Palacký University in Olomouc

Faculty of Science

Department of Botany

Genetic diversity of genus Pisum, its exploitation and pea domestication

Habilitation thesis

Botany

Ing. Petr Smýkal, Ph.D.

Olomouc August 2016

Declaration of originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it contains no material written by another person or material which to a substantial extent has been accepted for the award of any other degree or diploma of the university.

Olomouc, August 2016 Ing. Petr Smýkal, Ph.D.

The author grants permission to Palacký University in Olomouc to store and display this thesis and its electronic version in university library and on official website.

Poděkování

Na tomto místě bych chtěl poděkovat všem kolegům z katedry botaniky, PřF UP v Olomouci za podporu a vytvoření příjemného pracovního prostředí, a také kolegům ve svém předchozím zaměstnání ve společnosti Agritec Plant Research s.r.o. v Šumperku, kde jsem měl možnost začít pracovat na tématu analýzy a využití genetické diverzity genofondové sbírky hrachu. Děkuji rovněž mnoha zahraničním kolegům, díky nimž jsem mohl postupně proniknout do tajů fylogeneze, genetické diverzity a jejího využití ve šlechtění, především pak dr. Clarice Coyne, USDA-ARS, Pullman, USA, dr. Mike Ambrose, kurátorovi JIC Pisum kolekce, UK, dr. Robertu Reddenovi, kurátorovi Australian Grain Germplasm, dr. Aleksandarovi Mikičovi a jeho nezlomné vůli pro založení Legume Society a studium Vavilovie, dr. Gregovi Kenicerovi z Royal Botanical Garden, Edinburgh, který mě zasvětil do problematiky fylogeneze, prof. Noelu Ellisovi a prof. Andrew Flavellovi, Dundee, UK za dlouhodobou podporu při studiu diverzity a genetiky hrachu a mnoha dalším přátelům z celého světa.

V neposlední řadě děkuji svým rodičům, kteří mě vždy podporovali v práci, studiu i životě. Samozřejmě nemohu opomenout svou rodinu, především děti Karolínu, Elišku a Lukáše, kteří mě byly motivací a inspirací. Můj velký díky patří mé partnerce Andreje, bez jejíž podpory bych se možná k sepsání této práce neodhodlal.

Acknowledgements

I would like to express my thanks to collegues from Department of Botany, UP as well as from my former employment, Agritec Plant Research Ltd. where I have started my work on analysis of pea diversity, for friendship,support and stimulating discussions . I am gratefull to numerous colleagues worldwide, to whom I owe my knowledge of phylogeny, genetic diversity and its exploatation in breeding, particularly to dr. Clarice Coyne, USDA-ARS, Pullman, USA, dr. Mike Ambrose, JIC Pisum collection curator, UK, dr. Robert Redden, curator of Australian Grain Germplasm, dr. Aleksandar Mikič and his effort to establish Legume Society and study of mysterious Vavilovie, dr. Greg Kenicer from Royal Botanical Garden, Edinburgh, , prof. Noel Ellis and prof. Andrew Flavell, Dundee, UK for support of pea diversity analysis.

I am gratefull to my parents which supported me all the time as well as my family particularly to my childern Karolína, Eliška and Lukáš, which were sources of my motivation and inspiration. My great thanks belongs to my partner Andrea, without which this thesis might not come out.

Téma a členění práce

Hrách setý (*Pisum sativum*) je znám jako modelový biologický objekt, na němž v r. 1845 objevil opat brněnského augustiánského kláštera Johan Gregor Mendel (1822–84) zákonitosti dědičnosti a následně tak položil základy genetiky. Mendel využil již v té době dostupné diverzity kulturního hrachu pro pochopení přenosu znaků do dalších generací. Hrách patří mezi nejstarší kulturní plodiny, domestikované před přibližně 10 tisíci lety v oblasti Úrodného půlměsíce. V současnosti jde o celosvětově třetí nejpěstovanější luskovinu, jejíž semena bohatá na proteiny (20–25 % hmotnosti suchého semene) slouží jako potravina, zelenina i jako krmivo pro zvířata. Za oblast původu a počáteční domestikace se pokládá Středomoří a především Blízký východ. Planý hrách (*Pisum sativum* subsp. *elatius*) se v přírodě nachází v jižní Evropě, západní a Malé Asii a jeho areál zasahuje až do severní Afriky. Příbuzný druh *P. fulvum* se vyskytuje jen na Blízkém východě v oblasti Israele, Syrie, Libanonu a Jordánska. *P. abyssinicum* je pěstovaný v Etiopii a Jemenu. Hrách patří do podčeledi Fabeae, která je považována za vývojově pokročilou skupinu. Na základě sekvence chloroplastového genu maturázy K bylo její oddělení datováno do doby 20 milionů let, tedy do období středního miocénu. Podčeleď zahrnuje celkem pět rodů, včetně prastarých kulturních plodin, hrachoru (přibližně 160 druhů), čočky (4 druhy), hrachu (2-3 druhy) a vikve (140 druhů) a dále monotypický rod Vavilovia. Hlavním centrem diverzity podčeledi je východní Středomoří.

V důsledku staletí následného výběru a šlechtění dnes existují tisíce odrůd, genotypů hrachu. Celosvětově je v genových bankách uchováváno kolem 98 tisíc položek (genotypů) hrachu, z nichž je část zastoupena vícenásobně a naopak plané formy jen velmi nedostatečně. Česká národní kolekce hrachu čítající přes 2500 položek patří mezi 10 největších světových sbírek. Poznání genetické struktury je významné i z hlediska šlechtitelského využití, kdy je důležité znát míru genetické příbuznosti rodičovských komponent. Obecně lze říci, že s větší vzdáleností roste pravděpodobnost získání nových alel a tím kombinací vlastností. Poznání fylogenetických vztahů v rámci čeledi bobovitých (Fabaceae) je velmi důležité pro porozumění původu a rozrůznění těchto ekologicky a ekonomicky důležitých rostlin.

Práce je členěna na kapitoly týkající se analýzy genetické diverzity a biogeografie rodu *Pisum* a příbuzného monotypického rodu *Vavilovia formosa*. Následuje pohled do procesu domestikace hrachu, včetně archeogenetické analýzy DNA isolované ze semen hrachu z doby železné. Hrách je představen jako vhodný model pro porozumění procesu domestikace, především pak dvou klíčových vlastností tj. dormance semen a pukavosti lusku.

Na příkladu analýzy genu rezistence k virové mozaice přenosné semenem, je ukázána možnost využití genetické diverzity, podobně pak ve tvorbě a analýze introgresních liniích, umožňujících systematické využití planých předchůdců kulturních plodin, v tomto případě *P. fulvum*.

Tato práce předložená k habilitačnímu řízení si klade za cíl přispět k poznání genetické diverzity a částečně i fylogenetických vztahů v rámci rodu *Pisum*. Práce je založena na 22 vybraných publikovaných recenzovaných pracích, je členěna do 4 širších kapitol (1) Legumes phylogeny, 2) Pisum genus diversity, 3) Pea domestication, 4) Use of pea diversity for breeding.

V textu jsou tučně vyznačeny články na nichž jsem se autorsky podílel. Předkládané práce vznikly na pracovišti Agritec Plant Research s.r.o. v Šumperku (2004-2011) a na katedře botaniky PřF UP v Olomouci (2011-2016), ve spolupráci s celou řadou pracovišť v Evropě i ve světě.

Práce byla podporována řadou výzkumných projektů grantové agentury NAZVa Ministerstva Zemědělství ČR, Ministerstva školství, mládeže a tělovýchovy ČR (INGO, COST a KONTAKT), Grantové agentury ČR, Bioversity International, FP7 Evropské unie (Legato projekt) a stipendii Royal Society of Edinburgh.

