5 seeds per pod at 4, 7 and 15 days after anthesis of these generative tissues (Figure 10).

Seed weight, length and width increased during the entire pod development, even so weight increases were more prominent during the second week after anthesis (Figure 11). Differences were significant for all parameters both between immature and intermediate seeds as well as between intermediate and mature seeds (Wilcoxon test p=0 for weight; ANOVA p=0 for length and width). The lightest immature and the heaviest mature seeds showed a weight difference of 300 mg and length and width varied by a maximum of 1 cm.

Whole pod weight, including the seeds, increased up to 6000 mg between the immature and mature ripening stage (Figure 11). Differences in pod weight were most prominent between immature and intermediate pods (ANOVA p=0), while no significant differences were observed between intermediate and mature tissue (ANOVA p=0.67). Some of the intermediate pods were heavier than some of the mature pods. Similar results were observed for pod length with significant differences between immature and mature (ANOVA p=0) but not between intermediate and mature pods (ANOVA p=0.951).

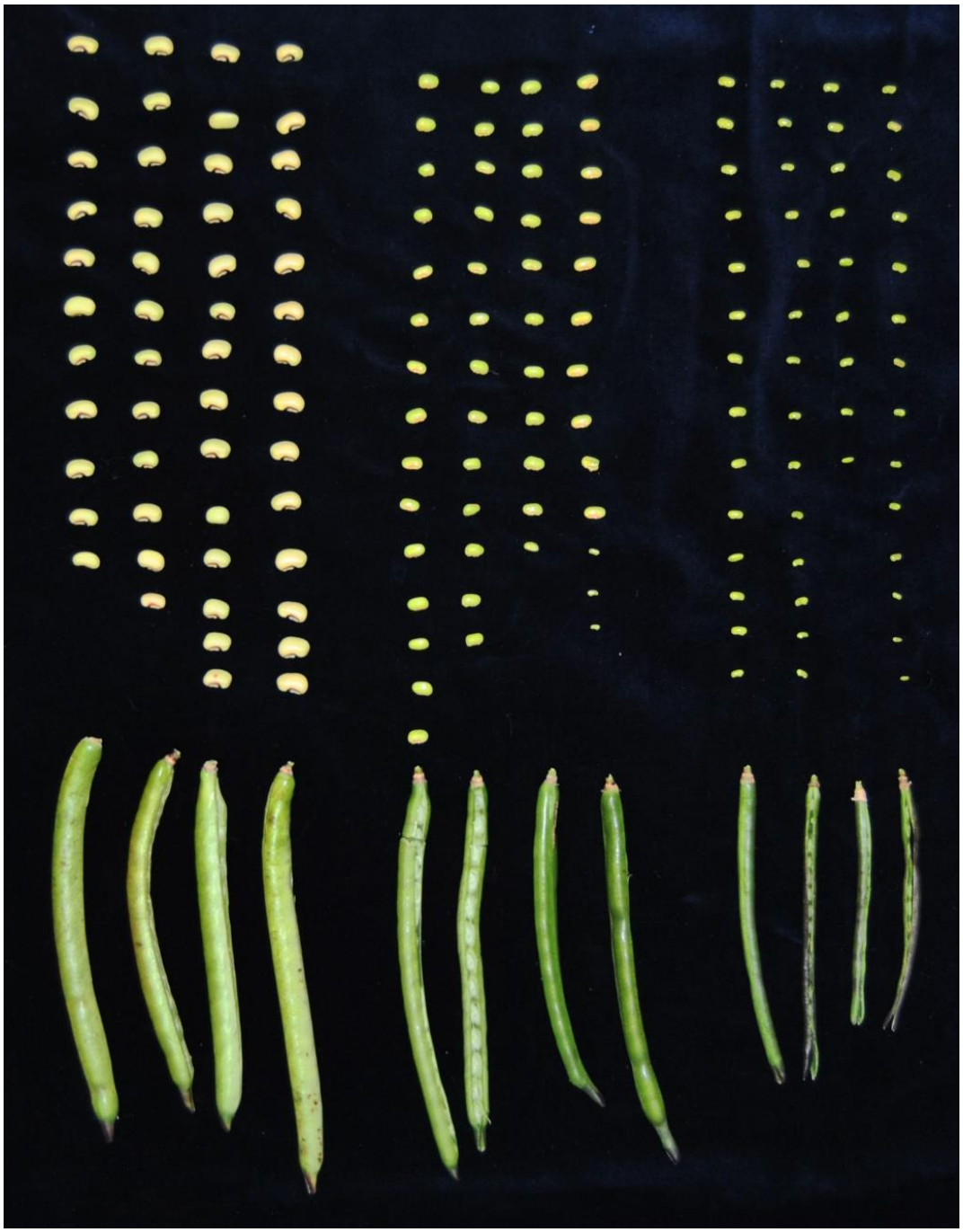


Figure 10. Pods of cowpea with their respective seeds used in the current study. From left to right completely grown pods and seeds before maturation started, intermediate pods and seeds and immature pods and seeds.

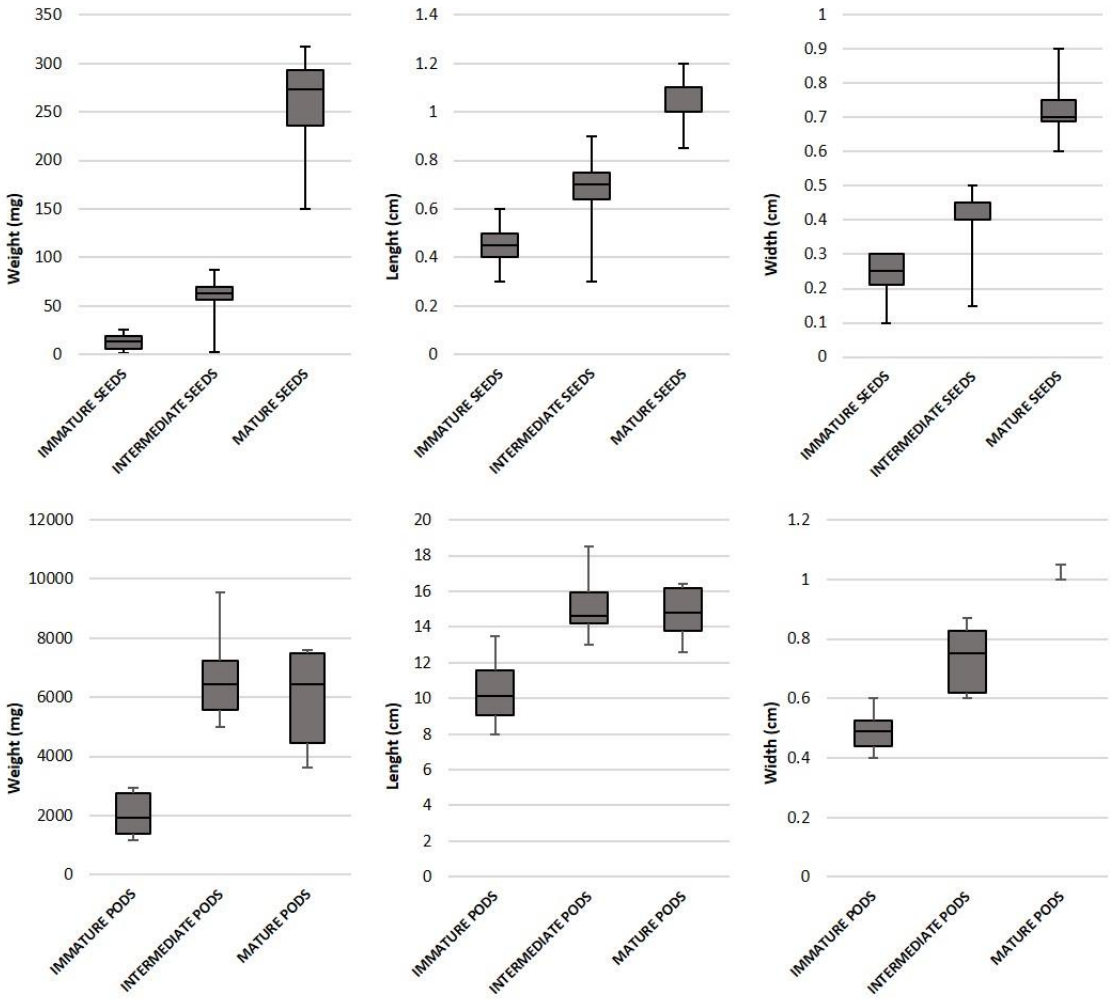


Figure 11. Boxplot of weight, length and width of seeds and pods in immature, intermediate and mature stage.

As already observed in case of weight, some intermediate pods were longer than the mature pods included in the analysis. These results suggest a variance in achieving maximal length that maybe up to a week. We measured up to 8 cm of difference in length between the shortest and the longest pods, while pod width varied by 0.6 cm from the thinnest to the widest pod. Pod width increased continuously during development with significant differences both between immature and intermediate (Wilcoxon test p=0.006) and between intermediate and mature pods (Wilcoxon test p=0.002). Mature pods had an extremely homogeneous width (Figure 11). Our data indicate that growth of cowpea pods, including the seeds, both in weight and length, occurs during the first week after pollination. Thereafter, pods further expand only in width. Prominent length growth of pods shortly after pollination is also observed in common bean, azuki bean and mung bean compared to whole soybean pods, which

grow very slow during the first week after flower opening (Zheng et al., 2004). The fact that whole pod weight increases prominently during the first week after anthesis but seed weight during the second week, indicates, that initial pod growth is the result of pod tissue growth.

3.3.2. Identification of genes for normalization, circadian clock genes and storage protein genes in cowpea

Primers, designed for the identified cowpea sequences of all genes, were used for end-point PCR on genomic DNA. All primers gave single clear amplification products with the expected size, except for *TUB4* that showed an apparent size of 500 bp as compared to the expected 250 bp. This, however, was the result of amplifying a short intron present in the gene (Supplementary Figure 2). *VuVic* was discarded from the analysis due to unspecific amplification products. Sequencing results also indicated a mixture of amplification products. We also tested the quality of the amplification in quantitative PCR to assess the melting profile of the PCR products, which gave single peaks at constant T_m for all tissues and developmental stages (Supplementary Figure 3).

Sequence alignments with the identified *V. unguiculata* clones derived from Noble VuGEA database confirmed the correct identity of amplification products for all genes. Sequences encoding putative circadian clock genes in cowpea were identified through alignment of amino acid sequences of core circadian clock genes from Arabidopsis and several legumes including *Glycine max*. We identified potential soybean homologs of *GI*, *ELF3, TOC1* and *LHY*. In order to establish the putative orthology of the different transcripts identified, we performed a phylogenetic analysis of the cowpea genes.

The gene *VuLHY* was found as a single scaffold suggesting that, as previously reported for model legumes (Hecht et al., 2005) and other species such as Petunia or *Solanum lycopersicum,* it is a single copy gene (Bombarely et al., 2016). The *LHY* orthologs of the legumes *Phaseolus vulgaris*, *Glycine max*, *Vigna radiata*, *V. unguiculata* and *V. angularis* appeared in a phylogenetic reconstruction on a single clade comprising genes from *Castanea* and *Populus* (Figure 12). They were separated from a second major clade comprising the Arabidopsis *LHY*, the paralog *CCA1* and other *MYB* genes such as

MIXTA from *Antirrhinum majus* (Noda et al., 1994) or *ENHANCER OF BENZENOID* from Petunia (Van Moerkercke et al., 2011). This indicates an early departure of these genes from a common ancestor previous to the duplication in giving rise to *CCA1* and *LHY*.

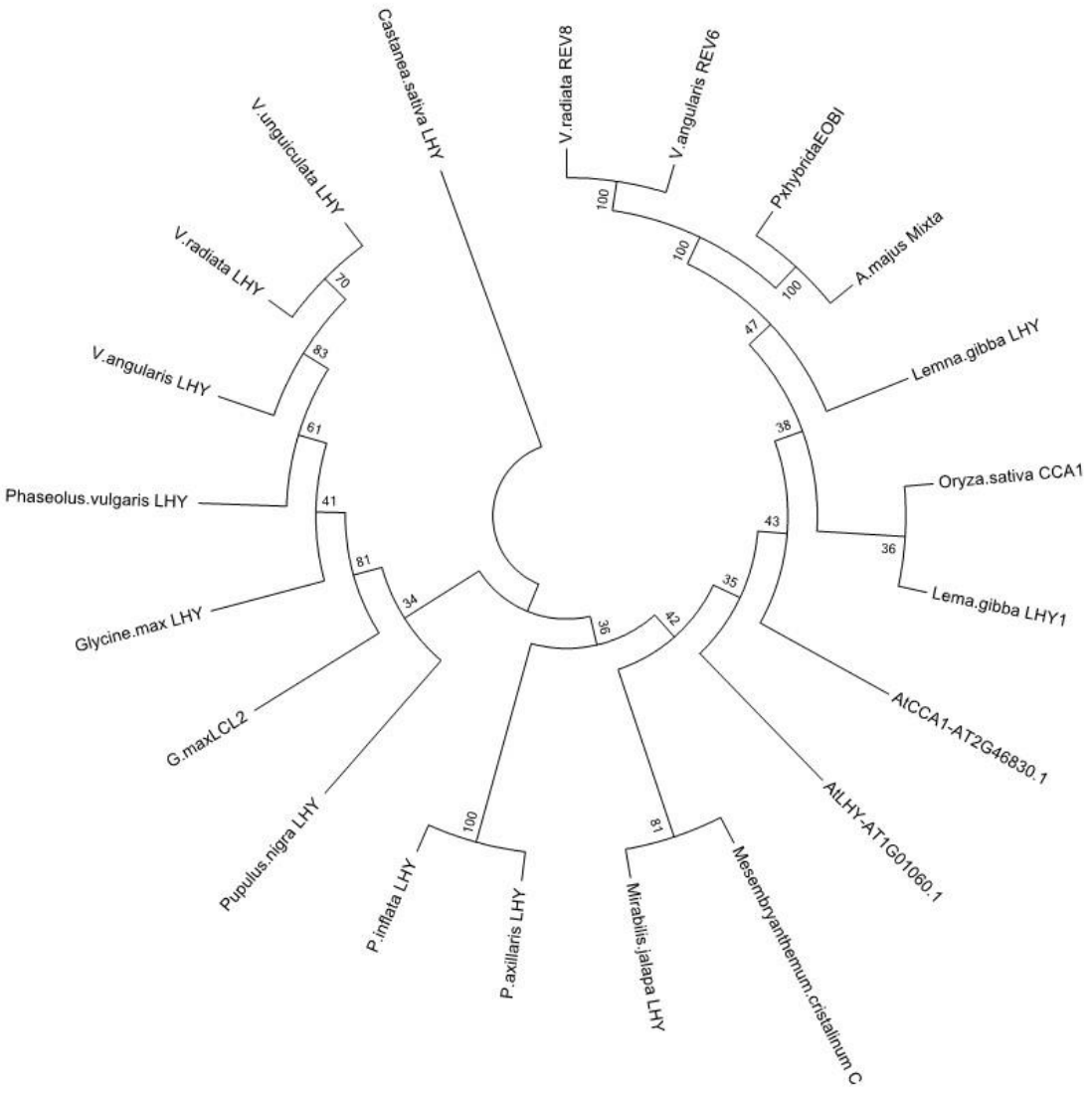


Figure 12. Phylogenetic tree of LHY related predicted proteins. The analysis involved 21 amino acid sequences. All positions containing gaps and missing data were eliminated.

The gene *ELF3* is a single copy gene in Arabidopsis but is present in two to four copies in other species such as Petunia or *Physcomitrella* (Bombarely et al., 2016; Ryo et al., 2016). Although *ELF3* had been previously reported as being absent from soybean (Hecht et al., 2005), a recent ORFeome analysis under drought conditions identified a *bona fide Glycine max ELF3* gene (Chai et al., 2015). The complete set of *ELF3* genes from legumes formed a distinct clade separated from a second one comprising the *ELF3*

paralogs found in *Solanaceae*, *Arabidopsis*, *Oryza* and *Physcomitrella*. This suggests an early separation of this gene in legumes (Figure 13).