Motto:

"Always do your best. What you plant now, you will harvest later. "

Og Mandino

"If you think in terms of a year, plant a seed; if in terms of ten years, plant trees; if in terms of 100 years, teach the people. "

Confucius

Content

Chapter 1

Legumes

1.3 Legumes and their use

Commenting on ***Smýkal P****, Coyne C and 14 co-authors (2015) Legume crops phylogeny and genetic diversity for science and breeding. Critical Reviews Plant Sciences 34: 43-104.*

Upadhyaya HD, Dwivedi SL, Ambrose M, Ellis N, Berger J, ***Smýkal P****, Debouck D, Duc G, Dumet D, Flavell A, Sharma SK, Mallikarjuna N, Gowda CL (2011) Legume genetic resources: management, diversity assessment, and utilization in crop improvement. Euphytica 180: 27-47.*

Legumes represent the second most important family of crop plants after *Poaceae* (grass family), accounting for approximately 27% of the world's crop production. While in cereals the major storage molecule is starch, which is deposited in the endosperm, in most of the grain legumes (pulses) the endosperm is transitory and consumed by the embryo during seed maturation, which contains a high proportion of proteins (20–40%), and either lipids (soybean, peanut) or starch (or both) as a further carbon source(**Smýkal *et al.* 2015**). In addition to traditional food and forage uses, legumes can be milled into flour, used to make bread, doughnuts, tortillas, chips, spreads, and extruded snacks. This importance was recognized by the United Nations General Assembly which declared this year (2016) as the 'International Year of Pulses'. Encouraging more pulses to be grown and eaten, and improving the protein content of the varieties under cultivation, is the goal of the international development and research communities. In addition to be a protein source, legumes (predominantly soybean and peanut) provide more than 35% of the world's processed vegetable oil and these are also rich sources of dietary protein for the chicken and pork industries. Nutritionally, they are generally deficient in sulfur-containing amino acids (cysteine and methionine), but unlike cereal grains, their lysine content is relatively high (**Smýkal *et al.* 2015**). The major storage proteins are globulins, which account for up to 70% of the total seed nitrogen. One of the most important attributes of legumes is their capacity for symbiotic nitrogen fixation, underscoring their importance as a source of nitrogen in both natural and agricultural ecosystems. Man used leguminous plants to enrich the soil centuries before he knew what made them useful. Records from the oldest civilizations of Egypt and eastern Asia demonstrate the

ancient use of various beans, peas, vetches, soybeans, and alfalfa. One of the early Greek botanists, Theophrastus, in third century before Christ, wrote of leguminous plants "reinvigorating" the soil and stated that beans were not a burdensome crop to the ground but even seemed to manure it. The Romans laid emphasis on the use of leguminous plants for green manuring; they also introduced the systematic use of crop rotations, a practice that was forgotten for a time during the early Middle Ages and partly also in todays agricultural practice. Alfalfa, also known as lucerne, is the most widely grown forage legume in the world. In the United States, alfalfa places among the top five crops in the nation in terms of both farmgate value and total acreage. In terms of protein production, alfalfa placed third, behind soybeans and corn. From a global perspective, alfalfa is among the top 10 crops for protein production. Legumes also accumulate natural products (secondary metabolites) such as isoflavonoids that are considered beneficial to human health through anticancer and other health-promoting activities. The ability of legumes to fix atmospheric nitrogen allows them to colonize poor soils; however adequate nitrogen reserves in the seed are vital to allow the seedling to survive the heterotrophic growth phase before nitrogen fixation is established in root nodules. Many tree-sized species in the legume family are valuable for their hard, durable timber. Species from the genera *Aeschynomene, Arachis, Centrosema, Desmodium, Macroptilium*, and particularly *Stylosanthes* offer promise for improved tropical pasture systems. Important gums are made from extracts of certain legume species including gum tragacanth from *Astragalus gummifer*, gum Arabic from *Acacia senegal* and *A. stenocarpa*, and tragasol from the carob (*Ceratonia siliqua*). The barks of some species of acacias (*Acacia dealbata, A. decurrens,* and *A. pycnantha*) are sometimes used as sources of tannins, chemicals that are mostly used to manufacture leather from animal skins. Some important dyes are extracted from species in the legume family. One of the world's most important, natural dyes is indigo, extracted from the foliage of the indigo (*Indigofera tinctoria*) of south Asia and to a lesser degree from American indigo (*I. suffruticosa*) of tropical South America. Derris or rotenone is a poisonous alkaloid extracted from *Derris elliptica* and *D. malaccensis* that has long been used by indigenous peoples of Southeast Asia as arrow and fish poisons. Rotenone is now used widely as a rodenticide to kill small mammals and as an insecticide to kill pest insects. Fenugreek (*Trigonella foenum graecum*) the seeds of which are used as spice in curries. Legumes include valuable fiber plant, the sunn-hemp of India (*Crotalaria juncea*) and Hemp sesbania (*Sesbania exaltata*) used by the Indians of the southwestern United States. Licorice (*Glycyrrhiza glabra*) is obtained from a legume of southern Europe and central Asia, the fibers of which have been used in making wallboard. Some legumes such as licorice (*Glycyrrhiza glabra*) and goatsrue (*Tephrosia virginiana*) have s medicinal value; many others rank among ornamental plants (for example *Lathyrus odoratus*), and legumes are of great importance for honey production.

Pea (*Pisum sativum* L.) was the original model organism used in Mendel´s discovery (1866) of the laws of inheritance, making it the foundation of modern plant genetics, but has been an object of experimental work already before Mendel (**Smýkal 2014**; **Smýkal et al. 2016 *accepted***). This might be attributed to the appearance and availability of large number of varieties with distinct traits, such as seed, pod and flower colours, seed shape, plant height etc. There were other plants with even higher variation like the cabbage family, but these were either biannual plants or displayed outcrossing pollination, apomixis and incompatibility. The first report, of using pea variation to gain insight into the transmission of traits among generations is from Thomas Andrew Knight (1759–1838). The garden pea (*Pisum sativum* L.) belongs to *Leguminosae* plant family, the third largest flowering plant family with 800 genera and over 19,000 species. *Papilionoideae* is the largest subfamily, with 476 genera and about 14,000 species (Lewis *et al.* 2005). The largest group of papilionoids, is Hologalegina, with nearly 4,000 species in 75 genera. This group includes the large galegoid tribes (Galegeae, Fabeae, Trifolieae). Tribe Fabeae Rchb. currently consists of five genera: *Lathyrus* (grasspea/sweet pea) (about 160 species); *Lens* (lentils) (4 species); *Pisum* (peas) (2-3 species); *Vicia* (vetches) (about 160-250 species) and the monotypic genus *Vavilovia formosa* (**Mikič *et al.* 2013**; **Smýkal *et al.* 2011, 2014, 2015**; Schaefer *et al.* 2012). Tribe Fabeae is considered one of the youngest groups in the legumes (Kupicha 1981; Steele and Wojciechowski 2003; Wojciechowski *et al.* 2004; Lock and Maxted 2005). A Bayesian molecular clock and ancestral range analysis suggest a crown age of 23-16 Mya, in the mid-Miocene (Lavin *et al.* 2005; Schaefer *et al.* 2012). The tribe is considered monophyletic, nested within the Trifolieae. The crown age of the *Pisum* clade is estimated to 2.3-0.8 Mya (Schaefer *et al.* 2012).

Reconstructing the phylogenetic relationship of the *Leguminosae* is essential to understanding the origin and diversification of this economically and ecologically important family. The monophyly of the family (*Leguminosae* / *Fabaceae*) as a natural group was confirmed by Kass and Wink (1996), Doyle *et al.* (1997) Lewis *et al.* (2005). Since then, molecular phylogenetic research has provided a solid understanding of relationships at all levels in the family (see Lewis *et al.* 2005 for review). Currently, based on morphological characters, the following three major groups are recognized and regarded as subfamilies: The mimosoid legumes, Mimosoideae (sometimes regarded as family Mimosaceae with four tribes and 3,270 species); the papilionoid legumes, Papilionoideae (or family Fabaceae/Papilionaceae with 28 tribes and 13,800 species); and the caesalpinioid legumes, Caesalpinoideae (or family Caesalpiniaceae with four tribes and 2,250 species) (Lewis *et al.* 2005). Estimates for the date of origin and early evolution of the legumes vary, but a rich Eocene macrofossil record shows that some lineages of the family existed by around 50 million years ago (Mya). The earliest known legume pollen remains date back to about 60-75 Mya (Lavin *et al.* 2005; Wojciechowski 2003), predating the macrofossils.

1.2 Fabeae tribe taxonomy and phylogeny

Tribe *Fabeae* (syn. *Vicieae*) is considered one of the most advanced groups in the legumes (Kupicha 1981; Steele and Wojciechowski 2003; Wojciechowski *et al.* 2004; Lock and Maxted 2005; Schaefer *et al.* 2012). It contains five genera, some of which are among the most important grain legumes: *Lathyrus* (grasspea)(about 160 species); *Lens* (lentils) (4 species); *Pisum* (peas) (2 species); *Vicia* (vetches) (about 140 species) and monotypic genus *Vavilovia formosa* (**Smýkal *et al.* 2015**). The tribe is morphologically characterized by paripinnate, often tendrillous leaves and a pubescent style or a pollen brush (Lavin and Delgado 1990). Stylar shapes and hair patterns (**Figure 1**) are one of the principal diagnostic characters within the genera of *Fabeae* (Gunn and Kluve 1976; Kupicha 1981; Choi *et al.* 2006).

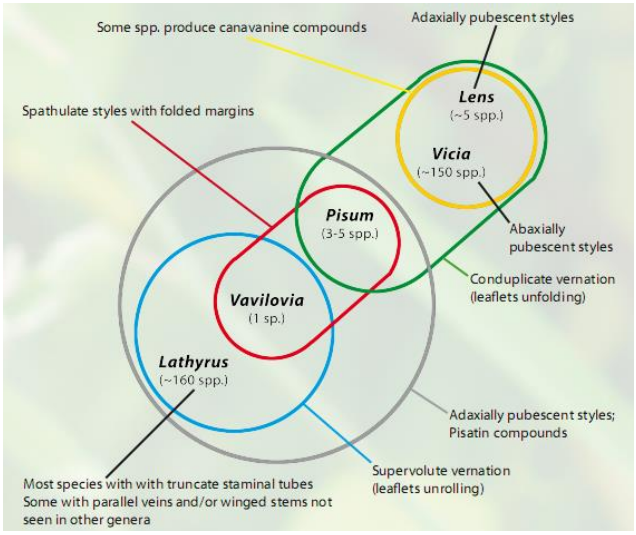


Figure 1

Key morphological and chemical character states in tribe Fabeae used in the classifications of Gunn & Kluve (1976) and Kupicha (1981) according to Kenicer 2005.

It is regarded as a rather advanced tribe in *Leguminosae* and has been considered monophyletic (Kupicha 1981; Lock and Maxted 2005; Wojciechowski *et al.* 2000, 2004; Steele and Wojciechowski 2003). However these studies, although having analyzed large number of species, were limited in coverage within the individual genus. Estimates based on rates of evolution in the *maturase* K (*matK*) chloroplast gene place the age of the crown clade at 17.5 Mya, in the mid-Miocene (Lavin *et al.* 2005). The centre of diversity and posited area of origin is the Eastern Mediterranean (Kupicha 1981; Kenicer 2007; Schaefer *et al.* 2012).

A wide range of molecular markers has been used to infer phylogenetic relationships within the group. In these, *Lathyrus* and *Pisum* formed one group, while *Vicia* was paraphyletic with *Lathyrus*, *Pisum* and *Lens* nested within it (Kenicer *et al.* 2005, 2008; Endo *et al.* 2008; **Smýkal *et al.* 2009**). *Pisum* and *Vavilovia* are both monophyletic and they in turn form a monophyletic pairing, which is sister to most

of *Lathyrus* (**Figure 2**). A comparative study of convicilin storage protein sequences of 29 species of *Fabeae* largely supports the chloroplast and ITS based phylogeny, in which *Pisum* are closer to *Lathyrus* and both are nested within *Vicia* species (Sáenz de Miera *et al.* 2008). Recent phylogenetic studies using chloroplast genes nested a monophyletic *Vicieae* within *Trifolieae*, with *Trifolium* sister to *Vicieae*. Similarly, the genus *Vicia* also appears to be paraphyletic, with *Pisum, Lens* and *Lathyrus* nested among its species (Steele and Wojciechowski, 2003; Choi *et al.* 2006; Schaefer *et al.* 2012).

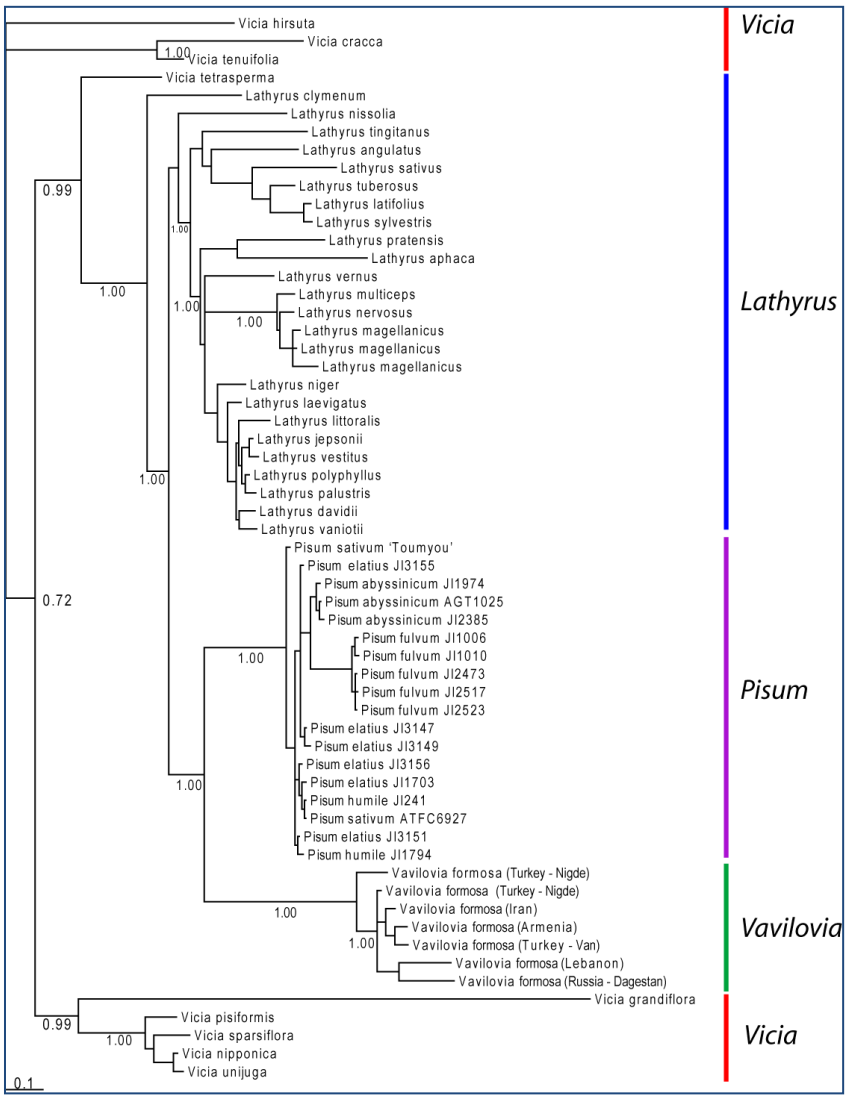


Figure 2

Bayesian analysis of cpDNA and ITS data on selected set of Lathyrus, Vicia, Pisum and Vavilovia species. adapted from Kenicer et al. 2009

Tribe Fabeae has a pantemperate distribution (excluding Australasia), with the centre of diversity in the Mediterranean and the adjacent Caucasus and Irano-Turanian floristic regions. *Lens*, *Pisum* and many species of both *Lathyrus* and *Vicia* are native to these areas. The latter two genera share

remarkably similar distributions and species numbers (Kupicha 1976, 1983) with both extending through temperate Eurasia and into the New World. Both *Lathyrus* and *Vicia* are found in temperate South America, with 20 to 30 species in each with most species in the temperate southern cone, and five or six species in each genus in the tropical Andean uplands (Schaefer *et al.* 2012). Humans have introduced at least six Mediterranean species now found naturalised in the New World and Australia.

1.2.1 *Lathyrus*

The largest genus in the tribe ***Lathyrus*** L. has the widest distribution of the genera; predominantly extra tropical in both northern and southern hemispheres. In the northern hemisphere its primary centre of diversity and inferred area of origin is the Eastern Mediterranean (Kupicha 1983, Kenicer *et al.* 2005; Schaefer *et al.* 2012). Most species of *Lathyrus* are mesophytes of open woodlands, forest margins, roadsides, with several drought tolerant, quasi-aquatic and halophytic species. The genetic diversity of the genus is of immense importance, particularly for rain-fed cropping systems (Vaz Patto and Rubiales 2014). *Lathyrus sativus, L. cicera* and *L. ochrus* are important animal fodders, with the first of these a key famine food for humans in Kenya, Ethiopia and India (Ochatt *et al.* 2007, **Smýkal *et al.* 2015**).

Unfortunately *Lathyrus* seeds, apart of being protein rich, contain the water soluble non-protein amino acid ODAP or OAP, which have been found to be neurotoxins, causing an irreversible neurological disorder, lathyrism (Barrow *et al.* 1974). Development of less toxic cultivars that retain palatability remains a holy grail in arid lands crop research as *L. sativus* has good agronomical potential (Vaz Patto and Rubiales 2014). The precise generic boundaries between *Lathyrus* and the other *Fabeae* genera have been much debated (Kupicha 1981). This taxonomic confusion has led to an abundant and complex synonymy. A relatively comprehensive molecular study of *Lathyrus* phylogeny based on nuclear (ITS) and chloroplast (*trnL-F, trnS-G*) markers using a large set of geographical and taxonomic samples was done by Kenicer *et al.* (2005). *Pisum sativum*, *Vicia cracca*, *V. nipponica* and *V. unijuga* were included for context, with the last of these constrained as outgroup. Many of Bässler's (1966, 1973), Czefranova's (1971) and Kupicha's (1983) infrageneric taxa were supported by this research. Several methods have been used to study the phylogenic relationships among different *Lathyrus* species, including morphological traits, crossability, karyotype analysis, chromosome banding and *in situ* hybridization and molecular markers (reviewed in Kumar *et al.* 2013). *Lathyrus* is predominantly a true diploid with a chromosome number of 2n=2x=14 also reconstructed as the ancestral number(Schaefer *et al.* 2012 and references therein). There are a few polyploid species among the perennials, including hexaploid (*L. palustris*, 2n = 6x = 42) and tetraploid (*L. venosus*, 2n = 4x = 28).