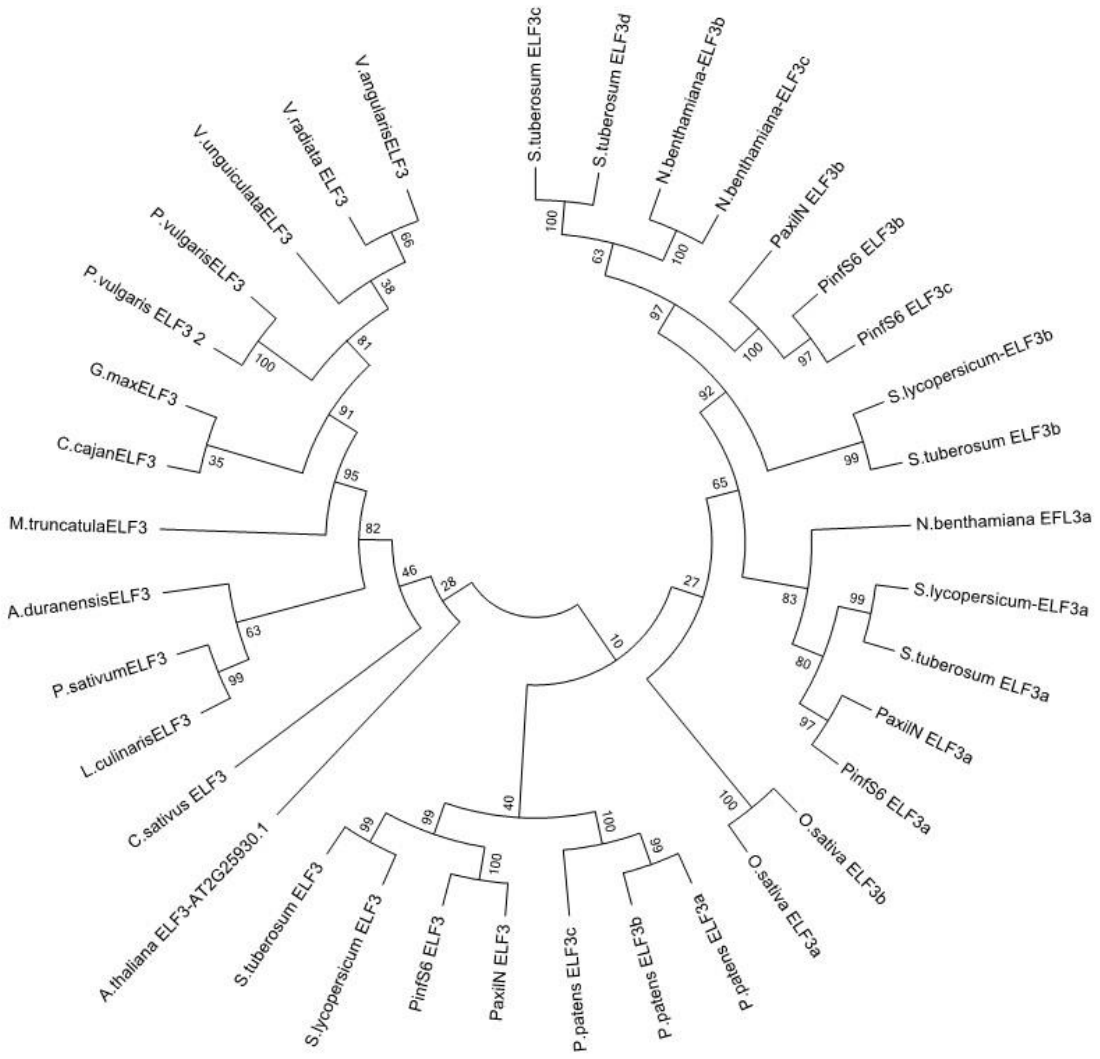


Figure 13. Phylogenetic tree of ELF3 related predicted proteins. The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated.

We analysed the phylogeny of *VuTOC1* and found that it formed a subclade with the rest of the *TOC1* genes (Figure 14), and clearly separated from the *PRR9/5* and *PRR7/3*. In contrast to the tree structures found for *VuLHY* (Figure 12) and *VuELF3* (Figure 13), the *TOC1* orthologs of legumes were closer to *AtTOC1*, indicating a strong degree of conservation.

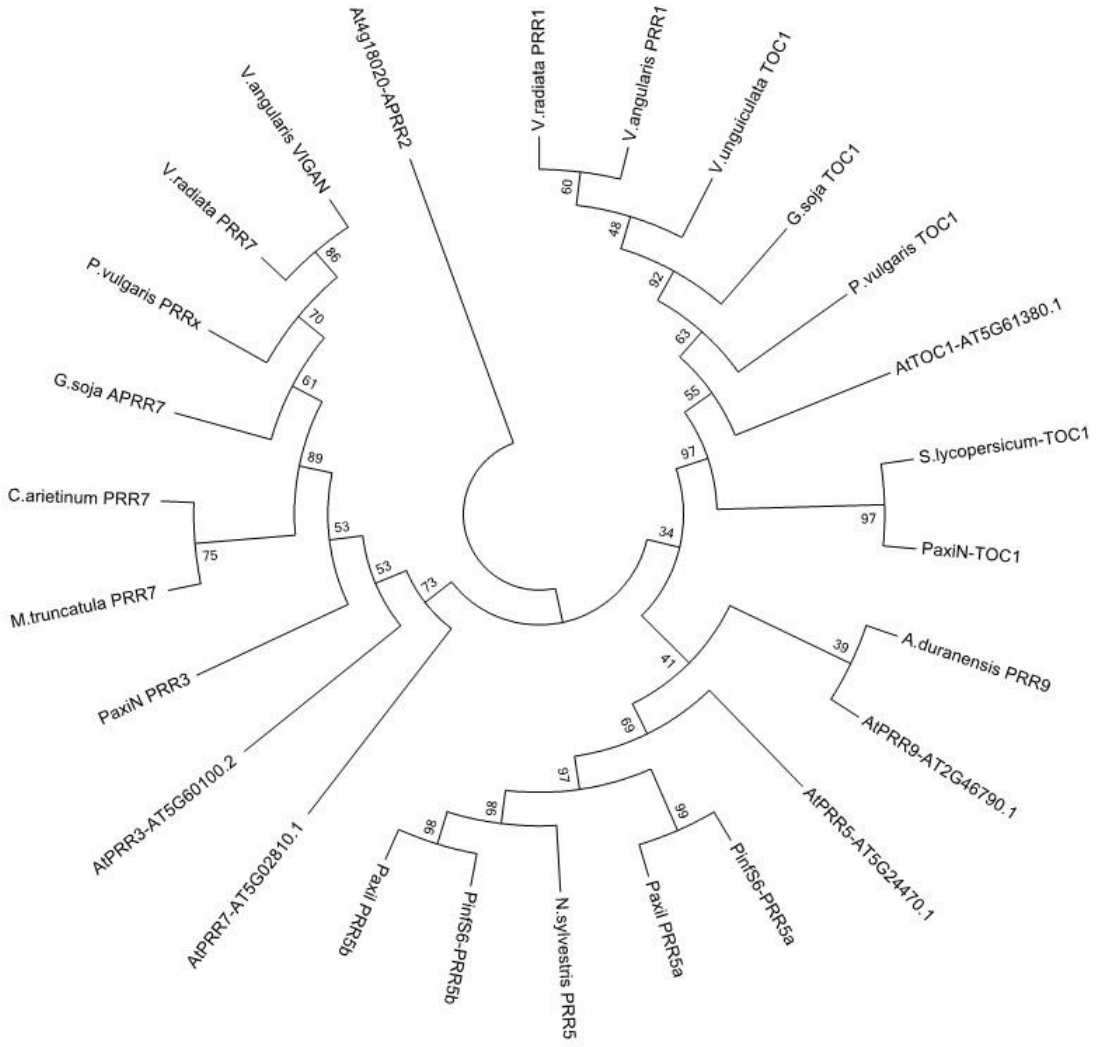


Figure 14. Phylogenetic tree of TOC1 related predicted proteins. The analysis involved 26 amino acid sequences. All positions containing gaps and missing data were eliminated.

The *GIGANTEA* gene is a single copy gene in Arabidopsis and is found in one to four copies in *Solanaceae* and legumes (Fowler et al., 1999; Marcolino-Gomes et al., 2014; Bombarely et al., 2016). We found one scaffold and one EST (Vun_T01130.1_6). The ORF giving high homology to GI and orthologs was found in the -3 frame suggesting that the aforementioned fragment had been annotated in the reverse position. The phylogenetic reconstruction of *VunGI* showed that, as previously found for *Solanaceae* and legumes (Bombarely et al., 2016), all GI orthologs clustered into three clades comprising monocots, dicots and basal angiosperms (Figure 6). As expected *VungGI* clustered together with the rest of the legume genes used for this phylogenetic reconstruction. Within the different clades, duplicated genes such as those found *Zea*

mays, *Glycine* or *Nicotiana benthamiana*, showed differing levels of divergence, probably depending on the timing of whole genome duplications that occurred in these species.

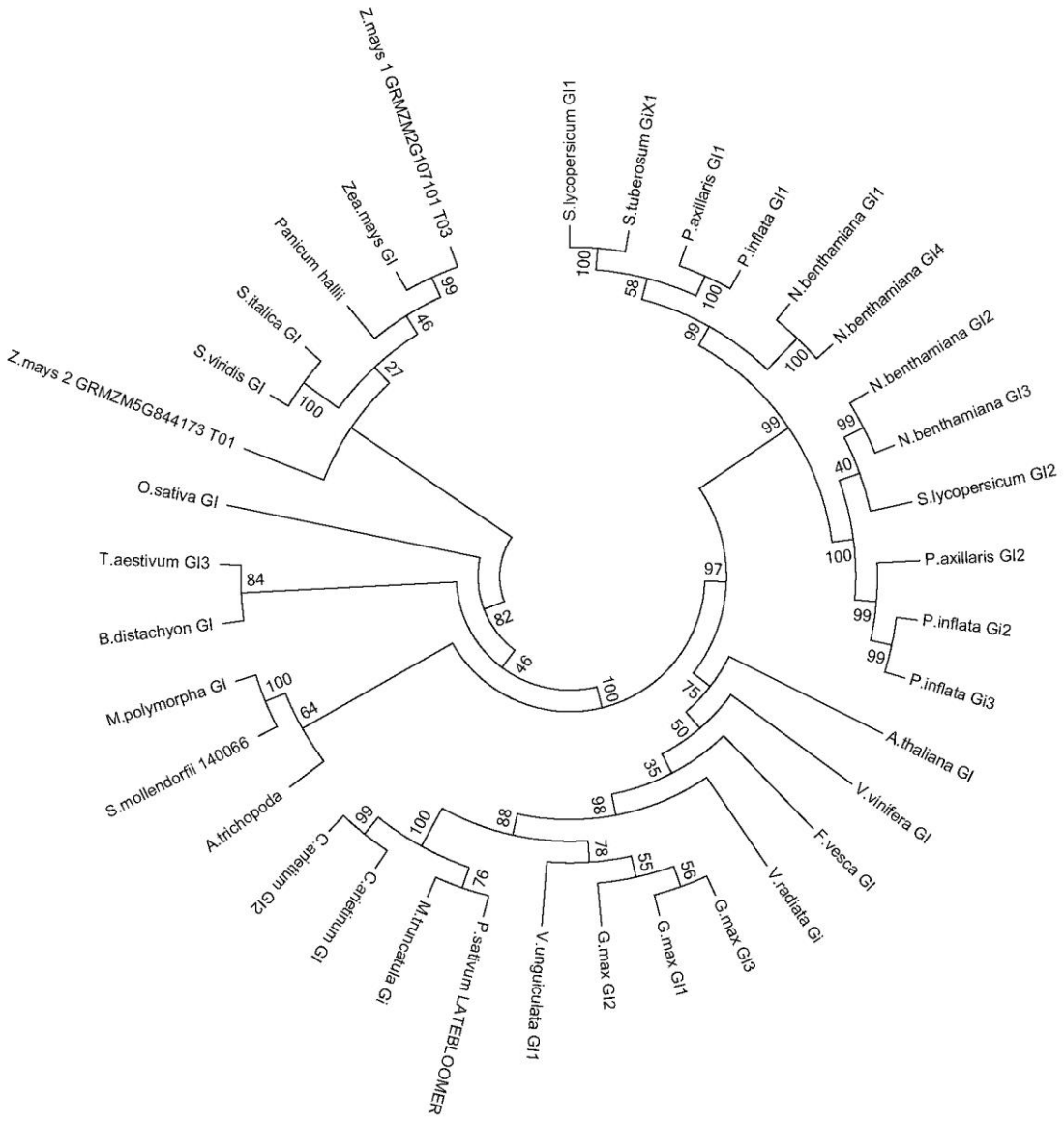


Figure 15. Phylogenetic tree of GI related predicted proteins. The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated.

3.3.3. Data mining for stable reference genes

The identification of reference genes for normalization of quantitative PCR is a prerequisite for reliable gene-expression analysis. We used *BETA-ACTIN (ACT), ACTIN 2/7 (ACT27), CYCLOPHYLIN (CYP), ELONGATION FACTOR 1-A (EF1A), ELONGATION*

FACTOR 1-B (EF1B), ALPHA TUBULIN (TUB4), BETA TUBULIN (TUB4), ASK-INTERACTING PROTEIN 16 (SKIP16) and a *HYPOTHETICAL PROTEIN UNKNOWN* from soybean *(UKN2).* There are a number of programs differing in the mathematical solution to identify stable genes including geNorm (Mestdagh et al., 2009), Normfinder (Andersen et al., 2004), Bestkeeper (Pfaffl et al., 2004) and the comparative delta Ct methods (Schmittgen & Livak, 2008). Table 15 shows the ranking of reference genes for the different analysis software and as result of the pooled PCR analysis software. Table 16 gives the Geomean of ranking values of the candidate reference genes based on the geometric mean of the weights of every gene calculating by each program (Xie et al., 2012). Ideally, expression of a reference gene should be independent of the morphogenetic process in order to validate transcriptomic changes of target genes during plant development. *ELONGATION FACTOR 1-A (EF1A)* was found to be the most suitable reference gene while *BETA TUBULIN (TUB4)* was the least suitable gene. *EF1A* was also shown to be a suitable reference gene for potato during biotic and abiotic stress conditions (Nicot et al., 2005) and for Petunia over a wide range of developmental stages, varieties extraction and qPCR procedure (Mallona et al., 2010) *EF1A* was further used for normalisation of transcript levels of clock genes and storage protein related genes in this study.

Table 15. Ranking of normalisation genes for cowpea transcriptomic analysis based on Rank-Aggreg.

Method	1	2	3	4	5	6	7
Delta CT	EF1A	Tua4	Act	Cyp	Skip16	Act27	Tub4
BestKeeper	Act	Tua4	EF1A	Skip16	Cyp	Act27	Tub4
Normfinder	EF1A	Tua4	Skip16	Act	Cyp	Act27	Tub4
Genorm	AF1A/Tua4		Skip16	Act	Cyp	Act27	Tub4
Recommended comprehensive ranking	**EF1A**	**Tua4**	**Act**	**Skip16**	**Cyp**	**Act27**	**Tub4**

Table 16. Geomean of ranking values of the candidate reference genes.

Gene	Geomean of ranking values
EF1A	1.32
Tua4	1.68
Act	2.63
Skip 16	3.66
Cyp	64.73
Act27	6.00
Tub4	7.00

3.3.4. Diel expression of the circadian clock genes VuGI, VuELF3, VuLHY, and VuTOC1 in leaf, pod and seed tissue under field conditions

In leaves, the genes *VuGI, VuTOC1* and *VuLHY* showed a significant rhythm during a time series of 24 hours, while significance in case of *VuELF3* was borderline (Figure 16; Table 17). The expression of *VuGI* and *VuELF3* was highest at 18 hours after dawn and lowest during morning and midday. This evening/night-phased expression is comparable to previous circadian oscillation reports (Alabadi et al., 2001; Egea-Cortines et al., 2013; Marcolino-Gomes et al., 2014). Transcript levels for *VuTOC1* in leaves showed a significant peak towards midday, which is in contrast to the usual found peak during the transition from day to night (Alabadi et al., 2001; Locke et al., 2006). As critical part of the clock function in Arabidopsis, *TOC1* participates in the negative regulation of the MYB transcription factors *LHY* and *CCA1* (Huang et al., 2012), while *LHY* and *CCA1* negatively regulate *TOC1* expression by binding to regulatory sequences in the *TOC1* promoter (Alabadi et al., 2001; Locke et al., 2006). Nevertheless, the relative expression amplitude over 24 hours resembled those found in soybean leaves (Marcolino-Gomes et al., 2014). *VuLHY* expression in leaves showed a marked increase towards midday compared to dawn, comparable to the typical morning peak observed over a wide range of tissues in Arabidopsis (Locke et al., 2006). This may be due to the fact that *CCA1/LHY* is regulated through a negative feedback loop not only by *TOC1*, but together with other genes, including *GI, ELF3, ELF4* and *LUX* (McClung CR, 2006). We observed the highest peak phase expression amplitude for *VuLHY*, followed by *VuELF3* and *VuTOC1.* These results are similar to soybean leaves, were relative expression amplitude at peak phase was highest for *GmLHY* followed by *GmGI*, *GmPRR* genes and *GmELF4* (Marcolino-Gomes et al., 2014).

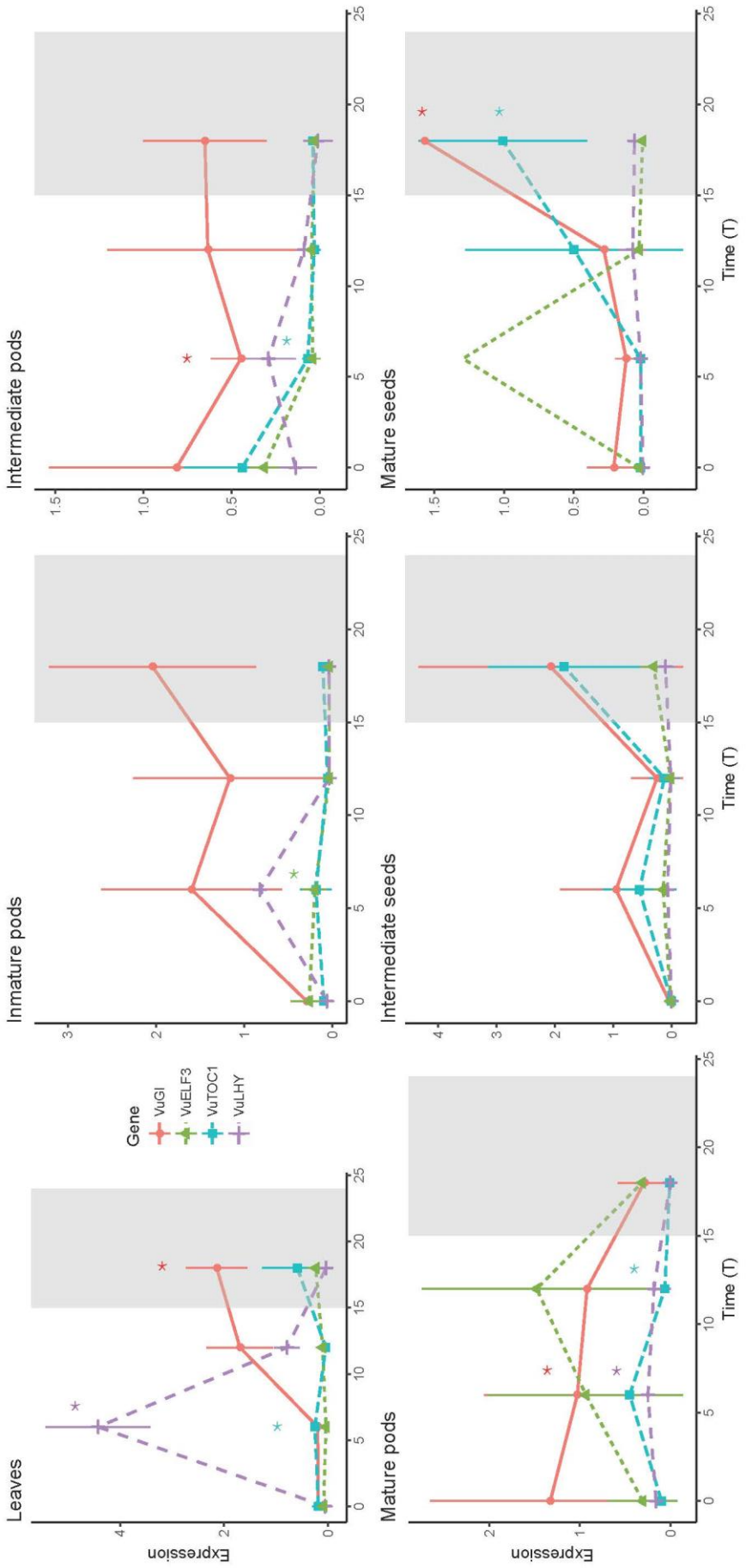


Figure 16. Expression of the circadian clock genes *VuGI*, *VuELF3*, *VuTOC1* and *VuLHY* in cowpea leaves, pods and seeds. Expression represents the normalized expression NE according to the formula (NE) = 2^-(Ct experimental – Ctn). Collection points represent Time 6, 12, 18 and 24 hours after dawn. Four samples were analyzed for each time point and error bars indicate the standard deviation. A significant rhythm in the time series according to RAIN (Rhythmicity Analysis Incorporating Nonparametric methods) is marked with an asterisk in position of peak phase. Error bars indicate standard deviations.

Table 17. Significance of 24 hour rhythms and phase of peak expression for circadian clock genes and protein storage genes in leaves, pods and seeds of cowpea sampled under field conditions.

Tissue	Gene	*p*-Value	Phase of peak	Significance level (*=$p \leq 0.05$; **=$p \leq 0.01$; ***$p \leq 0.001$)
LEAVES	GI	3.41E-05	18	***
	ELF3	0.078	18	
	TOC	4.86E-02	6	*
	LHY	9.78E-05	6	***
INMATURE PODS	GI	9.54E-01	18	
	ELF3	0.043	6	**
	TOC	4.20E-01	6	
	LHY	1.58E-01	6	
	CVC	2.33E-02	12	*
	LEG	1.66E-04	6	***
	LEGJ	4.84E-02	6	*
INTERMEDIATE PODS	GI	1.00E+00	6	**
	ELF3	0.149	18	
	TOC	4.80E-03	6	*
	LHY	3.94E-02	6	
	CVC	7.56E-01	12	
	LEG	9.22E-04	6	***
	LEGJ	5.35E-01	18	
MATURE PODS	GI	4.97E-04	6	***
	ELF3	0.273	12	
	TOC	6.60E-10	12	***
	LHY	5.93E-07	6	***
	CVC	2.63E-04	6	***
	LEG	1.48E-01	12	***
	LEGJ	1.66E-04	6	
INTERMEDIATE SEEDS	GI	7.11E-01	18	
	ELF3	0.149	18	
	TOC	7.80E-01	6	
	LHY	7.57E-01	18	
	CVC	8.33E-06	12	***
	LEG	3.95E-03	12	**
	LEGJ	8.33E-06	12	***
MATURE SEEDS	GI	1.44E-02	18	*
	ELF3	0.107	6	
	TOC	6.70E-03	18	**
	LHY	6.65E-01	12	
	CVC	6.50E-02	6	
	LEG	1.65E-01	12	
	LEGJ	3.66E-04	12	***

Except for *VuELF3*, immature pods did not have a significant rhythm in the clock genes analysed. As development advanced, we found significant rhythms in intermediate pods in *VuGI, VuTOC,* and at mature stages for *VuGI, VuTOC and VuLHY*. It is interesting that *VuTOC* and *VuLHY* regained a similar rhythm to that found in leaves, whereas *VuGI* and *VuELF3* showed a shift in the point of maximum expression. A similar picture was seen in seeds that showed a lack of rhythm at intermediate stages of

development, followed by a rhythm resetting during the mature seed stage for *VuGI* and *VuTOC,* but not *VuLHY and VuELF3*.

Our results indicate that the expression of some clock related genes in generative tissue resemble the oscillation pattern and amplitude observed in leaves while others diverge, depending on the developmental stage and tissue type. The rhythmicity observed in leaves for *VuTOC1* and *VuLHY,* but not of *VuGI and VuELF3,* was maintained in ripening pod tissue, but not in seeds. Our findings are in agreement with results from Arabidopsis, indicating the existence of autonomous and functionally independent circadian clocks in different plant tissues (Thain et al., 2000; Thain et al., 2002). The clock in pods and seeds differed from each other and from the leaf clock, indicating a complex network of signals that may be controlling inter tissue coordination of circadian gene expression.

On the other hand, doing a comparative analysis of gene expression at different times of the day during development (Table 18) we found that the gene *VuTOC* was significantly down regulated in immature versus mature pods at T18 (17.57 fold p=0.03) and intermediate versus mature pods at T12 (3.83 fold p=0.002). In addition gene *VuELF3* was significantly down regulated in immature versus intermediate pods at T6 (3.98 fold p=0.001); significantly up regulated in intermediate versus mature pods at T12 (17.79 fold p=0.001); significantly up regulated in immature versus mature pods at T24 (40.62 fold p=0.001); and significantly up regulated in intermediate versus mature pods at T24 (1354.72 fold p=0.001).

Table 18. Comparative analysis of gene expression at different times of the day during development of different tissues.

			Tissue											
			Immature pods vs mature pods			Immature pods vs intermediate pods			Intermediate pods vs mature pods			Intermediate seeds vs mature seeds		
	Gene	Time	*p*-Value	Expression factor	SE (±)	*p*-Value	Expression factor	SE (±)	*p*-Value	Expression factor	SE (±)	*p*-Value	Expression factor	SE (±)
Protein storage genes	*VuCvc*	T6	0.12	29.65	3.72	0.35	8.32	3.17	0.96	3.57	1.33	0.48	137.41	5.20
		T12	**0.02***	-16.99	2.29	0.97	1.25	0.14	0.06	-21.23	1.98	1.00	-1.16	0.57
		T18	0.30	-4.30	1.00	0.20	4.24	0.94	**0.03***	-21.69	1.63	0.52	32.03	5.12
		T24	1.00	-3.77	1.07	0.83	-1.48	0.18	0.84	-2.70	0.29	0.99	-1.98	0.36
	VuLeg	T6	0.47	-4.71	0.33	0.46	3.97	1.82	**0.02***	-18.68	1.72	0.92	-9.48	1.59
		T12	0.24	4.74	0.67	0.96	1.77	0.91	0.73	2.43	1.27	0.51	-31.14	2.41
		T18	0.70	-1.36	0.18	0.60	-2.15	0.33	0.77	1.61	0.97	0.91	5.76	2.66
		T24	0.92	-8.77	1.60	0.63	-2.43	0.71	0.82	-3.54	0.26	0.88	-8.20	1.17
	VuLegJ	T6	0.17	7.15	1.61	0.79	2.91	1.18	0.93	2.49	0.50	0.52	74.84	3.59
		T12	0.30	-11.12	1.80	0.92	-3.96	1.13	0.54	-2.81	0.38	0.86	-8.74	0.59
		T18	0.43	2.06	0.47	0.15	15.87	2.06	0.51	-7.72	0.82	0.86	3.50	3.05
		T24	0.97	2.15	0.95	0.93	1.71	0.38	0.92	1.26	0.75	0.95	-1.77	1.09
Circadian clock genes	*VuGI*	T6	0.82	-1.65	0.27	0.70	-3.23	0.21	0.91	1.96	1.02	1.00	-5.20	0.85
		T12	0.99	-1.21	0.31	0.57	-7.57	0.92	0.31	6.27	2.84	0.88	3.24	1.37
		T18	0.10	-6.80	1.85	0.69	-3.03	0.42	0.86	-2.24	0.13	0.86	51.12	3.17
		T24	0.08	3.50	1.07	0.80	1.58	0.19	0.40	2.21	1.34	0.99	6.35	1.23
	VuTOC	T6	0.42	2.67	1.72	0.61	-2.11	0.28	0.40	5.65	1.98	0.98	-11.37	1.39
		T12	0.99	1.60	1.02	0.97	-2.40	0.09	**0.002****	3.83	1.43	0.60	10.09	2.00
		T18	**0.03***	-17.57	2.22	0.65	2.09	1.52	0.56	-36.67	2.66	0.99	6.35	1.83
		T24	0.88	1.26	0.19	0.19	3.75	0.86	0.05	-2.97	0.08	0.83	5.97	0.58
	VuLHY	T6	0.43	-4.27	0.44	0.32	-3.15	0.19	0.85	1.32	0.50	1.00	-1.07	0.88
		T12	0.84	4.14	2.08	0.75	3.15	1.66	1.00	1.31	0.91	0.06	4.48	1.13
		T18	0.07	-8.28	1.93	0.99	-1.36	0.45	0.89	-6.11	1.10	0.88	10.39	1.48
		T24	0.37	3.04	1.77	0.68	2.75	1.80	0.96	1.10	0.07	0.99	1.80	1.47
	VuLHY	T6	0.43	-4.27	0.44	0.32	-3.15	0.19	0.85	1.32	0.50	1.00	-1.07	0.88
		T12	0.84	4.14	2.08	0.75	3.15	1.66	1.00	1.31	0.91	0.06	4.48	1.13
		T18	0.07	-8.28	1.93	0.99	-1.36	0.45	0.89	-6.11	1.10	0.88	10.39	1.48
		T24	0.37	3.04	1.77	0.68	2.75	1.80	0.96	1.10	0.07	0.99	1.80	1.47
	VuELF3	T6	0.49	-2.78	0.73	**0.001****	-3.98	0.33	0.51	1.42	8.89	**0.001****	105.26	7.68
		T12	0.51	6.74	7.90	0.47	-1.12	2.86	**0.001****	17.79	5.11	0.49	12.07	4.18
		T18	0.52	4.92	3.20	0.51	-2.64	1.62	0.53	5.49	2.83	0.51	-23.89	1.78
		T24	**0.001****	40.62	1.58	0.52	-33.35	1.28	**0.001****	1354.72	1.85	0.51	5.71	13.36

3.3.5. Expression profile of storage protein genes in seeds and pods during development under field conditions

The relative expression of storage proteins *VuCVC, VuLEG* and *VuLEGJ* in pods and seeds in six hour intervals is given in Figure 17. We did not detect expression of *VuCVC, VuLEG* and *VuLEGJ* in leaves. Peak gene expression amplitude for *VuCvc*, *VuLEG* and *VuLEGJ* was 49%, 54% and 68% lower in seeds of intermediate maturity compared to mature seeds. These differences were not significant due to high variability among samples. A higher expression level towards maturity is in contrast to the finding that storage protein accumulation in cowpea is most rapid during earlier stages of seed development (Awolumate, 1983). On the other hand, our results are in agreement with observations in *Lotus japonicus*, where storage protein production and peak expression levels of *vicilin*, *convicilin* and *legumin* like genes and precursor genes is concentrated towards the end of seed filling phase (Verdier et al., 2013). Results indicate that the strong seed weight increase observed during the second week after anthesis can be attributed prominently to the accumulation of storage proteins.

In cowpea pods, maximal expression levels of genes coding for storage globulins were 0.7-0.9% of those found in mature seeds. Legume pods are sources of food and feed and in *Lotus japonicus,* storage globulins form part of the pod enhanced transcripts (Verdier et al., 2013). A possible function of storage proteins in pods is the defence against predators as vilicins from cowpea and other legumes strongly bind to several chitin-containing structures and impair growth of seed feeding beetles (Sales et al., 2001). We compared expression levels during development considering the different sampling times (Table 18). We found that the gene *VuCVC* was significantly down regulated in immature versus mature pods at T12 (16.99 fold p=0.02) and intermediate versus mature pods at T18 (21.69 fold p=0.04). The expression of *VuLEG* was also significantly down regulated in intermediate versus mature pods at T6 (18.68 p=0.02). Despite the large changes in average gene expression found between intermediate and mature seeds, these differences were not significant as the variance in gene expression was high (Table 18). Higher expression levels of seed storage protein genes is in disagreement with observations in *Lotus japonicus*, were the quantity of storage

proteins decreases in pods during maturation (Nautrup-Pedersen et al., 2010). Furthermore, it indicates that the prominent pod growth observed here during transition from immature to intermediate ripening stage is probably not related to an increase in storage proteins.