1.2.2 *Vicia*

The second similarly species rich genus is ***Vicia*** L. (c. 140 sp.). The genus *Vicia* is primarily Euro-Asiatic, with other distributional centres in North America, South America, East Africa and Hawaii (Kupicha 1981; Schaefer *et al.* 2012). The worldwide distribution of *Vicia* taxa is given by Allkin *et al.* (1983) and a more detailed distributional survey for *Vicia* subgenus *Vicia* has been prepared by Maxted (1991). Its distribution and species numbers parallel that of *Lathyrus*, although it might be argued that *Vicia* shows less morphological diversity. The best know species of the genus is the faba bean, *V. faba*, an important pulse, fodder crop and vegetable (**Smýkal *et al.* 2015**). Kupicha (1976) undertook the most comprehensive revision of the genus, and divided four apparently natural groups of species into two subgenera; *Ervum* and *Cracca* in her subgenus *Vicilla*, and *Vicia* and *Faba* into subgenus *Vicia*, the distinction being based on near sessile flowers and the presence of nectariferous spots on the stipules in subgenus *Vicia*. Subgenus *Vicilla* is further divided into 17 sections, subgenus *Vicia* with 38 species into 5 sections. A biosystematic study of the genus was made by Hanelt and Mettin (1989), based on morphology and classical karyology, which largely agreed with Kupicha's work. The presence of pubescence only on the adaxial side of the style define group of *Lathyrus* and *Pisum* (Kupicha 1981), on the other hand *Vicia ervilia* and *V. koeieana* and species in the genus *Lens* have also this type. Problems over the taxonomic distinction within species can be attributed to large variation in morphology and karyotypes (Maxted *et al.* 1991). In order to avoid this bias, Choi *et al.* (2006, 2008) studied the molecular phylogeny of the genus based on ITS sequences, and demonstrated that *Vicia tetrasperma* was sister to a clade composed of *Pisum* and *Lathyrus* (Choi *et al.* 2006). They also agreed with Steele and Wojciechowski's (2003) work based on the plastid gene *matK* that *Lathyrus* is monophyletic and *Vicia* paraphyletic. Moreover, Steele and Wojciechowski (2003) suggested that *Lathyrus, Lens* and *Pisum* might all be nested in *Vicia*. Although the systematics of *Lathyrus* has been carefully revised (Kenicer *et al.* 2005; Kenicer 2006) the relationship between *Pisum* and *Lathyrus* is not completely clear (Ellis 2011).

These studies used a single representative of *P. sativum*, and the authors note that the support for the separation of *Pisum* and *Lathyrus* is weak. Kenicer *et al.* (2005) placed *Pisum* close to *Lathyrus nissolia* (grass pea) which as the common name suggests, has a very unusual leaf form. It is therefore interesting to note that a distinguishing feature of *Pisum* vs. *Lathyrus* is the way in which leaflets open (*Pisum* like a book vs. *Lathyrus* unrolling). The most common diploid number for *Vicia* is 2n=14, but the aneuploids 2n=10 and 12 are common, as well as a few records of 2n=16 and 18, with an ancestral number of 2n = 14 (Schaefer *et al.* 2012 and references cited therein). Notably, cultivated *V. faba* with a chromosome number of 12 and genome size of 1C= 13.3 pg is reproductively isolated from its closest relatives. Faba bean stands as an exception among the cultivated legumes, as there is no known wild progenitor

(Kosterin 2014). *Vicia faba* subsp. *paucijuga* from Pakistan and Afghanistan and close wild relative, *V. pliniana* from Algeria (Muratova, 1931), currently considered to be only a morphological variety of *V. faba* subsp. *faba* var. *Minor*, have been suggested as the progenitors, because their showed primitive characteristics. Morphological similarity led Hopf (1973) to propose *V. narbonensis* as the faba bean ancestor; however, its crossing barriers and phylogenetic results (Schaefer *et al.* 2012) do not support this hypothesis. In summary, we do not know the faba bean progenitor and cannot be sure it is not extinct (Hanelt 1972; Schäfer 1973; Abbo *et al.* 2013) or not yet found (De Wouw *et al.* 2001; Abbo *et al.* 2013). However, perhaps the simplest explanation could be that it had a restricted natural range and habitat and so was domesticated entirely (Kosterin 2014). Once more genomic information has been gathered, more light may be shed on this question.

1.2.3 *Lens*

The lentils (***Lens*** L.), a small genus of Mediterranean origin, closely allied to *Vicia*. Close phylogenetic relationship to genus *Vicia* has been deduced from molecular marker analysis (Schaefer *et al.* 2012). Different taxonomists have recognized different numbers of species within the genus (**Smýkal *et al.* 2015**). There were considered to be five lentil species: one cultivated lentil *Lens culinaris* Medik., *L. orientalis* Popow, *L. ervoides* Grande, *L. nigricans* (M. Bieb.) Godr. and *L. montbretii* (Fisch. & C. A. Mey.) P. H. Davis & Plitm. (Cubero 1981). *Lens montbretii* has been transferred from the genus *Lens* to the genus *Vicia* on the basis of its different morphology and cytology, with 2n = 12 chromosomes (Ladizinsky and Sakar 1982). *Lens lamottei* was distinguished on the basis of an herbarium specimen of *L. nigricans* (Czefranova 1971). An additional species *L. odemensis* was recognized by Ladizinsky *et al.* (1984) as a new species due to the difference in stipules from *L. nigricans* (Ladizinsky 1986). As the last discovered taxon, *L. tomentosus* was described as distinct from *L. orientalis* by tomentose pods, a minute satellite and one large, metacentric chromosome (Ladizinsky 1997). Lentil is a self-pollinated species with cleistogamous flowers and consequently usually has <0.8% natural cross pollination (Wilson and Law 1972). All species of *Lens* have a chromosome number of 2n = 2x = 14, which is also inferred as the ancestral number for the clade (Schaefer *et al.* 2012). The genome size is estimated to be 1C = 4.20 pg, corresponding to 4,063 Mb/C (Arumuganathan and Earle 1991). Current classification recognizes one cultivated lentil (*L. culinaris* subsp. *culinaris*) and six related taxa: *L. culinaris* subsp. *orientalis, L. culinaris* subsp. *tomentosus, L. culinaris* subsp. *odemensis, L. ervoides, L. nigricans* and *L. lamottei* (Ferguson *et al.* 2000). The wild relatives of the cultivated lentil have a wide distribution. *L. culinaris* subsp. *orientalis* (Boiss.) Ponert, naturally distributed from Turkey to Uzbekistan, is considered the putative progenitor of the cultivated lentils (Ladizinsky 1979a). *Lens culinaris* subsp. *tomentosus* is restricted to northern Syria,

Iraq and eastern Turkey; *L. ervoides* occurs along the eastern Mediterranean coast to former Yugoslavia, often in shady habitats, such as pine plantations; *L. lamottei* is found in Morocco, Spain and Southern France; and *L. nigricans* occurs from southwest Turkey to the southwestern Mediterannean (Ferguson and Erskine 2001). The cultivated lentils were divided into two subspecies by Barulina (1930) and two races by Cubero (1981), the largeseeded macrosperma and small-seeded microsperma race.