Rhythmic diurnal expression of core clock genes in maturing pods and seeds indicates the existence of a circadian system with possible output pathways like those related to carbohydrate or protein metabolism. Examples include the circadian regulation of transient starch in leaves (Zeeman et al., 2007), circadian regulation of genes involved in carbohydrate metabolism related to photosynthesis and lipid synthesis in soybean seeds (Hudson, 2010) or the diurnal changes in protein levels in rice leaves (Wang & Wang, 2011). There are no reports on diurnal changes in the expression of genes encoding storage proteins. In cowpea seeds of intermediate maturity, diurnal expression patterns were similar for the three storage proteins with a significant diurnal rhythm and peak expression at T12 for all proteins. In contrast, mature seeds displayed a significant rhythmicity only for *LegJ* at T12. The daily changes in light availability as affected by pod wall and seed coat transmittances may explain this differences, as light perception and photosynthesis creates circadian rhythms in plants (Allen et al., 2009). In pods, diurnal differences in the expression levels were not as pronounced as in seeds and significance of diurnal rhythmicity for a protein changed during development and phasing occurred mostly towards midday. Legume seed protein gene expression was shown to be metabolically regulated through changes in osmotic pressure or soluble sugar concentrations (Wobus et al., 1995) and these factors may contribute to changes observed here. It remains to be determined if the identified diel pattern is a direct effect of the circadian clock or the result of changes in general metabolism.

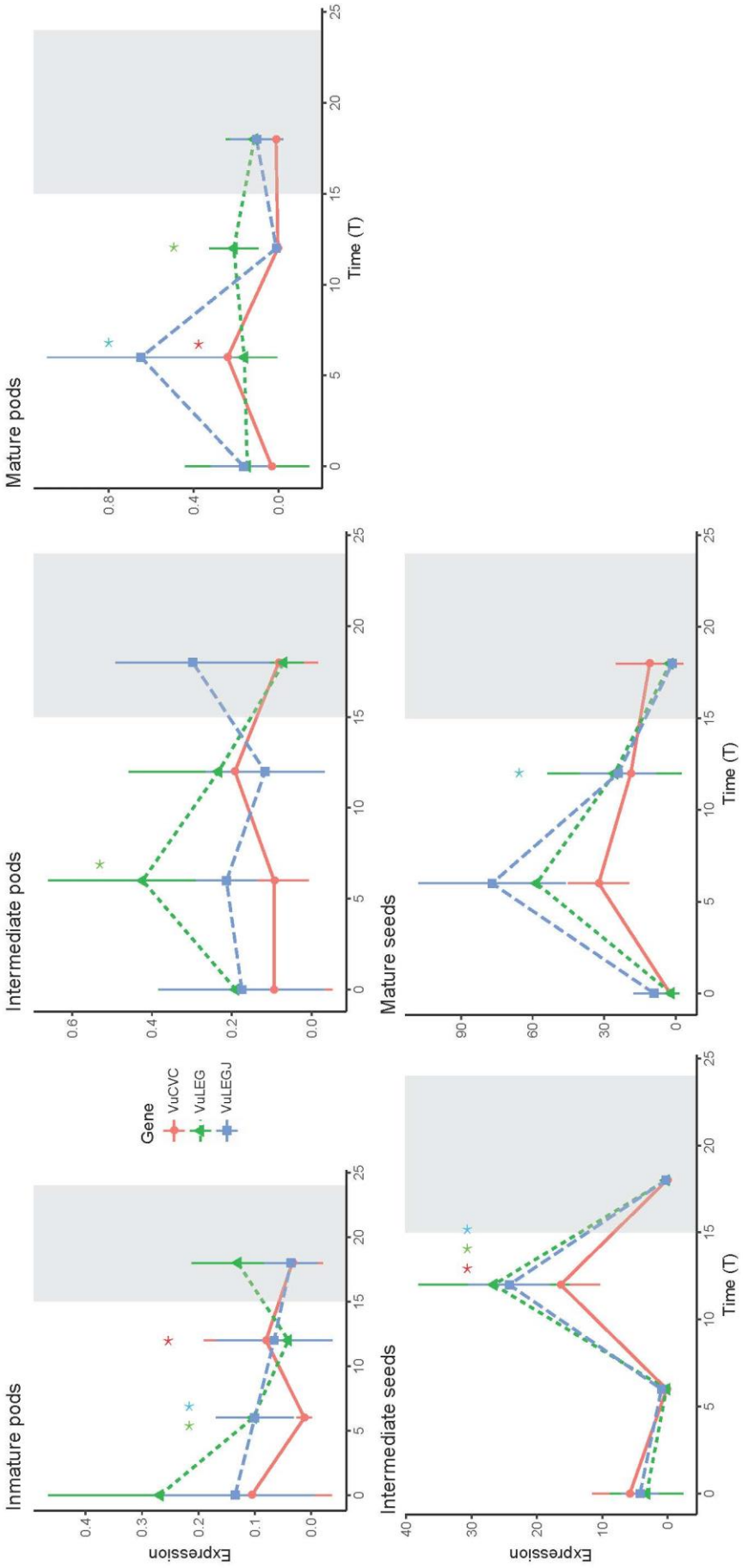


Figure 17. Expression of the protein storage genes *VuCVC*, *VuLEG* and *VuLEGJ* in cowpea pods and seeds. Expression represents the normalized expression NE according to the formula (NE) = 2^-(Ct experimental – Ctn). Collection points represent Time 6, 12, 18 and 24 hours after dawn. Four samples were analyzed for each time point and error bars indicate the standard deviation. A significant rhythm in the time series according to RAIN (Rhythmicity Analysis Incorporating Nonparametric methods) is marked with an asterisk in position of peak phase. Error bars indicate standard deviations.

3.4. Acknowledgments

This study was supported by EUROLEGUME project. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement nº 613781. We would like to thank Timothy J. Close for seeds supply.

GENERAL CONCLUSIONS

GENERAL CONCLUSIONS

- **CHAPTER 1**

The results indicate the existence of significant interactions among genotypes, locations and years, providing a useful knowledge about the breeding value of the genetic resources studied. INIAV could be the best place to grow these accessions because of the highest yield obtained. However, if we are looking for precocity, Cartagena and Vila Real are the most suitable places. Cartagena was the place where the filling of the seed was the fastest, probably due to the higher temperatures and radiation. The thermal time model (EDD) could be used to predict the period of cowpea development, therefore predict flowering and pod maturity dates, an important issue in harvest logistic and marketing strategies.

- **CHAPTER 2**

We found significant association of several SNPs to morphological traits such as: days to flowering, pod yield and protein content for *P. sativum* grown in Latvia and days to flowering, pod yield and protein content, pod length and plant height for *V. faba* grown in Spain. Our results thus demonstrate the feasibility of NGS-based genotyping using multiple barcoded samples.

- **CHAPTER 3**

During development, the circadian clock is reset between leaves and late pod development. Our results show that storage protein deposition may be circadian regulated. Our results emphasize on the importance of coordinating sampling time for comparative expression analysis of storage protein genes, for example when evaluating levels of protein storage gene expression as marker tor protein content in cowpea varieties.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Origin of *Pisum sativum* and *Vicia faba* accessions utilized for genotyping.

Nº	Origin	*Pisum sativum*	Origin	*Vicia faba*
1	Spain	BGE001453	Spain	BGE011663
2	Spain	BGE027120	Spain	BGE038412
3	Spain	BGE001564	Spain	BGE038411
4	Spain	BGE022214*	Spain	BGE011660
5	Spain	BGE001121	Spain	BGE013482
6	Spain	BGE043922	Spain	BGE013480
7	Latvia	Lasma**	Spain	BGE011631
8	Spain	BGE035455	Spain	BGE011694
9	Latvia	k-5738**	Spain	CM 10-00543
10	Spain	BGE032226	Spain	BGE011703*
11	Estonia	Ami	Spain	BGE013463
12	Spain	BGE023488	Spain	BGE010837
13	Spain	BGE025267	Spain	BGE001671
14	Spain	BGE026433	Spain	BGE011622
15	Latvia	k-4826	Spain	BGE011669
16	Latvia	k-4903**	Spain	BGE001544
17	Spain	BGE029026*	Spain	BGE011649
18	Spain	BGE026432	Spain	BGE011692
19	Estonia	Virges	Spain	Vc41-UMU.S
20	Latvia	k-6973	Spain	Vc44-UMU.S
21	Latvia	k-6968	Spain	BGE013516
22	Estonia	Kirke	Spain	BGE010839
23	Spain	BGE026431	Spain	BGE013450
24	Spain	CM 13-00785	Spain	Vc39-UMU.S
25	Spain	BGE038866	Spain	BGE011665
26	Spain	BGE027123	Spain	BGE013464
27	Latvia	k-4172**	Spain	BGE010835
28	Latvia	k-4821**	Spain	Vc42-UMU.S
29	Spain	BGE030159	Spain	BGE013487
30	Estonia	Looming	Spain	BGE026424
31	Spain	BGE029019	Spain	BGE035490
32	Spain	BGE033608*	Spain	BGE010812
33	Spain	BGE035451	Spain	BGE038409
34	Estonia	Valma	Spain	BGE010745
35	Spain	BGE029017	Spain	BGE013477
36	Spain	BGE025265	Spain	BGE035506
37	Latvia	k-5700	Spain	Vc12-UMU.S
38	Latvia	Bayava	Spain	BGE027149
39	Spain	BGE027118*	Latvia	BGE001633*
40	Spain	BGE023644*	Latvia	BGE013525*
41	Spain	BGE027117*	Spain	BGE001670*
42	Spain	BGE033609	Estonia	BGE027144*
43	Spain	BGE027115*	Latvia	BGE001038*
44	Spain	BGE033610*	Spain	BGE010764*
45	Spain	BGE001122*	Latvia	BGE010744*
46	Portugal	Gp3263	Latvia	Fb2939
47	Portugal	Gp3491	Spain	Fb395
48	Portugal	Gp3497	Estonia	Favel
49	Portugal	Grisel	Latvia	Fb3293
50	Latvia	k-4829, priekulu L-91	Spain	Priekulu vietejas
51	Latvia	k-6975	Spain	VF_014

52	Latvia	k-6389	Spain	Valmiera
53	Latvia	k-4824	Spain	VF_001
54	Latvia	k-5698	Latvia	Priekulu 32
55	Latvia	Aina	Latvia	VF_004
56	Latvia	k-6967	Latvia	Bauska
57	Latvia	k-4823, Priekulu linija, Priekulu peluskis	Portugal	Lielplatones
58	Latvia	k-6969	Portugal	VF_005
59	Latvia	Capulos**	Portugal	Tolea
60	Latvia	k-6972**	Latvia	VF_003**
61	Latvia	k-4171**	Latvia	VF_002**
62	Latvia	k-4827**	Latvia	VF_008**
63	Latvia	k-4825**	Latvia	VF_009**
64	Latvia	Balva**	Portugal	VF_007
65	Latvia	k-4174, Raunas peluskis**	Latvia	Polli**
66	Latvia	k-4828**	Latvia	Jogeva**
67	Latvia	Alma**	Spain	Vc3-UMU.S
68	Latvia	Retrija**	Spain	BGE011645
69	Latvia	Rota**	Spain	BGE011687
70	Latvia	Vitra**	Spain	Vc16-UMU.S
71	Latvia	Selga**	Spain	Vc37-UMU.S
72	Latvia	Zaiga**	Spain	Vc34-UMU.S

*Accessions phenotyped in Spain; ** Accessions phenotyped in Latvia

Supplementary Table 2. SNP primer sequence, primer melting temperature (Tm) and PCR product length, utilized in *Pisum sativum* and *Vicia faba*.

Pisum sativum					
SNP number	**SNP id[a]**	**Forward Primer**	**Reverse Primer**	**Tm (°C)**	**Product length (bp)**
1	SNP_100000443	TCCAGGTGCTCAACA	AGGACCACCAGACCA	49.4/49.8	135
2	SNP_100000341	CCCGAACCATCTGTT	GATGCCGGAGCTTAC	49.8/49.9	110
3	SNP_100000597	TGTGCTTCCAATTAACC	TTGGTTTGCAATGCTAT	49.8/50.5	106
4	SNP_100000346	GTGAATCTCCGTTGGAT	GAGGCTGAAGATCAGGT	51.0/50.3	100
5	SNP_100000414	CTTAACAATGGCACGAT	GATGAAGGGGCTCTTAC	49.8/50.0	115
6	SNP_100000460	CCTGAAGCCAATGAAG	ATGCTGACCCAAGAAA	50.1/49.8	119
7	SNP_100000648	CAGTCCCCAAACCTG	TGTGATTCCAGGAGGA	50.3/50.4	108
8	SNP_100000616	TGCAATACACTTGCATAAT	TCACGGGTTTTAGCC	49.9/50.5	116
9	SNP_100000543	TAACGAGCGCAAAAG	GTGGGTTCTTGTCAGC	49.4/49.5	117
10	SNP_100000664	AGGCGAAATTGAAGC	ACTCTAACACCCCTTTCA	49.8/50.1	114
11	SNP_100000360	AAGGCTAGGAGCTAATCA	CCCGGAACTTGGAG	50.0/50.4	116
12	SNP_100000659	AGGGAGAGCCAGAAAG	CGTTATCCCAAAATGC	50.5/50.2	114
13	SNP_100000091	CCTCACTCAATGACCAG	TGCTCATGGCAAGAC	49.6/49.1	135
14	SNP_100000095	GATCATGGAGCAGATGA	CCATGGTGAATGAATTT	50.0/49.2	102
15	SNP_100000288	CATTGATGGTCTTTTTCA	GGCCATAACTTTGGAG	49.7/49.0	101
16	SNP_100000395	ATTCACCCCACCTGA	AGCTGCACCACCAGT	49.9/50.5	158
17	SNP_100000035	GTGCGCTCGCTTTA	ACCACTTTCGGTGTTG	50.1/50.1	135
18	SNP_100000786	GACCACACGGAAACC	TCGTGCTCTTCGTGA	50.0/49.9	113
19	SNP_100000282	CACTGCCAAGTCTTCTG	ACTGATGGAGGAGAAGG	50.2/50.1	118
20	SNP_100000071	GGGACGTACTGGATCA	GATTCAAGCAGGGTGA	49.8/50.2	130
21	SNP_100000171	CCAGAAGCTCTGATGC	TCCATCCTCCAAAGC	50.1/50.0	138
22	SNP_100000408	CACGGACCTCAAACC	CATTCAGCCCATCCT	50.3/49.8	130

23	SNP_100000264	GCGATGCAACAAACTAT	AGACACGCTGGTAGTGA	50.0/50.0	120
24	SNP_100000822	CGAAAAACTCTCCATCA	TGAATGCTTCTTTTGA	49.7/50.0	120
25	SNP_100000042	TAATGTGACCCATGAGC	AGAATGGCTTGTTGTGA	50.3/50.0	106
26	SNP_100000485	GGAATTCCTGCCAAA	CCAGGCTGTGGTGA	50.2/49.6	121
27	SNP_100000746	ATTGGGAATCAAATCCT	CATGTACATCGGGGTTA	40.6/50.3	102
28	SNP_100000584	CCTTGATTGGCTGGT	GTGCTCATGCTATTGCT	50.0/50.0	124
29	SNP_100000736	GGGTGCGTAGTCGAT	ACAAGGTGCCACAAAC	50.0/50.1	113
30	SNP_100000302	GCAAGAATACCAGAGCA	TTGGCTGCTAGAGATTG	50.0/50.2	119
31	SNP_100000639	GTGTCCGAGGAAACAA	TGTCATGCCAGTTGC	50.1/50.1	132
32	SNP_100000350	GAGGAGGAGGAAGACAA	AACCAAAACCAAACAAA	50.2/49.3	102
33	SNP_100000700	GGCTTTTGACGTTGTTA	CCAAATTCATACAAGGTG	50.1/49.2	99
34	SNP_100000526	TTTCTGCCAGTTTCTTG	TTAACGAAGCATTGGAC	50.1/50.0	115
35	SNP_100000702	TCTCCCCAGGTACAAA	CCTGATGGTCCTCTCA	49.7/49.8	149
36	SNP_100000847	TGGCGTAACGAACAC	TGACCTAAACGACGC	50.5/49.8	100
37	SNP_100000655	AGCGTGATCCCTGAG	CGTTACAGGGCGATT	50.0/50.5	127
38	SNP_100000058	GAAGCTTTGCCACCT	GGTGCCTCCTCTTCA	49.5/50.0	132
39	SNP_100000328	CGCAAGTACCAGAAGAG	GAAGCGGAGATCAGTCT	49.6/50.5	100
40	SNP_100000175	TGCTGGGAAAGAGTATG	TGAAAAACAGCACCAGT	49.8/50.2	103
41	SNP_100000255	TGATCCAACGATCCA	TGTGGACGCTACGAA	49.1/50.0	138
42	SNP_100000228	ATCGATGTTCCCAATTT	TCAATTCCATCGTTTGT	50.7/49.9	120
43	SNP_100000045	CTTCCGAGAAGGAGAAT	GGTGACTCACTGCAAGA	49.7/50.3	118