1.2.4 *Vavilovia*

Commenting on: Mikic A, ***Smykal P,*** *Kenicer G et al. (2013) The bicentenary of the research on 'beautiful' vavilovia (Vavilovia formosa), a legume crop wild relative with taxonomic and agronomic potential. Botanical Journal of the Linnean Society 172: 524-531.*

Smýkal P, *Chaloupská M, Bariotakis M, Marečková L, Sinjushin A, Gabrielyan I, Akopian J, Toker C, Kenicer G, Kitner M, Pirintsos S (2016) Spacial patterns and intraspecific diversity of the glacial relict legume species Vavilovia formosa (Stev.) Fed. in Eurasia. Plant Systematics and Evolution (in revision)*

Vavilovia formosa has the typical intricate taxonomic history of Eurasian tribe Fabeae, having been placed in various genera: first as *Orobus* L. (today a synonym of *Lathyrus*) and described as *Orobus formosus* Stev. (Steven 1812), Schott and Kotschy (1856) considered it a part of *Lathyrus*. It was assigned to *Pisum* as *P. aucheri* (Jaub et Spach, 1842) or *P. frigidum* (Alef. 1860). Grossheim (1949) treated it as a new genus based on *Alophotropis formosa* (Steven) Grossh. More than a century after its discovery, it gained the status of a separate genus in tribe Fabeae as the monospecific *Vavilovia* Fed. (Fedorov 1952), in honour of N. I. Vavilov for his promotion of the importance of crop wild relatives (**Mikič *et al.* 2013, 2014**). In most regional floras, this monotypic genus was included in *Pisum* L., as *Pisum formosum* (Stev.) Alef., (Alefeld 1861; Boissier 1872; Bobrov 1972; Townsend 1974; Rechinger 1979; Makasheva 1983; Pakravan 2000; Maxted and Ambrose 2001), but it was again treated as a separated genus in other works (Davis 1970; Kupicha 1981; Lock and Simpson 1991; Lock and Maxted 2005). Segregate taxa have been established based on leaf morphology; Govorov (1937) separated two intraspecific subspecies: *Pisum formosum* subsp. *typicum* Gov. with larger 7 – 17 mm leaves with entire margins, and subsp. *microphyllum* Ser. with smaller 5 – 7 mm leaves and serrate stipules. Grossheim (1949) distinguished two species (*Alophotropis aucheri* (Jaub. & Spach) Grossh. and *A. formosa* (Stev.) Grossh., differing in obovate versus ovate leaflets. Fedorov, in the second edition of Flora of the Caucasus, also distinguished two species: *V. formosa* and *V. aucheri* with existing transitory forms (Fedorov 1952).

There are reports on variability of flower colour and leaf pubescence (resulting in separation of *P. formosum* var. *pubescens* C.C.Towsend 1968). Recently Sinjushin and Belyakova (2010) and Smýkal *et al. under revision*) have made morphometric analysis of leaves from herbarium specimens. Although differences were found between samples from different origins, with Armenian samples morphologically most diverged, no distinct intraspecific separation was observed.

Vavilovia combines the morphological traits at least of *Lathyrus* and *Pisum* genera. Among the facts that support closer relationship to pea are: 1) it has the same number of chromosomes (2n=14); 2) it has proven to be susceptible to the pea-specialized fungi; 3) its hybridization with pea is possible (Golubev 1990; **Atlagič *et al.* 2010**).

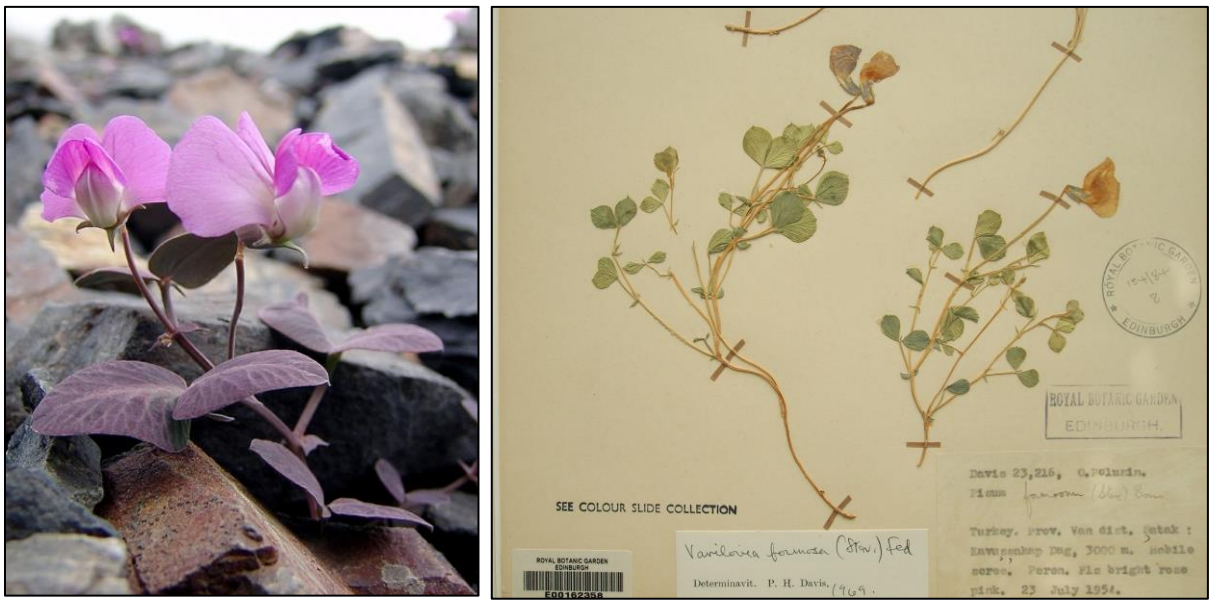


Figure 3

Flowering Vavilovia formosa found at Dagestan (A) and original herbarium specimen collected by P. Davis in 1954, deposited at herbarium of Royal Botanical Garden Edinburgh (B).

Steele *et al.* (2004) suggested for the first time a close relationship between *Vavilovia* and *Pisum* using chloroplast gene *matK*. Moreover, Lock and Maxted (2005), without analysing any DNA sequences for *Vavilovia*, postulated that the species are sister taxa, reflecting *Vavilovia* earlier classification in *Pisum*. Due to its rarity and unclear status, several recent studies have addressed its phylogenetic status (**Kenicer *et al.* 2009**; Sinjushin *et al.* 2009; Oskoueiyan *et al.* 2010; Schaefer *et al.* 2012; **Mikič *et al.* 2013**). Our results of ITS sequences in comparison to *Pisum* (*P. fulvum* JI1006, *P. abyssinicum* JI1974, *P. elatius* JI3147 and *P. sativum* JI1794) show 11 parsimony informative, 48 variable and 37 singleton sites and one 2 bp parsimony informative insertion-deletion, with the closest phylogenetically related species *P. fulvum*. In the case of *rbcL* and *trn-FL* there are 3 and 11 variable sites between *Vavilovia* and *Pisum*

within 578 or 581 bp of coding sequence respectively. In case of *trn-SG* fragment there were one 5 bp, two 2 bp, one 34 bp indels and 11 variable sites (Smýkal, unpubl. res.). Previous molecular genetic analysis based on these sequences demonstrated that the genus *Vavilovia* belongs to the same clade as *Lathyrus* and *Pisum* within Fabeae, that *Lathyrus*, *Lens*, *Pisum* and *Vavilovia* are all monophyletic, that *Vicia* is paraphyletic with all other genera of the tribe nested inside, and that *Vavilovia* is sister to *Pisum*, while both are sisters to *Lathyrus*, and that tribe Fabeae is a monophyletic relative to its neighbouring tribe Trifolieae (**Kenicer *at al.* 2009**; Schafer *et al.* 2012; **Mikič *et al.* 2013**). The results of Oskoueiyan *et al.* (2010) based on nuclear DNA (ITS) and plastid (*trn-FL, trn-SG*) DNA sequences, suggested that *V. formosa* is not distinct enough from pea and therefore they considered it a pea species, *Pisum formosum*. Other studies, based on four phylogenetically informative regions, have taken the alternative view that *V. formosa* is distinct based on its monophyly within tribe Fabeae and should be retained as a distinct genus (**Kenicer *at al.* 2009**; Schafer *et al.* 2012; **Smýkal *et al.* 2013; Mikič *et al.* 2013, 2014**) together with *Pisum, Lens*, *Vicia* and *Lathyrus* in tribe Fabeae.

Several attempts have been made at *ex situ* conservation of *Vavilovia*, especially in the former USSR, (Makasheva, 1973; Zhukovskyi 1971). Some success was achieved in the United Kingdom (Cooper & Cadger, 1990), but these did not result in the production of new seeds or in multiplication of the plants. More promising results were produced in the Vavilov Institute during 1974–1981. Some plants survived for years, bloomed and even formed fruits with seeds (reviewed in Akopian *et al.* 2010). *Vavilovia* has periodically been grown in the Yerevan Botanic Garden since 1940, as well as is being recently cultivated *in vitro*; nevertheless, this particular species in currently vulnerable to habitat destruction and climate change, and no seeds have been preserved *ex situ* to ensure its longer term conservation (Akopian *et al.* 2010; **Mikič *et al.* 2013, 2014**). We are reporting the development for the first time of a range of biotechnology approaches for *Vavilovia formosa in vitro* propagation, regeneration from callus derived from explants, protoplast isolation and culture to callus proliferation and differentiation, coupled with rooting of vavilovia shoots derived from callus or shoot cultures (**Figure 4**), and we also verified the trueness-to-type of regenerants by flow cytometry and provided the first record of its relative nuclear DNA content, compared to the model legume barrel medic (*Medicago truncatula* Gaertn.) and to the cultivated pea (*Pisum sativum* L.) (**Ochatt *et al.*** *submitted*).

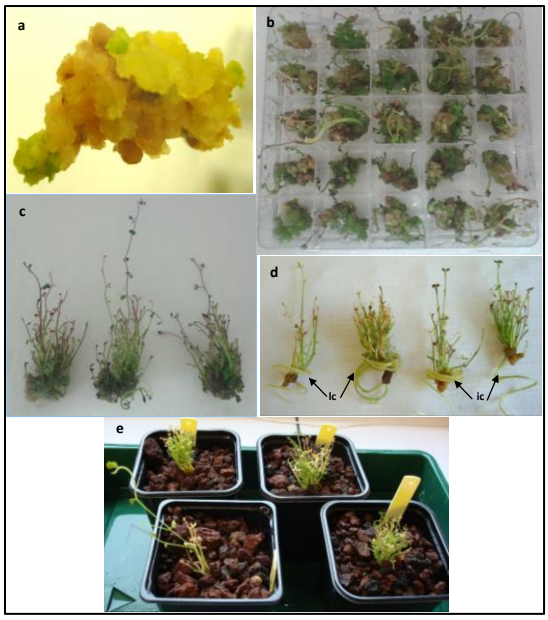


Figure 4

Plant regeneration from callus of Vavilovia formosa. A) callus derived from leaves of in vitro culture shoots; B) typical responses of shoot bud regeneration in a multiwell dish of callus derived from internode explants; C) clusters of regenerated shoot buds showing internode elongation D) a sample of rooted plants regenerated from leaf callus (lc) or internode callus (ic); E) regenerated plants after acclimatization to in vivo conditions. Ochatt et al. submitted

We have analyzed 51 *V. formosa* samples from various herbarium collections for ITS and cpDNA (*trn-FL, trn-SG, matK, rbc*L and *psbA-trnH*) sequences (**Smýkal *et al.*** *in revision*). These represent in total up to 2551 bp of chloroplast and 664 bp of nuclear sequences per sample. These were complemented with published sequences from 5 samples (Sinjushin *et al.* 2009; Oskoueiyan *et al.* 2010; Schaefer *et al.* 2012) along with samples of the two populations we analysed in detail. These represent in total 17 samples from Turkey, 6 from Iran, 17 from Armenia, 1 from Lebanon and 11 from Caucasus (Georgia, Russia and Daghestan). The only polymorphism found was in the *trnK* 3´intron with a 7 bp indel (GATTGGT). This has separated two haplotypes, 13 with this as insertion originating from Armenia, Daghestan, Nakhichevan and Iran (**Figure 5**), while the remaining samples had deletion (all Turkish and part of Armenian). In case of 664 bp of ITS sequences, in 8 of samples we detected G instead of A at position 500 bp. There was no clear geographical assignment of these samples.

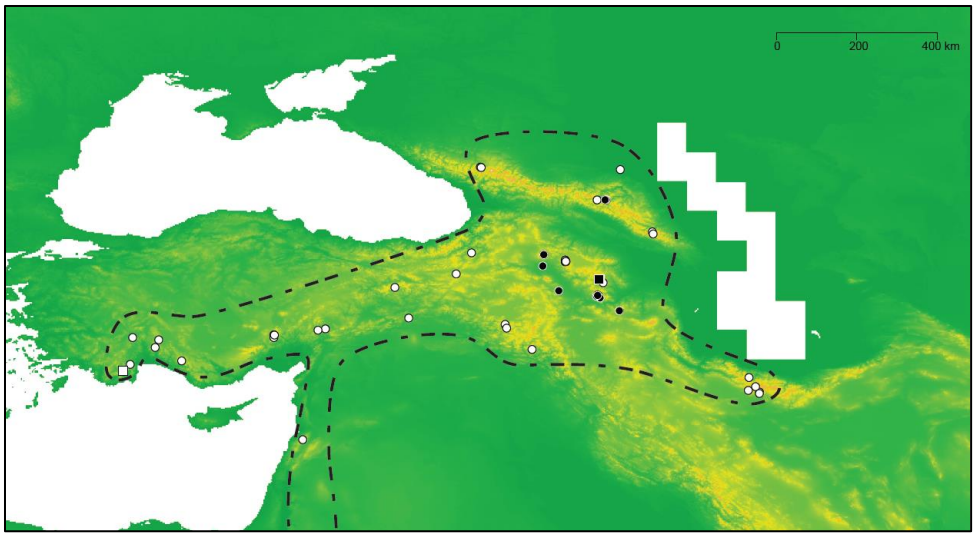


Figure 5

Specimen distribution map with indication of two matK detected haplotypes. In squares are labelled studied populations in Turkey (white with black outline) and Armenia (black with white outline). White space corresponds to sea water areas. Dashed line shapes the recorded distribution area of the species.

Two population samples were studied in details (Smýkal *et al. in revision*). One population represented by 30 samples originated from Mt. Ughtasar, Armenia from locality between 3,305–3,453 m a.s.l. in area of about 75,000 m^2 on the south-western slope and corresponded to study of Akopian *et al.* (2010). These habitats are well sunlit and sun-warmed, and due to volcanic nature of rocks it retains heat during the night. There is clear altitudinal optimum, with upper limit likely dictated by temperature e.g. length of vegetation period, while the lower seems to be related to competing plant species once they stabilize the volcanic gravel. We have not confirmed the statement of Akopian *et al.* (2010) that plants reproduce mainly clonally. Plants 1 m apart are already genetically distinct, as revealed by AFLP. The rough estimate of population size is between 1,000 to 3,000 individuals, i.e. one plant per 25 to 75 m^2. There is however no homogenous distribution within the area, but plants are in patches of 6 to 20 m^2. They seem to be distributed in vertical lines corresponding to edges of moving gravel. The second population was located on north-western slopes of Mt. Kizlar Sivresi, western Taurus, Turkey at an altitude between 2,000 to 2,150 m in limestone scree (Denüz and Sümbül 2004). This is by our knowledge the most western distribution range of *Vavilovia*.

A total of 143 unambiguous AFLP fragments were obtained from the set of 47 *V. formosa* samples (Smýkal *et al. in revision*). The total gene diversity was high (H_T = 0.252) and reflect the differences between the two populations each from one site of the geographical range. The mean intrapopulation gene diversity was low (H_E = 0.088) with lower H_E value observed for Turkish than in Armenian

population. Also other genetic variability values were slightly lower in Turkish population, but this is likely caused by imbalanced sampling. The hierarchical AMOVA analysis revealed that most of the observed molecular variance (78.96 %) is present among the two populations, while a minor part (21.04 %) represents variation observed within populations. Both clustering methods (PCoA and Bayesian clustering) clearly divided the analysed *V. formosa* samples into two major groups according their population origin (**Figure 6**).

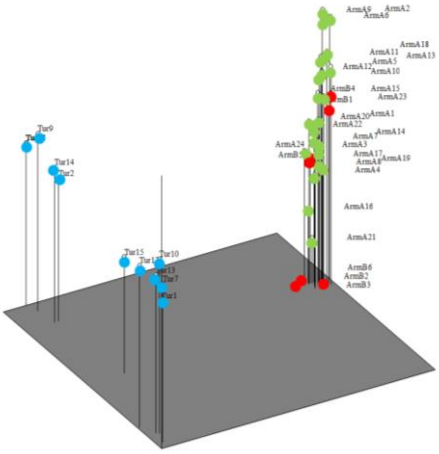


Figure 6

Principal coordinate analysis of pair-wise Jaccard's similarity matrix of 17 Turkish (TUR) and 30 Armenian (ARM) Vavilovia formosa samples. The first three principal coordinates accounted for 71.65% of the total variation

In case of Armenian population composed of two subpopulations, there were similar but distinct. Finally, significant genetic differentiation between the two populations was observed, with the genetic differentiation measure, Fst = 0.78, being significantly different from zero.

The results of the niche modelling support the hypothesis that *V. formosa* survived and expanded as a cryoxerophilic species in steppe and shrub tundra vegetation during LGM. The niche modelling results indicated that *V. formosa* was favored during the LGM compared to the Last Interglacial Period (~ 130,000 ybp) and to the current environmental conditions (**Figure 7**).

These findings are in line with the view that oroxerophytic floristic elements of Caucasus are not only Boreal and Arcto-Alpine elements, which penetrated into this area during Pleistocene, but part represents ancient elements of autochthonous origin (Nakhutsrishvili 2013 and references therein). This view supports the hypothesis that *V. formosa* had been formed in the Miocene-Pliocene as a more thermophilous element, and after, in the Pliocene-Pleistocene the element fully changed in the direction of cryo- and xerophytization. *V. formosa* has suffered a range reduction following climate warming after LGM, which testify and classify this species as cold-adapted among the Fabeae species.