Vicia faba

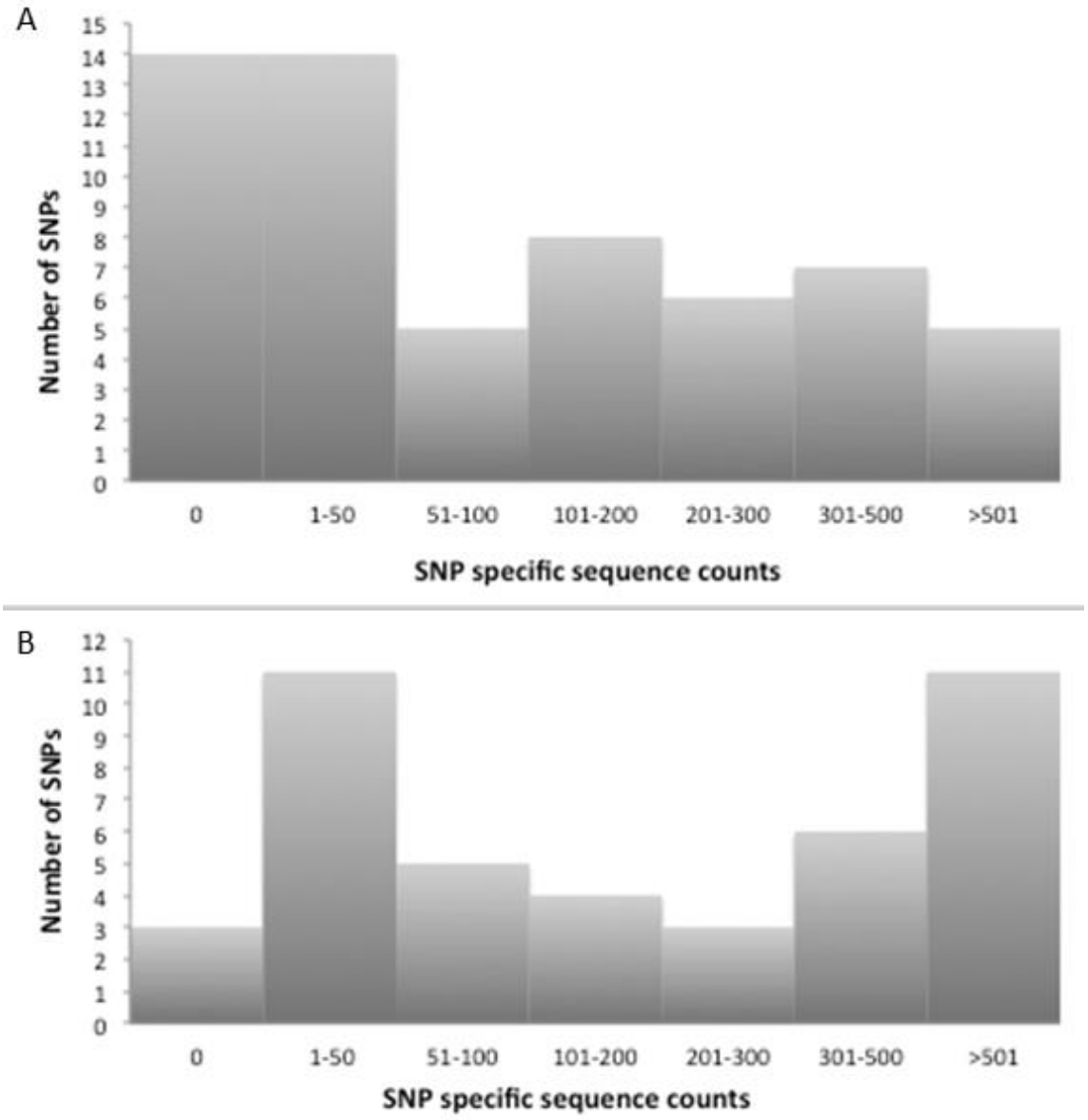
SNP number	SNP id[a]	Forward Primer	Reverse Primer	Tm (°C)	Product length (bp)
1	SNP_50000201	TGTCATGGAAAGCACA	AAGCATTGAAGCCTGA	49.7/50.3	112
2	SNP_50001182	CCCTTTTCTCTCTCAAAA	ACAATAGCCACAACCAA	50.2/50.1	108
3	SNP_50002158	TCATTTGGATAATCCTTTC	ATGGGAACGTTGTGAG	49.9/50.0	121
4	SNP_50002306	GTGGTGGCCAAATCT	ATCAAATGAGCCTTGC	49.7/49.5	124
5	SNP_50002318	GCAGCGTTGATTATGTT	CCAATGCCTCAGTCC	50.0/50.4	158
6	SNP_50000197	TGCTTGCTCATACGC	TGGGCCAATATGATTC	49.9/50.1	114
7	SNP_50000487	CAGAACCAGTGGCAGT	AAGGTGGAGGAAAAATG	50.1/50.5	141
8	SNP_50000557	CTCCATTGAGCAGCA	TGGTCAAAAGCACCA	49.4/49.7	147
9	SNP_50001252	CCGTGGATTCCTCAC	TTGGCAACAATGTCAG	50.1/49.3	119
10	SNP_50000760	TTACCATGGGCCTCT	TGGTATCAATTCCCTTG	49.5/49.4	111
11	SNP_50000819	TTCCATGGCTGATCTC	CCAGGAGCATCTTGTT	50.0/49.3	100
12	SNP_50000432	GCTGGGACAGACTCC	CCTGAACATCTGACTCG	49.0/49.8	143
13	SNP_50001365	TTGGCCTTGATAGCC	GGGTTTGCATTTCGT	50.5/50.2	145
14	SNP_50001987	GGTTTTGTGGGCATT	ACCAGCATTAGCATGTC	50.0/49.5	143
15	SNP_50002207	CCCCACTGTCGTTTT	GCATGTTGTGGCAAA	49.9/50.3	104
16	SNP_50000307	GAAGTTGCGGAAAGC	TGGAGCAGCAAAGTG	50.3/49.6	120
17	SNP_50001872	GGTGTGAAGGTTGGAA	TGGCCTTGAAGAGTGT	49.9/49.8	118
18	SNP_50001826	CACCTTCCTCTCACGA	CCCAGGCTTTGTAGG	50.2/49.9	102
19	SNP_50001916	CAGTAGCTTCCATACCG	AGCTCTGCTGTGGAGA	49.5/50.0	122
20	SNP_50000883	TCTTTGCTGGAACCAT	CCCATCATGCCATAAC	49.8/50.5	116
21	SNP_50000057	GGGTCTTCTCGACCTT	ACCGACAATTTTACCG	50.0/49.6	152
22	SNP_50000069	AAAATGGCTGAAACAAA	CCAGCTCTTCCACAAT	49.6/49.5	119
23	SNP_50000225	GCAGATGAGAGGTGGA	TGGCTATCCAAGAAGC	49.9/49.8	152
24	SNP_50000911	GGCAGCAAACACCTT	TGCGAAGAGGTTGATT	50.4/50.1	146
25	SNP_50001679	GCTTCTGGCAGTTAGG	TGCCTTTGATCTCTGC	49.4/50.7	105
26	SNP_50002062	GGTACTCCTGCCATTG	GCATCTTCTCCCATGA	49.7/50.0	102

27	SNP_50000965	CAGTTGTGTGATGAGCA	CGAAGCAACGAAAAA	49.3/49.2	150
28	SNP_50001040	CCGGAAGAACAGTGG	TGGGAATCCTTGGAA	50.6/50.4	113
29	SNP_50001146	GAGAAGTTGCAGCAAGA	ATGGAGCCAGTGGAA	50.1/50.4	123
30	SNP_50000089	TGCATTCAGCAAAGC	TGACAGTGGCAGTGAA	49.8/49.6	108
31	SNP_50000217	GCATTGTGGGAGAAGA	GTGATCACCCTTCACAA	50.2/49.9	145
32	SNP_50000310	TTGTAGCTGAGCCTGTT	CTCCATATGCCAGTGAC	49.6/49.7	106
33	SNP_50000356	TTTGCTCATCCTATCACA	TCATAGGAGCCTTTGATT	50.3/50.4	114
34	SNP_50000332	TTCCCTTCTGCTTCTG	GACAGGGAAACACGAA	49.7/50.1	103
35	SNP_50000787	ACCCTGTCCACAAGC	GAGGGACGGTTCTTG	49.7/49.4	104
36	SNP_50000084	CAAAAACAAATGGTAGGA	CGACAATGTCCGTCA	49.1/49.9	149
37	SNP_50000308	TTACCAGATCGCCTGT	TGGACAGATCGGAAAC	49.9/50.0	106
38	SNP_50000440	AACTGAGCTTTCGAACAT	CTTAAATTCCGCAAGG	50.5/49.7	140
39	SNP_50001023	AGAGCCAACACCATTG	GATTCCCTCACTCCAAC	50.2/50.4	106
40	SNP_50002450	GGCCGATACTTCTGC	AGGAGTCCCCAAACC	49.9/50.5	120
41	SNP_50000764	CACCTTCCTTTTGGTG	GGAGGAGTTGAAGAAGTG	49.8/50.1	109
42	SNP_50001532	TAAGCCTCGGTAATGG	CCAAGGAGAGGAACAA	49.9/49.3	107
43	SNP_50001531	CTGGCAACATTCATCA	TGGAGACCTTTTCTGC	49.2/49.4	143
44	SNP_50001647	TGGAGTTGTTCAGGGTA	CAAACCACACGAGCA	49.8/50.3	143
45	SNP_50001880	TGGAGAACCATTACCAA	ATAGGCCAACCTAATGG	49.9/50.6	148
46	SNP_50000181	AACTCCATCGGCAAC	CCAGGAAGCTGTGTGT	50.7/50.1	140
47	SNP_50000203	CAGAAAACTGCATAGCC	GCAGCAGTCCTTGGT	49.7/50.2	119
48	SNP_50000402	GGGTTTCCAAATCTCC	AACTCCATCGGCAAC	50.4/50.0	109
49	SNP_50000824	CAAAAAGACTCCAGTGC	GGATCAGCAGGAAGC	49.3/49.6	112
50	SNP_50001586	TCCCAGGTGATTTCAT	TCTAGGATAAGGGACCAA	49.2/50.3	100
51	SNP_50000208	GAGGAATGAAAGGCTGT	CTGAGGACGCTTGAAT	50.5/49.5	125
52	SNP_50001889	CCCTGGCTTCTCAAG	AGCTTCAGCGATGACT	49.9/49.1	145
53	SNP_50002378	TCCAGCTCTGTTACCC	CATGGGCAAATGAAGT	49.3/50.2	102
54	SNP_50001498	TTAACCAACGCCAAA	CGTCCAGACGTGGAT	49.3/50.8	151
55	SNP_50002155	CCCTCAGTCGGAGAA	ATACCCGGATGTTGG	50.0/49.7	155
56	SNP_50001725	CCATGAAGCCACTTTC	TGCCTTCTCCTTCTCA	49.8/50.0	102
57	SNP_50000436	TTTCACAATGGGAACAT	GCCATTCCAGAAACC	49.6/49.4	141
58	SNP_50001725	TGGGATCTGACAAGGA	CCCACCTGGGTTGTA	50.4/50.8	120
59	SNP_50002190	AGCATTTGCGATTACC	TTGCCAGATCGTTTG	49.8/49.4	121

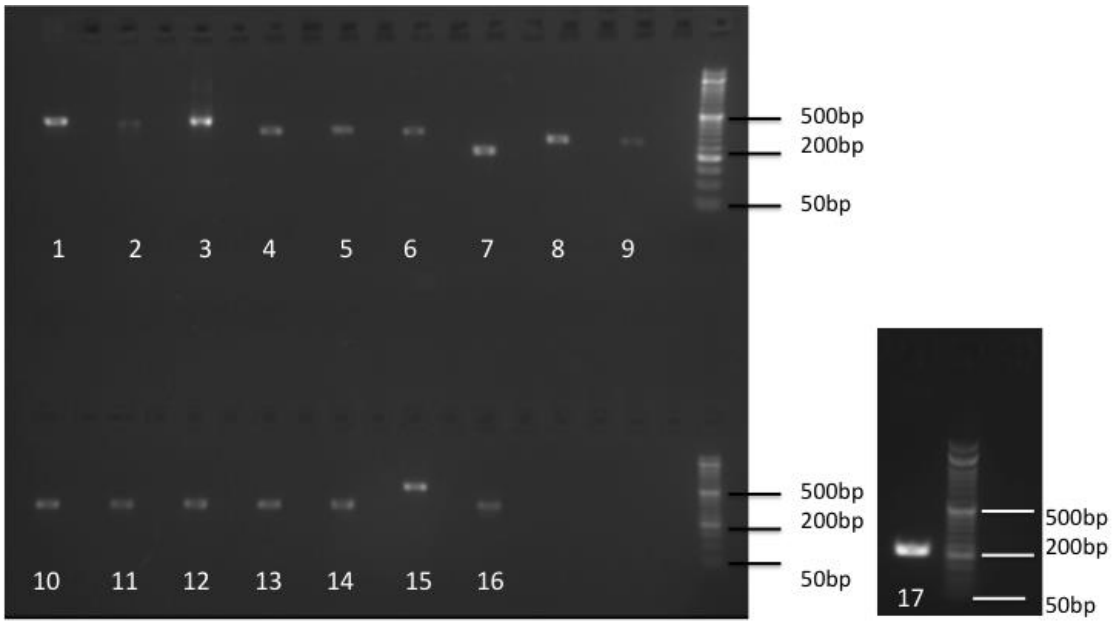
[a]SNP id according to Leonforte et al., 2013 and Kaur et al., 2014

Supplementary Table 3. List of analysed reference genes, clock genes and protein storage genes.