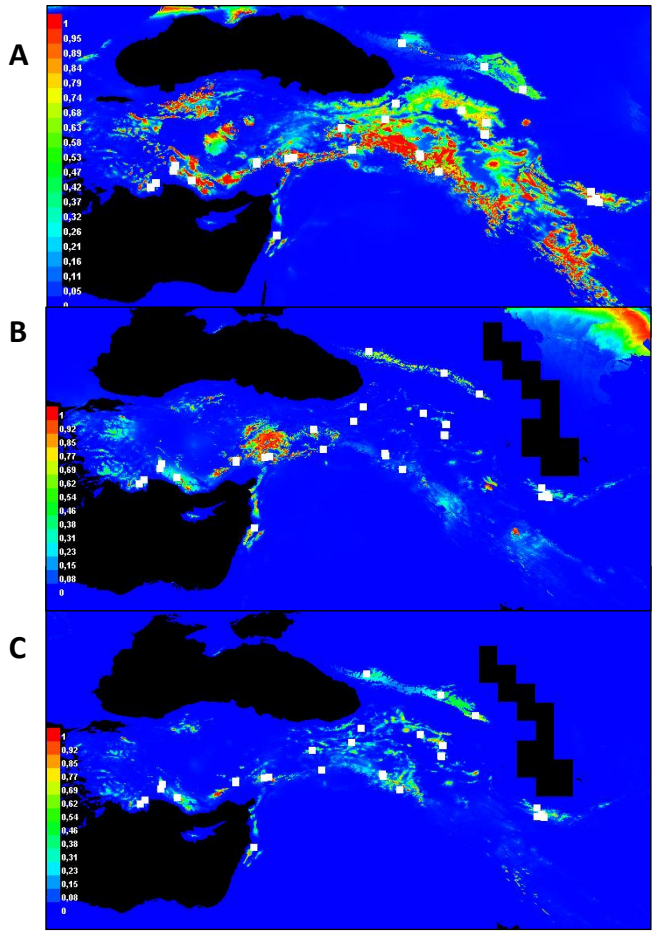


Figure 7

Predicted distribution of V. formosa in past Last Glacial Maximum (A), Last Interglacial (B) and present (C) climatic conditions. Predictions resulted from applying thresholds on the outputs of maxent models predictions. For Last Glacial Maximum three separate global climate models were used, and green to red areas represent predicted presence for a higher number of the used models.

This species is also classified as a Glacial relict following Hampe and Jump (2011), who define that Glacial relicts were more widely distributed during Quaternary cold stages and have experienced significant range contractions since the LGM. This finding is also supported by the results of Joannin *et al.* (2010), who report that in the regional Caucasus arboreal plants mainly developed during interglacials whilst herbaceous ones were dominant during glacials. Despite there is no doubt that the climate in the area of occurrence of *V. formosa* is colder than surrounding regions and the species as a cold-adapted species has been able to survive unusual warm periods in these refugia (sensu Hampe and Jump 2011) in the last ca 10,000 years (see also Ohlemüller *et al.* 2008), it remains open whether other factors can likewise exert significant effects on local scale, such as rhizobia, which are well known for their contribution in mineral nutrition of legumes especially in severe and drought conditions (Zahran 1999). Taking also into consideration that range reduction of a species is most likely to cause loss of intraspecific genetic diversity (Aguilar *et al.* 2008; Alsos *et al.* 2012), we can hypothesize that *V. formosa* has suffered a loss of intraspecific genetic diversity. Of course the loss of intraspecific genetic diversity due to range reduction is not a simple case, as many parameters are involved in this relationship, such as the distance dispersal.

1.2.5 *Pisum*

Commenting on ***Smýkal P,*** *Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, Ford R, Coyne CJ, Maxted N, Ambrose MJ, Ellis THN (2011) Phylogeny, phylogeography and genetic diversity of the Pisum genus. Plant Genetic Resources 9: 4-18.*

Jing R, Vershinin A, Grzebyta J, Shaw P, ***Smýkal P****, Marshall D, Ambrose MJ, Ellis THN, Flavell AJ (2010) The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. BMC Evolutionary Biology 10: 44.*

Phylogenetic relationships of *Pisum* within tribe Fabeae was clarified by Schaefer *et al.* (2012) based on analysis of several cpDNA and ITS regions sampling of 262 of the c. 380 species currently accepted in the tribe. In this study, the pea *Pisum sativum* s. l. is sister to *Pisum fulvum* Sibth. & Sm. and both are sister to *Vavilovia formosa* (Stev.) Fed. The genus *Pisum* is distinguished morphologically from the related genera *Lathyrus* and *Vavilovia* by the presence of large, leafy stipules, which are semi-amplexicaul. The genus *Pisum* contains the flavonoid phytoalexin pisatin, which is shared by the genus *Lathyrus* (**Figure 1**) but not found in *Vicia* species (Bisby *et al.* 1994) which contain wyerone instead. The genus *Pisum* L., originally described to be distinct from *Lathyrus* L. (Linnaeus 1753), has recently been shown to be included in the *Lathyrus/Vicia* complex (Schaefer *et al.* 2012) rendering *Lathyrus* to be non-monophyletic. This leaves two options based on the current data: either *Pisum* has to be subsumed into synonymy in a larger *Lathyrus,* or *Lathyrus* has to be divided into more than one genus (G. Lewis personal communication). Interestingly Lamarck (1778), who was aware of Linnaeus`s description, designated pea as *Lathyrus oleraceus* Lam., the name which we might eventually return to after generic re-circumscription of the tribe (**Smýkal *et al.* 2015**). Chemosystematic studies were made by Harborne (1971) and Pate (1975). Studies of leaf flavonol glycosides showed that *P. fulvum* contains quercetine 3-glucoside, primitive cultivars from Nepal and *P. abyssinicum* contain kaempferol and quercetine 3-sophoroside, while modern pea cultivars contain kaempferol and quercetine 3 (coumaroyl-sophorotrioside). Harborne (1971) reported that petals of wild peas contain delphinidin, petunidin and malvidin 3-rhamnoside-5-glucosides, while petals of garden pea contain in addition pelargonidin, cyanidin and peonidin 3-rhamnoside-5-glucosides. Unfortunately, the yellow petals of *P. fulvum* were not studied. Electrophoretic patterns of albumin and globulin (Waines 1975) and chloroplast DNA polymorphism (Palmer *et al.* 1985) have led taxonomists to consider *P. fulvum* to be a distinct species and *P. sativum* to be an aggregate of *P. humile, P. elatius* and *P. sativum*. The work of Ben-Ze'ev and Zohary (1973) has become the standard text for pea species relationships and was based on classical

species definitions using hybridization barriers along with ecological aspects of distribution. Their work followed the taxonomy of Boissier (1872), which recognized three wild pea species: *P. fulvum, P. elatius* and *P. humile*, and the domesticated pea *P. sativum*.

As a direct result of broad phenotypic diversity, a large number of different Latin names at different ranks have been proposed for various forms of pea (Ellis 2011; **Smýkal *et al.* 2011**) and the classification of *Pisum* L., based on morphology and karyology, has changed over time from being considered a genus with five species (Govorov 1937), three taxa recognised in comprehensive study of Makasheva (1979) to a monotypic genus (Lamprecht 1966; Marx 1977). Later Davis (1970) and Kupicha (1981) recognised two species, *P. fulvum* Sibth. & Sm. and *P. sativum* L. The *P. elatius* M. Bieb. was first described at the rank of species in 1808, and first reduced in rank to a subspecies by Schmalhausen (1895), although many authors ascribe the down-ranking to Ascheron and Graebner (1910). *Pisum humile* was described by Boissier and Noë (1856) but their name is illegitimate because it is a later homonym of *Pisum humile* Miller (1768), a form of cultivated pea. Berger (1928) downgraded the rank of the taxon to that of subspecies and gave it a new name: *P. sativum* subsp. *syriacum* A. Berger, but its status was raised again to species by Lehmann (1954) as *Pisum syriacum* (A. Berger). C.O. Lehm., though this nomenclatural change remained unsupported. Ben-Ze´ev and Zohary (1973) suggested that there are two wild populations of *Pisum*: *P. sativum* subsp. *elatius* Bieb. and *P. humile* Boiss & Noë (= *P. syriacum* (A. Berger) C.O. Lehm.). These two wild groups were described as being morphologically, ecologically and genetically distinct. According to Ben-Ze'ev and Zohary (1973), the two taxa differ by internode length, peduncles and pods and flower sizes. Makasheva (1979) characterised these differences numerically. However, all these differences are quantitative and might simply represent adaptations to two types of habitat. These subspecies were shown to be polyphyletic by Ben-Ze'ev and Zohary (1973), where two karyological classes coincided only partially with morphological characters. This led Townsend (1968) and Davis (1970) to consider all wild forms of *Pisum sativum* as belonging to the same subspecies under the priority name *Pisum sativum* subsp. *sensu lato*. In the review of Yarnell (1962) were considered *P. humile* and *P. sativum* to be conspecific, even though they differ by chromosomal inversions and translocations. Recently, Ladizinsky and Abbo (2015) recognised two species: *P. fulvum* Sibth.&Sm. and *P. sativum* L., the later divided into three subspecies: the domesticated pea subsp. *sativum* and two wild forms: subsp. *elatius* (M. Bieb.) Asch. & Graebn. and subsp. *humile* (Holmboe) Greuter, Matthäs & Risse. They delimited the southern form as subsp. *humile* var. *humile* (Boiss et Noë) Ladizinsky, and the northern form as subsp. *humile* var. *syriacum* (A. Berger) Ladizinsky. Mostly small and restricted populations of wild pea (*Pisum sativum* subsp. *elatius* and *Pisum sativum* subsp. *humile* as recognized by Ladizinsky and Abbo (2015) are scattered over a great area of the Mediterranean basin in the broad sense, from Portugal in the west to Iran in the east and from Hungary in the north to