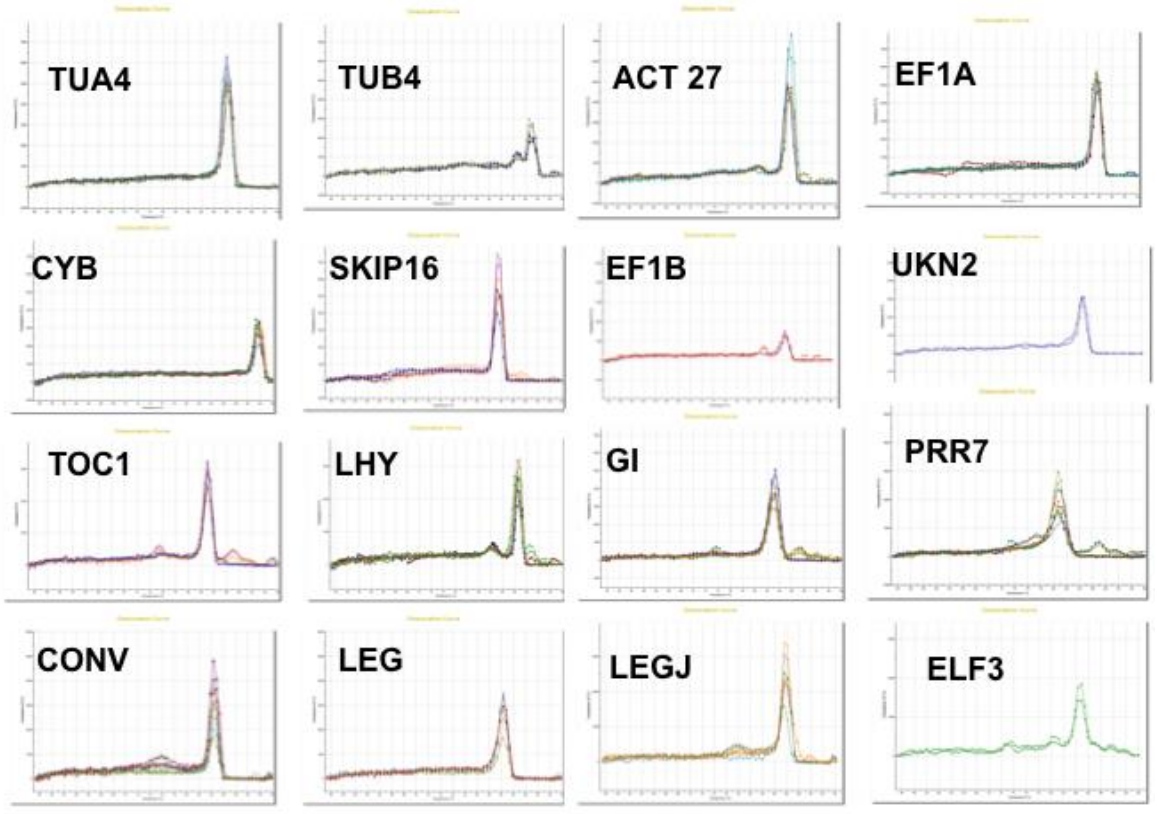
Name	Short name	NCBI accession number Query	Accession number *V. unguiculata* (Noble VuGEA)	Amplicon size EST (bp)	Forward Primer sequence	Reverse Primer sequence
B-Actin	Act	FR839671	FR839671	295	TTCGTCTTGATT TGGCTGGT	TGCCTACTGCTTC CATTCCT
Actin 2/7	Act27	XP_006578804	Vun_T14730.1	295	GAGCACCCTGTT CTCCTCAC	TAGCCACGCTCGG TTAAGAT
Cyclophylin	Cyp	NP_001235223	Vun_T05892.1	294	GCAAACTTCGA GCCGTAAAT	AATTCAATTGCAA CGCAACA
Elongation factor 1-A	EF1A	XP_003553292	Vun_T23609.1	300	GCCTGGTATGG TGGTGACTT	GCGAACTTCACTG CAATGTG
Elongation factor 1-B	EF1B	XP_006588020	Vun_T21512.1	300	ATCTCACCCCTT GAGCATTG	TCTGGGTTTTCAG GGATCAG
Alpha tubulin	Tua4	XP_006606007	Vun_T07354.1	290	GTTTGATGGTG CATTGAACG	TGCGCTTGGTCTT GATAATG
Beta tubulin	Tub4	XP_003554108	Vun_T13531.1	193	TATCAGAAGGG TTCCCATGC	TAACTTCGTCTTC GGGCAGT
ASK-interacting protein 16	Skip16	NP_001242370	Vun_T09332.1	296	ACAGCCGTTGA ACAAAAAGG	GTGGCTTCTTCGT CCACACT
Hypothetical protein unknown	Ukn2	XP_003528257	Vun_T00647.1	297	TGAAGAAATTG GCCTTTGGA	AGCAATCGGCTGC ACTATCT
Convicilin	CVC	CAB82855	Vun_T03561.1	295	CTCCAACACGTT GGGAAGAT	AGCGTTGATGGCT ACTGGAT
Legumine B	LEGB	XP_003539704	Vun_T10886.2	300	TTCCCTTGGTTG GTGATGTT	TTCCAGCCAAGTG GGTAAAT
Legumine J	LEGJ	XP_003520601	Vun_T05146.1	264	AGAGGACAACA TACTGGGCACT	GCCACGAGAAGA GTTGGAAA
Early Flowering 3	ELF3	ABP81864		217	GCGTCTGTAGA GCTCGCTTC	GCTGCTACAGGAC CTTGCTT
Timing of CAB Expression1	TOC1	Q9LKL2	Vun_T21607.1	296	ACTTACCCAGCA GCAGCAGT	AGGAGGACATTG GGGAAGAC
Late Elongates Hypocotyl	LHY	Q6R0H1	Vun_T17583.1	292	ACACAACACAA CCAGGACG	CACACAAAGGAA GCAGTCCA



Supplementary Figure 1. Frequency distribution of the number of SNP specific sequences identified in (A) *V. faba* and (B) *P. sativum*. A total of 56 and 43 SNPs were analyzed for (A) and (B), respectively.



Supplementary Figure 2. (1) Conv (2) Act (3) Vic (4) Cyp 5) EF1A (6) EF1B (7) GI (8) LegJ (9) Leg(10) LHY (11) Skip16 (12) Toc1 (13) Tua4 (14) Tub4 (15) Act27 (16) Ukn2 (17) ELF3.



Supplementary Figure 3. Dissociation curve of the genes applied in this study for selected tissues

REFERENCES

REFERENCES

Adewale BD, Okonji C, Oyekanmi AA, Akintobi DAC & Aremu CO (2010) Genotypic variability and stability of some grain yield components of Cowpea. *African Journal of Agricultural Research* **5**, 874-880. http://doi.org/10.5897/AJAR09.481

Ahmad S, Kaur S, Lamb-Palmer ND, Lefsrud M & Singh J (2015) Genetic diversity and population structure of Pisum sativum accessions for marker-trait association of lipid content. *The Crop Journal*, **3**, 238–245. http://doi.org/10.1016/j.cj.2015.03.005

Akande SR (2007) Genotype by environment interaction for cowpea seed yield and disease reactions in the forest and derived savanna agro-ecologies of south-west Nigeria. *American-Eurasian Journal of Agricultural & Environmental Science* **2**, 163-168. ISSN 1818-6769.

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P & Kay SA (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science*, **293**, 880–883. http://doi.org/10.1126/science.1061320293/5531/880 [pii]

Allen ON & Allen EK (1981) *The Leguminosae, a source book of characteristics, uses, and nodulation*. Univ of Wisconsin Press. ISBN 0-333-32221-5

Allen DK, Ohlrogge, JB & Shachar-Hill Y (2009) The role of light in soybean seed filling metabolism. *The Plant Journal : For Cell and Molecular Biology*, **58**, 220–34. http://doi.org/10.1111/j.1365-313X.2008.03771.x

Alves-Carvalho S, Aubert G, Carrère S, Cruaud C, Brochot AL, Jacquin F, ... & Silva C (2015) Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *The Plant Journal*, **84**, 1-19. http://doi.org/10.1111/tpj.12967

Amede T, Schubert S & Stahr K (2003) Mechanisms of drought resistance in grain legumes I: Osmotic adjustment. *SINET: Ethiopian Journal of Science*, **26**, 37-46. ISSN: 0379–2897.

Andersen CL, Jensen JL & Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to

identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, **64**, 5245–5250. Journal Article. http://doi.org/10.1158/0008-5472.CAN-04-0496

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP & Zondervan KT (2010) Data quality control in genetic case-control association studies. *Nature Protocols*, **5**, 1564–73. http://doi.org/10.1038/nprot.2010.116

Ashraf M (2010) Inducing drought tolerance in plants: Recent advances. *Biotechnology Advances*, **28**, 169–183. http://doi.org/10.1016/j.biotechadv.2009.11.005

AOAC (1990) In "Official methods of analysis". (Eds Helrich K.), 15th edition. (Arlington, VA, USA).

Awolumate EO (1983) Accumulation and quality of storage protein in developing cowpea, mung bean and soya bean seeds. *Journal of the Science of Food and Agriculture*, **34**, 1351–1357. http://doi.org/10.1002/jsfa.2740341206

Aykroyd WR & Doughty J (1982) *Legumes in human nutrition* (Vol. 20). FAO. Rome. ISBN 92-5-101181-8

Barrera-Figueroa BE, Gao L, Diop NN, Wu Z, Ehlers JD, Roberts PA, ... & Liu R (2011) Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. *BMC plant biology*, **11**, 127. https://doi.org/10.1186/1471-2229-11-127

Beach LR, Spencer D, Randall PJ & Higgins TJV (1985) Transcriptional and post-transcriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. *Nucleic acids research*, **13**, 999-1013. https://doi.org/10.1093/nar/13.3.999

Bendix C, Marshall CM & Harmon FG (2015) Circadian clock genes universally control key agricultural traits. *Molecular plant*, **8**, 1135-1152. https://doi.org/10.1016/j.molp.2015.03.003

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, … Udvardi MK (2008) A gene expression atlas of the model legume Medicago truncatula. *The Plant Journal : For Cell and Molecular Biology*, **55**, 504–13. http://doi.org/10.1111/j.1365-313X.2008.03519.x

Birney E, Clamp M & Durbin R (2004) GeneWise and Genomewise. *Genome Research*, **14**, 988–95. http://doi.org/10.1101/gr.1865504

Blanco FM (1877) Flora de Filipinas. (Eds Naves FA) (de Plana y C[a]). Manila.

http://bibdigital.rjb.csic.es/ing/Libro.php?Libro=14

Bogdanova VS & Berdnikov VA (2000). A study of potential ability for cross-pollination in pea originating from different parts of the world. *Pisum Genetics*, *32*, 16–17. Retrieved from http://hermes.bionet.nsc.ru/pg

Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, … & Kuhlemeier C (2016). Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nature Plants*, *2*(6), 16074. http://doi.org/10.1038/nplants.2016.74

Bonhomme R (2000) Bases and limits to using 'degree. day'units. *European journal of agronomy* **13**, 1-10. https://doi.org/10.1016/S1161-0301(00)00058-7

Bordage S, Sullivan S, Laird J, Millar AJ & Nimmo HG (2016). Organ specificity in the plant circadian system is explained by different light inputs to the shoot and root clocks. *New Phytologist*, **212**, 136–149. http://doi.org/10.1111/nph.14024

Box MS, Coustham V, Dean C & Mylne JS (2011) Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from Arabidopsis. *Plant Methods*, **7**, 7. http://doi.org/10.1186/1746-4811-7-7

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y & Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635. http://doi.org/10.1093/bioinformatics/btm308

Bremner JM (1960) Determination of nitrogen in soil by the Kjeldahl method. *The Journal of Agricultural Science*, **55**, 11. http://doi.org/10.1017/S0021859600021572

Campbell L, Euston SR & Ahmed MA (2016) Effect of addition of thermally modified cowpea protein on sensory acceptability and textural properties of wheat bread and sponge cake. *Food Chemistry* **194**, 1230-1237. https://doi.org/10.1016/j.foodchem.2015.09.002

Casey R, Christou P, Domoney C, Hedley C, Hitchin E, Parker M, … & Zasiura C (2001) Expression of legumin and vicilin genes in pea mutants and the production of legumin in transgenic plants. *Die Nahrung*, **45**, 385–7. http://doi.org/10.1002/1521-3803(20011001)45:6<385::AID-FOOD385>3.0.CO;2-K

Chai C, Wang Y, Joshi T, Valliyodan B, Prince S, Michel L, … & Nguyen HT (2015)

Soybean transcription factor ORFeome associated with drought resistance: a valuable resource to accelerate research on abiotic stress resistance. *BMC Genomics*, **16**, 596. http://doi.org/10.1186/s12864-015-1743-6

Chaves MM, Maroco JP & Pereira JS (2003) Understanding plant responses to drought: from genes to the whole plant, *Functional Plant Biology,* **30**, 239–264. https://doi.org/10.1071/FP02076

Chung JW, Kim TS, Suresh S, Lee SY & Cho GT (2013) Development of 65 novel polymorphic cDNA-SSR markers in common vetch (*Vicia sativa* subsp. *sativa*) using next generation sequencing. *Molecules*, **18**, 8376-8392. https://doi.org/10.3390/molecules18078376

Collado E, Klug TV, Martínez-Sánchez A, Artés-Hernandez F, Aguayo E, Artés F, Fernández JA & Gómez PA (2017) Immature Pea Seeds. Effect of Storage Under Modified Atmosphere Packaging and Sanitation with Acidified Sodium Chlorite. *Journal of the Science of Food and Agriculture*. http://doi.org/10.1002/jsfa.8513

Cousin R (1997) Peas (Pisum sativum L.). *Field Crops Research*, **53**, 111-130.

Craufurd PQ, Summerfield RJ, Ellis RH & Roberts EH (1997) Photoperiod, temperature, and the growth and development of cowpea. In "Advances in Cowpea Research". (Eds B.B. Singh, D.R. Mohan Raj, K.E. Dashiell and L.E.N. Jackai) pp. 75-86. (IITA, Ibadan and JIRCAS, Tsukuba). ISBN 978131 110 X

Cubero JI (2011) The faba bean: a historic perspective. *Grain Legumes*, **56**, 5–7.

Davis DW, Oelke EA, Oplinger ES, Doll JD, Hanson, CV & Putnam DH (1991). Cowpea. *Alternative Field Crops Manual; University of Wisconsin-Extension*.

de Montaigu A, Toth R, Coupland G & Tóth R (2010) Plant development goes like clockwork. *Trends Genet*, **26**, 296–306. http://doi.org/S0168-9525(10)00084-3 [pii]10.1016/j.tig.2010.04.003

De Pace C, Delre V, Scarascia Mugnozza, GT, Maggini F, Cremonini R, Frediani M & Cionini PG (1991) Legumin of *Vicia faba* major: accumulation in developing cotyledons, purification, mRNA characterization and chromosomal location of coding genes. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, **83**, 17–23. http://doi.org/10.1007/BF00229221

Derbyshire E, Wright DJ, & Boulter D (1976) Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry*, **15**, 3-24. https://doi.org/10.1016/S0031-

9422(00)89046-9

Deschamps S, Llaca V & May GD (2012) Genotyping-by-Sequencing in Plants. *Biology*, **1**, 460–483. http://doi.org/10.3390/biology1030460

Domínguez-Perles R, Carnide V, Marques G, de Castro I, de Matos, Carvalho M & Rosa E, (2015) Relevance, constraints and perspectives of cowpea crops in the Mediterranean Basin. *Legume Perspectives* **10**, 40-42. ISSN 2340-1559

Doumbia IZ, Akromah R & Asibuo JY (2013) Comparative study of cowpea germplasms diversity from Ghana and Mali using morphological characteristics. *Journal of Plant Breeding and Genetics* **1**, 139-147. ISSN: 2305-297X

Drabo I, Redden R, Smithson JB & Aggarwal VD (1984) Inheritance of seed size in cowpea (*Vigna unguiculata* (L.) Walp.). *Euphytica* **33**, 929-934. https://doi.org/10.1007/BF00021923

Duc G (1997) Faba bean (*Vicia faba* L.). *Field Crops Research*, **53**, 99-109. https://doi.org/10.1016/S0378-4290(97)00025-7

Egbadzor KF, Dadoza M, Danquah EY, Yeboah M, Offei SK & Ofori K (2013) Genetic control of seed size in cowpea (*Vigna unguiculata* (L.) Walp). *International Journal of Agriculture Sciences*, **5**, 367-371. https://doi.org/10.9735/0975-3710.5.2.367-371

Egea-Cortines M, Ruíz-Ramón F, Weiss J, Ruiz-Ramon F & Weiss J (2013) Circadian Regulation of Horticultural Traits: Integration of Environmental Signals in Plants. In J. Janick (Ed.), *Horticultural Reviews* (Vol. 41, pp. 1–46). Wiley. http://doi.org/10.1002/9781118707418.ch01

Ehlers JD & Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Research* **53**, 187-204. https://doi.org/10.1016/S0378-4290(97)00031-2

Endo M (2016) Tissue-specific circadian clocks in plants. *Current Opinion in Plant Biology*, **29**, 44–49. http://doi.org/10.1016/j.pbi.2015.11.003

Endo M, Shimizu H, Nohales Ma, Araki T & Kay Sa (2014) Tissue-specific clocks in Arabidopsis show asymmetric coupling. *Nature*, **515**, 419–422. http://doi.org/10.1038/nature13919

Erskine W, Smartt J & Muehlbauer F (1994) Mimicry of lentil and the domestication of Common Vetch and Grass Pea. *Economic Botany*, **48**, 326-332. http://doi.org/10.1007/BF02862334

European Commission: Trade: Export Helpdesk: EU Customs Union. European Commission 344 2016. http://exporthelp.europa.eu/thdapp/index.htm (17 September 2016).

Falconer DS & Mackay TFC (1996) Introduction to Quantitative Genetics, Ed 4. Longmans Green, Harlow, Essex, UK. ISBN 978-0-582-24302-6

FAO (2004) *Protein sources for the animal feed industry*.

FAOStat, Food and Agriculture Organization (FAO), World Production. (On line) (2017). Available: http://faostat3.fao.org/browse/Q/QC/E [2 May 2017].

Farré EM & Weise SE (2012) The interactions between the circadian clock and primary metabolism. *Current Opinion in Plant Biology*, **15**, 293–300. http://doi.org/10.1016/j.pbi.2012.01.013

Felsenstein J (1985) Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution*, **39**, 783–791. https://doi.org/10.1093/oxfordjournals.molbev.a026203

Filichkin SA, Breton G, Priest HD, Dharmawardhana P, Jaiswal P, Fox SE, ... & Mockler TC (2011) Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PloS one*, **6**, e16907. https://doi.org/10.1371/journal.pone.0016907

Fourmont R (1956) Les variétés de pois (*Pisum sativum*) cultivés en France. INRA, Paris, 253 pp. + 70 planches.

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Coupland G, … & Putterill J (1999) GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO Journal*, **18**, 4679–4688. https://doi.org/10.1093/emboj/18.17.4679

Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E & Weber AP (2011) Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC genomics*, **12**, 227. https://doi.org/10.1186/1471-2164-12-227

Ganal MW, Altmann T & Röder MS (2009) SNP identification in crop plants. *Current Opinion in Plant Biology*, **12**, 211–217. http://doi.org/10.1016/j.pbi.2008.12.009

Garcia-Vallvé S,Palau J & Romeu A (1999) Horizontal gene transfer in glycosyl

hydrolases inferred from codon usage in Escherichia coli and Bacillus subtilis. *Molecular Biology and Evolution*, **16**, 1125–34.

Gatehouse JA, Evans IM, Bown D, Croy RRD & Boulter D (1982) Control of storage-protein synthesis during seed development in pea (Pisum sativum L.). *Biochemical Journal*, **208**, 119–127. http://doi.org/10.1042/bj2080119

Gonçalves A, Goufo P, Barros A, Domínguez-Perles R, Trindade H, Rosa EAS, Ferreira L, & Rodrigues M (2016) Cowpea (*Vigna unguiculata* L. Walp), a renewed multipurpose crop for a more sustainable agri-food system: nutritional advantages and constraints. *Journal of the Science of Food and Agriculture* **96**, 2941-2951. https://doi.org/doi: 10.1002/jsfa.7644

Graham PH & Vance CP (2003) Legumes: importance and constraints to greater use. *Plant physiology*, **131**, 872-877. https://doi.org/10.1104/pp.017004

Gupta P, Singh R, Malhotra S, Boora KS & Singal HR (2010) Characterization of seed storage proteins in high protein genotypes of cowpea [*Vigna unguiculata* (L.) Walp]. *Physiology and Molecular Biology of Plants : An International Journal of Functional Plant Biology*, **16**, 53–8. http://doi.org/10.1007/s12298-010-0007-9

Gupta S, Nadarajan N & Gupta DS (Eds.) (2014) Legumes in the omic era. *Springer*, **348**. https://doi.org/10.1016/j.jclepro.2015.05.002

Harmer SL, Hogenesch JB, Straume M Chang HS, Han B, Zhu T, … & Kay SA (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science (New York, N.Y.)*, **290**, 2110–3.

Hayes LD, Bickerstaff GF & Baker JS (2010) Interactions of cortisol, testosterone, and resistance training: influence of circadian rhythms. *Chronobiology international*, **27**, 675-705. http://dx.doi.org/10.3109/07420521003778773

He J, Zhao X, Laroche A, Lu Z-X, Liu H & Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, *5*(September), 484. http://doi.org/10.3389/fpls.2014.00484

Hecht V, Foucher F, Ferrandiz C, Macknight R, Navarro C, Morin J, … & Weller JL (2005) Conservation of Arabidopsis flowering genes in model legumes. *Plant Physiology*, **137**, 1420–1434. Journal Article. http://doi.org/Doi 10.1104/Pp.104.057018

Hecht V, Knowles CL, Vander Schoor JK, Liew LC, Jones SE, Lambert MJ & Weller JL

(2007) Pea *LATE BLOOMER1* is a *GIGANTEA* ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant physiology*, **144**, 648-661. https://doi.org/10.1104/pp.107.096818

Henning JA, Coggins J & Peterson M (2015) Simple SNP-based minimal marker genotyping for Humulus lupulus L. identification and variety validation. *BMC Research Notes*, **8**, 542. http://doi.org/10.1186/s13104-015-1492-2

Higgins TJV, Newbigin EJ, Spencer D, Llewellyn DJ & Craig S (1988) The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. *Plant Molecular Biology*, **11**, 683-695. https://doi.org/10.1007/BF00017468

Hu R, Fan C, Li H, Zhang Q & Fu Y-F (2009) Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Molecular Biology*, **10**, 93. http://doi.org/10.1186/1471-2199-10-93

Hu X, Ren J, Ren X, Huang S, Sabiel SAI, Luo M, … & Sun D (2015) Association of Agronomic Traits with SNP Markers in Durum Wheat (Triticum turgidum L. durum (Desf.)). *PLOS ONE*, **10**, e0130854. http://doi.org/10.1371/journal.pone.0130854

Huang W, Perez-Garcia P, Pokhilko A, Millar a J, Antoshechkin I, Riechmann JL, … & Pérez-García P (2012) Mapping the Core of the Arabidopsis Circadian Clock Defines the Network Structure of the Oscillator. *Science*, **336**, 75–79. Journal Article. http://doi.org/Doi 10.1126/Science.1219075

Hudson KA (2010) The Circadian Clock-controlled Transcriptome of Developing Soybean Seeds. *The Plant Genome Journal*, **3**, 3. http://doi.org/10.3835/plantgenome2009.08.0025

International Board for Plant Genetic Resources (IBPGR) (1983) Descriptors for cowpea. Rome.

Iqbal A, Khalil IA, Ateeq N & Khan MS (2006) Nutritional quality of important food legumes. *Food Chemistry*, **97**, 331-335. https://doi.org/10.1016/j.foodchem.2005.05.011

James AB, Monreal JA, Nimmo GA, Kelly CL, Herzyk P, Jenkins GI & Nimmo HG (2008) The Circadian Clock in Arabidopsis Roots Is a Simplified Slave Version of the Clock in Shoots. *Science*, **322**, 1832–1835. Journal Article. http://doi.org/Doi 10.1126/Science.1161403

Jones HG (2014) Plants and microclimate. A quantitative approach to environmental plant physiology. 3rd edition. Cambrige University Press. Cambridge, United Kingdom. ISBN 978-0-521-27959-8

Johnson HW, Robinson HF & Comstock RE (1955) Estimates of genetic and environmental variability in soybean. *Agronomy Journal* **47**, 274-318.

Jünemann S, Prior K, Szczepanowski R, Harks I, Ehmke B, Goesmann A, … & Harmsen D (2012). Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PloS One*, **7**, e41606. http://doi.org/10.1371/journal.pone.0041606

Karapanos I, Papandreou A, Skouloudi M, Makrogianni D, Fernández JA, Rosa E, Ntatsi G, Bebeli PJ & Savvas D (2017) Cowpea fresh pods: A new legume for the market. Assessment of their quality and dietary characteristics of 37 cowpea accessions grown in southern Europe. *Journal of the Science of Food and Agriculture*. https://doi/10.1002/jsfa.8418/full

Kaur S, Kimber RBE, Cogan NOI, Materne M, Forster JW & Paull JG (2014) SNP discovery and high-density genetic mapping in faba bean (*Vicia faba* L.) permits identification of QTLs for ascochyta blight resistance. *Plant Science : An International Journal of Experimental Plant Biology*, **217–218**, 47–55. http://doi.org/10.1016/j.plantsci.2013.11.014

Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, … & Ecker JR (2016) Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell*, **166**, 492–505. http://doi.org/http://dx.doi.org/10.1016/j.cell.2016.06.044

Khalid II, Elhardallou SB & Elkhalifa EA (2012) Composition and functional properties of cowpea (*Vigna ungiculata* L. Walp) flour and protein isolates. *American Journal of Food Technology* **7**, 113-122. https://doi/10.3923/ajft.2012.113.122

Kinmonth-Schultz HA, Golembeski GS & Imaizumi T (2013) Circadian clock-regulated physiological outputs: dynamic responses in nature. In *Seminars in cell & developmental biology* (Vol. 24, No. 5, pp. 407-413). Academic Press. https://doi.org/10.1016/j.semcdb.2013.02.006

Koboldt DC, Steinberg KM, Larson DE, Wilson RK & Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell*, **155**, 27-38. http://dx.doi.org/10.1016/j.cell.2013.09.006

Koivunen E, Partanen K, Perttilä S, Palander S, Tuunainen P & Valaja J (2016) Digestibility and energy value of pea (*Pisum sativum* L.), faba bean (*Vicia faba* L.) and blue lupin (narrow-leaf) (*Lupinus angustifolius*) seeds in broilers. *Animal Feed Science and Technology*, **218**, 120-127. https://doi.org/10.1016/j.anifeedsci.2016.05.007

Kreps JA & Kay SA (1997) Coordination of plant metabolism and development by the circadian clock. *The Plant Cell*, ***9***, 1235. http://dx.doi.org/10.1105/tpc.9.7.1235

Kumar S, Stecher G & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870–1874. http://doi.org/10.1093/molbev/msw054

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, … & Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, **23**, 2947–8. http://doi.org/10.1093/bioinformatics/btm404

Lavania D, Siddiqui MH, Al-Whaibi MH, Singh AK, Kumar R & Grover A (2015) Genetic approaches for breeding heat stress tolerance in faba bean (*Vicia faba* L.). *Acta physiologiae plantarum*, **37**, 1737. https://doi.org/10.1007/s11738-014-1737-z

Lazányi J (2002) Trends in Dry Pea (*Pisum sativum* L.) Production. *J, Agr. Microbiol*, **5,** 19-24.

Lee S, Abecasis GR, Boehnke M & Lin X (2014) Rare-variant association analysis: study designs and statistical tests. *American Journal of Human Genetics*, **95**, 5–23. http://doi.org/10.1016/j.ajhg.2014.06.009

Leonforte A, Sudheesh S, Cogan NO, Salisbury PA, Nicolas ME, Materne M, … & Mitchell S (2013) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). *BMC Plant Biology*, **13**, 161. http://doi.org/10.1186/1471-2229-13-161

Li JY, Wang J & Zeigler RS (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. *GigaScience*, **3**, 8. http://doi.org/10.1186/2047-217X-3-8

Liew LC, Hecht V, Laurie RE, Knowles CL, Vander Schoor JK, Macknight RC & Weller JL (2009) *DIE NEUTRALIS* and *LATE BLOOMER 1* contribute to regulation of the pea circadian clock. *The Plant Cell*, **21**, 3198-3211. https://doi.org/10.1105/tpc.109.067223

Li Y, Sidore C, Kang HM, Boehnke M & Abecasis GR (2011) Low-coverage sequencing:

implications for design of complex trait association studies. *Genome*, **21** 940–951. http://doi.org/10.1101/gr.117259.110.individuals

Locke JCW, Kozma-Bognár L, Gould PD, Fehér B, Kevei E, Nagy F, … & Millar AJ (2006) Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Molecular Systems Biology*, **2**, 59. http://doi.org/10.1038/msb4100102

Lynch M (2009) Estimation of allele frequencies from high-coverage genome-sequencing projects. *Genetics*, **182**, 295–301. http://doi.org/genetics.109.100479 [pii]10.1534/genetics.109.100479

Mallona I, Egea-Cortines M & Weiss J (2011) Conserved and divergent rhythms of CAM-related and core clock gene expression in the cactus Opuntia ficus-indica. *Plant Physiology*, **156**, 1978–1989.

Mallona I, Lischewsky S, Weiss J, Hause B & Egea-Cortines M (2010) Validation of endogenous genes as controls for quantitative real-time PCR during leaf and flower development in Petunia hybrida. *BMC Plant Biology*, **10**, 4.

Mallona I, Lischewsky S, Weiss J, Hause B & Egea-Cortines M (2010) Validation of reference genes for quantitative real-time PCR during leaf and flower development in Petunia hybrida. *BMC Plant Biology*, **10**, 4. http://doi.org/10.1186/1471-2229-10-4

Mallona I, Weiss J & Egea-Cortines M (2011) pcrEfficiency: a Web tool for PCR amplification efficiency prediction. *BMC Bioinformatics*, **12**, 404. http://doi.org/10.1186/1471-2105-12-404

Manggoel W & Uguru MI (2011) Comparative study on the phenology and yield components of two photoperiodic groups of cowpea (*Vigna unguiculata* (L.) Walp.) in two cropping seasons. *African Journal of Agricultural Research* **6,** 5232-5241. https://doi/10.5897/AJAR11.156

Manggoel W, Uguru MI, Ndam ON & Daskak MA (2012) Genetic variability, correlation and path coefficient analysis of some yield components of tem cowpea [*Vigna unguiculata* (L.) Walp.] accessions. *Journal of Plant Breeding and Crop Science* **4**, 80-86. https://doi/10.5897/JPBCS12.007

Marcolino-Gomes J, Rodrigues FA, Fuganti-Pagliarini R, Bendix C, Nakayama TJ, Celaya B, … & Nepomuceno A (2014) Diurnal oscillations of soybean circadian clock and

drought responsive genes. *PloS One*, **9**, e86402. http://doi.org/10.1371/journal.pone.0086402

Martos-Fuentes, M, Sánchea-Navarro V, Ruiz-Hérnandez MV, Weiss J, Egea-Gilabert C, Zornoza R, … & Egea-Cortines M (2015). Genetic and growth conditions determine the protein content in cowpea (*Vigna unguiculata*). In *Eucarpia International Symposium on Protein Plants* (pp. 143–144).

McClung CR. (2006). Plant circadian rhythms. *Plant Cell*, **18**, 792–803.

Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, … & Mesirov J (2009) A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biology*, **10**, R64. http://doi.org/10.1186/gb-2009-10-6-r64

Mendes-Moreira P, Alves ML, Satovic Z, Santos JP, Santos JN, Souza JC, Pêgo SE, Hallauer AR & Vaz Patto MC (2015) Genetic architecture of ear fasciation in maize (*Zea mays*) under QTL scrutiny. *PLoS ONE* **10**, 1-31. https://doi.org/10.1371/journal.pone.0124543

Michael TP, Mockler TC, Breton G, McEntee C, Byer A, Trout JD, … & Chory J (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genetics*, **4**, e14. http://doi.org/10.1371/journal.pgen.0040014

Mishra P & Singh P (2014) Genetic variability, heritability, genetic advance, correlation coefficient and path analysis in gladiolus. *IOSR Journal of Agriculture and Vetenary Science* **7**, 23-6. ISSN: 2319-2380

Nautrup-Pedersen G, Dam S, Laursen BS, Siegumfeldt AL, Nielsen K, Goffard N, … & Stougaard J (2010) Proteome analysis of pod and seed development in the model legume Lotus japonicus. *Journal of Proteome Research*, **9**, 5715–26. http://doi.org/10.1021/pr100511u

Newbigin EJ, deLumen BO, Chandler PM, Gould A, Blagrove RJ, March JF, Kortt AA & Higgins TJV (1990) Pea convicilin: structure and primary sequence of the protein and expression of a gene in the seeds of transgenic tobacco. *Planta* **180**, 461. https://doi.org/10.1007/BF02411442

Ng NQ & Marechal R (1985) Cowpea taxonomy, origin and germplasma. In "Cowpea research, production and utilization". (Eds Singh SR, Rachie KO) pp.11-21. (John Wiley & Sons Ltd, London).

Nicot N, Hausman J-F Hoffmann L & Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany*, **56**, 2907–14. http://doi.org/10.1093/jxb/eri285

Noda K, Glover BJ, Linstead P & Martin C (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature*, **369**, 661–4. http://doi.org/10.1038/369661a0

Novák P, Neumann P & Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC bioinformatics*, **11**, 378. https://doi.org/10.1186/1471-2105-11-378

Nybom H, Weising K & Rotter B (2014) DNA fingerprinting in botany: past, present, future. *Investigative Genetics*, *5*(1), 1. http://doi.org/10.1186/2041-2223-5-1

Ochoa J, Niñiirola D, Conesa E, Lara L, López-Marín J & Fernández JA (2011) Predicting purslane (*Portulaca oleracea* L.) harvest in a hydroponic floating system. *Acta Horticulturae* **898**, 205–209. https://doi.org/10.17660/ActaHortic.2011.898.25

Ogunkanmi LA, Fakoya AO & Adekoya KO (2013) Breeding for long pod trait in Cowpea (*Vigna unguiculata* L. Walp). *Journal of Sci. Res. Dev*, **14**, 9-14.

Oliveira HR, Tomás D, Silva M, Lopes S, Viegas W, Veloso MM, … & Prance G (2016) Genetic Diversity and Population Structure in Vicia faba L. Landraces and Wild Related Species Assessed by Nuclear SSRs. *PLOS ONE*, **11**, e0154801. http://doi.org/10.1371/journal.pone.0154801

Oluwatosin OB (1997) Genetic and environmental variation for seed yield, protein, lipid and amino acid composition in cowpea (*Vigna unguiculata* (L) Walp). *Journal of the Science of Food and Agriculture* **74**, 107-116. https://doi.org/10.1002/(SICI)1097-0010(199705)74:1<107::AID-JSFA776>3.0.CO;2-L

Omoigui LO, Ishiyaku MF, Kamara AY, Alabi SO & Mohammed SG (2006) Genetic variability and heritability studies of some reproductive traits in cowpea (*Vigna unguiculata* (L.) Walp.). *African Journal of Biotechnology* **5**, 1191-1195. ISSN 1684–5315

O'Neill HM, Rademacher M, Mueller-Harvey I, Stringano E, Kightley S & Wiseman J

(2012). Standardised ileal digestibility of crude protein and amino acids of UK-grown peas and faba beans by broilers. *Animal Feed Science and Technology*, **175**, 158-167. https://doi.org/10.1016/j.anifeedsci.2012.05.004

Osborne TB (1924) The Vegetable Proteins. (Longmans, Green and Co., London). http://dx.doi.org/10.5962/bhl.title.18912

Oton M. (2017) *Elaboración industrial de haba mínimamente procesada en fresco. Optimización de su calidad y vida comercial*. Tesis doctoral. Universidad Politécnica de Cartagena. España.

Padulosi S & Ng NQ (1997) Origin, taxonomy, and morphology of *Vigna unguiculata* (L.) Walp. In "Advances in cowpea research". (Eds Singh BB, MohanRaj D., Dashiell KE, Jackai LEN) pp. 1-12. (IITA-JIRCAS, Ibadan).

Palander S, Laurinen P, Perttilä S, Valaja J & Partanen K (2006). Protein and amino acid digestibility and metabolizable energy value of pea (*Pisum sativum*), faba bean (*Vicia faba*) and lupin (*Lupinus angustifolius*) seeds for turkeys of different age. *Animal feed science and technology*, **127**, 89-100. https://doi.org/10.1016/j.anifeedsci.2005.07.003

Paoletti MG & Pimentel D (1996) Genetic engineering in agriculture and the environment. *BioScience*, **46**, 665-673. https://doi.org/10.2307/1312896

Pawluczyk M, Weiss J, Links MG, Egaña Aranguren M, Wilkinson MD & Egea-cortines M. (2015). Quantitative evaluation of bias in PCR amplification and nex-generation sequencing derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry*, **407**, 1841–1848. http://doi.org/10.1007/s00216-014-8435-y

Peksen E & Artik A (2004) Comparison of some cowpea (*Vigna unguiculata* L. Walp.) genotypes from Turkey for seed yield and yield-related characters. *Journal of Agronomy* **3**, 137-140. https://doi.org/10.3923/ja.2004.137.140

Pfaffl MW, Horgan GW & Dempfle L (2002) Relative expression software tool (REST(C)) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucl. Acids Res.*, **30**, e36-. http://doi.org/10.1093/nar/30.9.e36

Pfaffl MW, Tichopad A, Prgomet C & Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters*,

26, 509–515.

Pihur V & Datta SS (2009) RankAggreg, an R package for weighted rank aggregation. *BMC Bioinformatics*, **10**, 62.

Ravelombola WS, Ainong S, Yuejin W, Dennis M, Pengyin C, Vibha S & Clay W (2016) Evaluation of total seed protein content in eleven Arkansas cowpea (*Vigna unguiculata* (L.) Walp.) Lines. *American Journal of Plant Sciences* **7**, 2288-2296. https://doi.org/10.4236/ajps.2016.715201

Red Andaluza de Semillas (2017) Descriptor Guisante (*Pisum sativum* L.) http://www.redandaluzadesemillas.org/IMG/pdf/Guion_Descriptor_RAS_Guisante.pdf

Redden B, Leonforte T, Ford R, Croser J & Slattery J (2005). Pea (Pisum sativum L.). *Genetic resources, chromosome engineering, and crop improvement*, **1**, 49-83.

Riely BK, Ane JM, Penmetsa RV & Cook DR (2004) Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. *Current opinion in plant biology*, **7**, 408-413. https://doi.org/10.1016/j.pbi.2004.04.005

Robinson HF, Comstock RE & Harvey PH (1949) Estimates of heritability and the degree of dominance in corn. *Agronomy Journal* **41**, 353-359.

Ryo M, Matsuo T, Yamashino T, Ichinose M, Sugita M & Aoki S (2016) Diversity of plant circadian clocks: Insights from studies of Chlamydomonas reinhardtii and Physcomitrella patens. *Plant Signaling & Behavior*, **11**, e1116661. http://doi.org/10.1080/15592324.2015.1116661

Sáenz de Miera LE, Ramos J & Pérez De La Vega M (2008) A comparative study of convicilin storage protein gene sequences in species of the tribe Vicieae. *Genome*, **51**, 511-523. https://doi.org/10.1139/G08-036

Saitou N & Nei M (1987) The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**, 406–425. Journal Article.

Sales MP (2000) Do Legume Storage Proteins Play a Role in Defending Seeds against Bruchids? *PLANT PHYSIOLOGY*, **124**, 515–522. http://doi.org/10.1104/pp.124.2.515

Sales MP, Pimenta PP, Paes NS, Grossi-de-Sá MF & Xavier-Filho J (2001) Vicilins (7S storage globulins) of cowpea (*Vigna unguiculata*) seeds bind to chitinous

structures of the midgut of Callosobruchus maculatus (Coleoptera: Bruchidae) larvae. *Brazilian Journal of Medical and Biological Research = Revista Brasileira de Pesquisas Médicas E Biológicas / Sociedade Brasileira de Biofísica ... [et Al.]*, **34**, 27–34.

Sathe SK, Deshpande SS, Salunkhe DK & Rackis JJ (1984) Dry beans of *phaseolus* . A review. Part 1. Chemical composition: Proteins. *C R C Critical Reviews in Food Science and Nutrition*, **20**, 1–46. http://doi.org/10.1080/10408398409527382

Scaife A, Cox EF & Morris GEL (1987) The relationship between shoot weight, plant density and time during the propogation of four vegetable species. *Annals of Botany* **59**, 325-334. https://doi.org/10.1093/oxfordjournals.aob.a087321

Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M & Wisman E (2001) Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. *The Plant Cell*, **13**, 113–23.

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

Scholz G, Manteuffel R & Richter J (1974) Studies on Seed Globulins from Legumes: I. Separation and Purification of Legumin and Vicilin from *Vicia faba* L. by Zone Precipitation. *Biochemie und Physiologie der Pflanzen*, **166**, 163-172. https://doi.org/10.1016/S0015-3796(17)30025-2

Schuster-Gajzágó I (2004) Nutritional aspects of legumes. *Encyclopedia of Food and Agricultural Sciences, Engineering and Technology Resources (Cultivated Plants, Primarily as Food Sources, 1. Grains and Cereals. Encyclopedia of Life Support System (EOLSS). Developed under the auspices of the UNESCO, Eolss Publishers, Oxford, UK. http://www. eolss. net*.

Serraj, R, Bidinger FR, Chauhan Y S, Seetharama N, Nigam SN & Saxena NP (2003). Management of drought in ICRISAT cereal and legume mandate crops. *Water productivity in agriculture: Limits and opportunities for improvement*, 127-144. https://doi.org/10.1079/9780851996691.0127

Shewry PR, Napier JA & Tatham AS (1995) Seed storage proteins: structures and biosynthesis. *The Plant Cell Online*, **7**, 945-956. https://doi.org/10.1105/tpc.7.7.945

Shimelis H & Shiringani R (2010) Variance components and heritabilities of yield and agronomic traits among cowpea genotypes. *Euphytica* **176**, 383-389. http://doi.org/10.1007/s10681-010-0222-z

Shiringani RP & Shimelis HA (2011) Yield response and stability among cowpea genotypes at three planting dates and test environments. *African Journal of Agricultural Research* **6**, 3259-3263. http://doi.org/10.5897/AJAR11.496

Silva AC, Morais OM, Santos JL, D'Arede LO, Silva CJ & Rocha M (2014) Estimativa de parâmetros genéticos em *Vigna unguiculata*. *Sociedade de Ciências Agrárias de Portugal* **37**, 399-407. ISSN 0871-018X.

Simmonds NC (1995) The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture* **67**, 309-315. http://doi.org/10.1002/jsfa.2740670306

Singh BB (2002) Recent genetic studies in cowpea. In "Cowpea genetic and Breeding" (Eds Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamò M) pp. 3-13. (International Institute of Tropical Agriculture, Ibadan, Nigeria). ISBN 978 131 110 X

Singh BB, Ajeigbe HA, Tarawali SA, Fernandez-Rivera S & Abubakar M (2003) Improving the production and utilization of cowpea as food and fodder. *Field Crops Research*, **84**, 169-177. https://doi.org/10.1016/S0378-4290(03)00148-5

Smartt J (1990) Grain legumes: evolution and genetic resources. *Cambridge University Press*, Cambrige, UK, 379 pp.

Smith AM, Heisler LE, St Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, … & Nislow C (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Research*, **38**, e142. http://doi.org/10.1093/nar/gkq368

Somerville C & Briscoe J (2001) Genetic engineering and water. *Science*, **292**, 2217-2226. http://doi.org/10.1126/science.292.5525.2217

Spiess AN, Feig C & Ritz C (2008) Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. *BMC Bioinformatics*, **9**, 221. http://doi.org/10.1186/1471-2105-9-221

Staiger D, Shin J, Johansson M & Davis SJ (2013) The circadian clock goes genomic. *Genome Biology*, **14**, 208. http://doi.org/10.1186/gb-2013-14-6-208

Takahashi N, Hirata Y, Aihara K & Mas P (2015) A hierarchical multi-oscillator network

orchestrates the Arabidopsis circadian system. *Cell*, **163**, 148-159. https://doi.org/10.1016/j.cell.2015.08.062

Tarawali SA, Singh BB, Fernandez-Rivera S, Peters M, Smith JW, Schultze-Kraft R & Ajeigbe HA (1997) Optimizing the contribution of cowpea to food and fodder production in crop–livestock systems in West Africa. Grasslands 2000, Proceedings of the XVIII International Grassland Congress, Winnipeg and Saskatoon, Canada, pp. 19-53.

Tayeh N, Thaben PF & Westermark PO (2014) Detecting rhythms in time series with RAIN. *Journal of Biological Rhythms*, **29**, 391–400. http://doi.org/10.1177/0748730414553029

Thaben PF & Westermark PO (2014) Detecting rhythms in time series with RAIN. *Journal of biological rhythms*, **29**, 391-400. https://doi.org/10.1177/0748730414553029

Thain SC, Hall A & Millar AJ (2000) Functional independence of circadian clocks that regulate plant gene expression. *Current Biology : CB*, **10**, 951–6.

Thain SC, Murtas G, Lynn JR, McGrath RB & Millar A J (2002) The circadian clock that controls gene expression in Arabidopsis is tissue specific. *Plant Physiology*, **130**, 102–10. http://doi.org/10.1104/pp.005405

Thomé OW (1885) Flora von Deutschland, Österreich und der Schweiz in Wort und Bild für Schule und Haus. (Eds Köhler E) Vol 1. 366 pp. (Gera-Unthermaus). http://dx.doi.org/10.5962/bhl.title.5360

Thompson AJ, Bown D, Yaish S & Gatehouse JA (1991) Differential expression of seed storage protein genes in the pea *legJ* subfamily; sequence of gene *legK*. *Biochemie und Physiologie der Pflanzen*, **187**, 1-12. https://doi.org/10.1016/S0015-3796(11)80177-0

Tiedemann J, Neubohn B & Müntz K (2000) Different functions of vicilin and legumin are reflected in the histopattern of globulin mobilization during germination of vetch (*Vicia sativa* L.). *Planta*, **211**, 1-12. https://doi.org/10.1007/s004250000259

Tilman D & Clark M (2014) Global diets link environmental sustainability and human health. *Nature*, **515**, 518–522. http://doi.org/10.1038/nature13959

Timko MP, Ehlers JD & Roberts PA (2007) Cowpea. In "Genome mapping and molecular breeding in plants: Pulses, sugar and tuber crops". (Eds Kole C) pp. 49-67. (Springer-Verlag, Berlin).

Timko MP & Singh BB (2008) Cowpea, a multifunctional legume. In "Genomics of tropical crop plants". (Eds Moore PH, Ming R) pp. 227-258. (Springer Science + Business Media LCC, New York).

van Dijk EL, Jaszczyszyn Y & Thermes C (2014) Library preparationmethodsfornext-generation sequencing: Tone down the bias. *Experimental cell research*, **322**, 12-20. https://doi.org/10.3390/v3101849

Van Moerkercke A, Haring MA & Schuurink RC (2011) The transcription factor EMISSION OF BENZENOIDS II activates the MYB ODORANT1 promoter at a MYB binding site specific for fragrant petunias. *The Plant Journal : For Cell and Molecular Biology*, **67**, 917–28. http://doi.org/10.1111/j.1365-313X.2011.04644.x

Vance CP, Graham PH & Allan DL (2000) Biological nitrogen fixation: phosphorus-a critical future need?. In *Nitrogen fixation: From molecules to crop productivity* (pp. 509-514). Springer Netherlands. https://doi.org/10.1007/0-306-47615-0_291

Vasconcelos IM, Maia FMM, Farias DF, Campello CC, Carvalho AFU, de Azevedo Moreira R & de Oliveira JTA (2010) Protein fractions, amino acid composition and antinutritional constituents of high-yielding cowpea cultivars. *Journal of Food Composition and Analysis*, **23**, 54–60. http://doi.org/10.1016/j.jfca.2009.05.008

Klug TV, Martínez-Sánchez A, Gómez PA, Collado E, Aguayo E, Artés F & Artés-Hernandez F (2017) Improving quality of an innovative pea puree by high hydrostatic pressure. *Journal of the Science of Food and Agriculture*. http://doi.org/10.1002/jsfa.8454

Verdier J, Torres-Jerez I, Wang M, Andriankaja A, Allen SN, He J, … & Udvardi MK (2013) Establishment of the Lotus japonicus Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *The Plant Journal : For Cell and Molecular Biology*, ***74***, 351–62. http://doi.org/10.1111/tpj.12119

Wang Z & Wang T (2011) Dynamic proteomic analysis reveals diurnal homeostasis of key pathways in rice leaves. *Proteomics*, **11**, 225–38. http://doi.org/10.1002/pmic.201000065

Weller JL, Liew, LC, Hecht VF, Rajandran V, Laurie RE, Ridge S, ... & Dalmais M (2012) A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proceedings of the National Academy of Sciences*, **109**, 21158-21163. http://doi.org/10.1073/pnas.1207943110

Wienkoop S, Weiß J, May P, Kempa S, Irgang S, Recuenco-Munoz L, ... & Weckwerth W (2010) Targeted proteomics for Chlamydomonas reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. *Molecular BioSystems*, **6**, 1018-1031. http://doi.org/10.1039/B920913A

Wobus U, Borisjuk L, Panitz R, Manteuffel R, Bäumlein H, Wohlfahrt T, … & Weschke W (1995). Control of Seed Storage Protein Gene Expression: New Aspects on an Old Problem. *Journal of Plant Physiology*, **145**, 592–599. https://doi.org/10.1016/S0176-1617(11)81269-1

Xie F, Xiao P, Chen D, Xu, L & Zhang B (2012) miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Molecular Biology*. **80**, 75-84. http://doi.org/10.1007/s11103-012-9885-2

Xiong M, Zhao Z, Arnold J & Yu F (2010) Next-Generation Sequencing. *Journal of Biomedicine and Biotechnology*, **2010**, 370710. http://doi.org/10.1155/2010/370710

Yang T, Bao SY, Ford R, Jia TJ, Guan JP, He YH, ... & Zong XX (2012) High-throughput novel microsatellite marker of faba bean via next generation sequencing. *BMC genomics*, **13**, 602. https://doi.org/10.1186/1471-2164-13-602

Young MW & Kay SA (2001) Time zones: a comparative genetics of circadian clocks. *Nature reviews. Genetics*, **2**, 702.

Zeeman SC, Smith SM & Smith AM (2007) The diurnal metabolism of leaf starch. *The Biochemical Journal*, **401**, 13–28. http://doi.org/10.1042/BJ20061393

Zheng SH, Maeda A, Kashiwagi Y & Nakamoto AFM (2004). Simultaneous growth of pods and seeds set on different racemes in soybean. In *4th International Crop Science Congress*.

Zuckerkandl E & Pauling L (1965) Evolutionary Divergence and Convergence, in Proteins. In *Evolving Genes and Proteins* (pp. 97–166).