Tunisia, Morocco and Jordan in the south, with the greatest diversity in the Near East (Turkey, Syria, Israel), the center of pea diversity, while the distribution of *P. fulvum* Sibth.&Sm. is restricted to the Middle East (**Smýkal *et al.* 2011**, **2015**; Ladizinsky and Abbo 2015).

P. sativum* subsp. *elatius (M. Bieb.) Asch. & Graebn. grows as a tall climber (up to 3 m) in humid forested valleys from the Caspian coast through the Caucasus to the Mediterranean region, including its islands and northern African coast, extending north to the Black Sea coast and Hungarian plains. It is found at altitudes from 0 to 1700 m above sea level (asl) (Maxted and Ambrose 2001). It has large (20–30 mm), often bicolour flowers and long peduncles (2–4× longer than stipules) most often with two flowers (1–3), producing large pods (50–80×10–12 mm). Leaflets are two to four paired, ovate-elliptic, entire or subdentate. This subspecies has a chromosomal translocation difference from cultivated *P. sativum*, but it is interfertile, although some nucleo-cytoplasmatic conflict has been reported in specific crosses (Bogdanova, Galieva and Kosterin 2009).

P. sativum* subsp. *humile (Holmboe) Greuter, Matthäs & Risse is former subspecies *pumilio* (*P. sativum* subsp. *elatius* var. *pumilio*) has shorter internodes (20–40 cm stem length), shorter peduncles, smaller (40–45×7–10 mm) often pigmented pods and small flowers (15– 18 mm). It is distributed from the Mediterranean through Turkey, Syria and Israel to Iran in steppe habitats. Compare to *Pisum subsp. elatius* found in higher altitudes, from 700– 1800 m at least in Syria (Maxted and Ambrose 2001). Comparison of data from the expeditions to Syria and previous herbarium passport data from Turkey reveals differences in circumstances. For example, in Syria discrete variation exists in altitude, rainfall and parent rock or soil type, correlated with an allopatric association between subsp. *elatius* and var. *pumilio*. However, in Turkey, where these varieties have been found sympatric, mild and overlapping climatic conditions have been reported (Mumtaz *et al.* 2002).

P. fulvum Sibth. & Sm. is distinguished by its weak slender stems (10–45 cm), one to two paired dentate leaflets, peduncle as long as the incised-dentate stipules, usually with single small (10–15 mm), yellow to orange flowers. Pods are small (30–40×5–10 mm) and pigmented, seeds are dark brown to velvet black with subpapillose testa. Some *P. fulvum* accessions possess amphicarpic character, with basal pods growing into the ground. It grows on open arid (300–450 mm annual rainfall) rocky limestone slopes (30–1500 m asl).

P. abyssinicum A.Br. has been described from Ethiopia and Yemen as cultivated type. It is 30–60 cm tall, with ovate, obtuse, irregularly dentate 4–5 cm long stipules up to the top and also along the inner margin, with semicordate acute basal lobes. The stipules are as long as internodes. Peduncles are shorter (1/3 to 1/2) than the stipules at the time of flowering, but prolonged thereafter, one-flowered

with small flowers. Flowers are pale, calyx lobes narrow lanceolate, standard only half open, whitish, wings shorter bright or pale purple-red, keel shorter than wings and narrow. Pods 40–50 mm long, with four to six seeds. Seeds globular-cubic, brownish red, violet, brown or grayish green. Most with one pair of leaflets and branched tendrils. Leaflets ovate, elliptical or obovate, obtuse, mucronulate, sharply or incisely dentate except of lower third, 3–4 cm long. Entire plants often have a bluish green colour.

Cultivated pea is attributed to ***P. sativum* subsp. *sativum***, diagnosed by characters resulting from domestication, namely: non-dehiscing pods, absence of seed dormancy (Abbo *et al.* 2013, **Smýkal *et al.* 2014**) and seeds without a rough testa. Some authors continue to recognise other subspecies such as *P. sativum* subsp. *asiaticum* (Govorov), a vague aggregate of forms from Egypt to Central Asia, *P. sativum* subsp. *transcaucasicum* (Govorov), a vetch-like fodder crop from Transcaucasia, and *P.s.* subsp. *abyssinicum* from Ethiopia (Govorov 1937; Makasheva 1979; Kosterin and Bogdanova 2008; Westphal 1974). The actual diversity of wild forms of *P. sativum*, as well as the associated taxonomy thus has a confusing history.

Intra-generic relationships with in the genus *Pisum* have traditionally been assessed morphologically (Lehman 1954), by using seed proteins (Waines, 1975), flavonoids (Harborne 1971; Pate 1975), allozymes (Hoey *et al.* 1996,), chloroplast DNA polymorphism (Palmer *et al.* 1985; Polans and Moreno, 2009) and gene derived sequences (Jing *et al.* 2007; Zaytseva *et al.* 2015). All these studies separated *P. fulvum* as a distinct species and *P. sativum* as an aggregate of *'P. humile', P. sativum* subsp. *elatius* and *P. sativum* subsp. *sativum*. Serological studies of *Pisum* taxa by Kloz (1963) indicated a close relationship of all studied taxa, except for *P. fulvum* and *P. abyssinicum*. He was possibly the first to indicate that *P. abyssinicum* might have originated from hybridization between *P. sativum* subsp. *elatius* and *P. fulvum*. The possible hybrid origin of *P. abyssinicum* was also revealed by retrotransposon based diversity analysis (Ellis *et al.* 1998; Vershinin *et al.* 2003; Jing *et al.* 2010) and is clearly shown in our recent genome-wide study (**Smýkal *et al.* 2016** *submitted*). Zaytseva *et al.* (2015) have detected rare intragenic recombination events in histone H5 subtype between *P. fulvum* and *P. sativum* subsp. *elatius* in geographically limited samples. *P. abyssinicum* histone sequences were placed within one of the two *P. sativum* subsp. *elatius* groups in this study. Due to the presence of indehiscent pods, moderately large seeds and a lack of seed dormancy, it has been identified as partially domesticated. The intriguing question is: was it domesticated independently? The crosses between *P. abyssinicum* and cultivated *P. sativum* did not show any segregation in domestication traits, suggesting that identical loci/genes are involved (Holden 2009). In this study, evidence of separate domestication was inferred from fixation of early domestication traits at separate loci in the two species. This approach yielded mixed results: some domestication traits were found to segregate transgressively, others segregated non-transgressively in both populations, and

others showed different segregation patterns in the two mapping populations. In summary, observed transgressive segregation of domestication traits (seed weight and seed number) in the wide crosses suggests that these taxa do not share domestication history (Holden 2009). This question can be solved once all of the genes involved in pea domestication are identified (Weeden 2007). *P. abyssinicum* has been used as a bridge between *P. fulvum* and *P. sativum* as it crosses reasonably well with both (Kosterin and Bogdanova 2014; **Warkentin *et al.* 2015**) supporting further the hybrid origin. Therefore *P. abyssinicum* qualifies for species status on the basis of phenotype (early flowering and strongly serrate leaflets) and biological isolation (see **Warkentin *et al.* 2015**). Based on our comprehensive DNA analysis, we proposed to keep species rank for *P. abyssinicum* as suggested by Maxted and Ambrose (2001). Another discussed group of cultivated peas are the so called Afghan types or *P. sativum* subsp. *asiaticum* as defined by Govorov (1937) and Makasheva (1979). These authors further subdivided this group into 34 varieties and convarieties, based largely on geographical origin and seed characters. Afghan pea types were rarely included in published analyses, except in Tar´an *et al.* (2005), Jing *et al.* (2010) and Kwon *et al.* (2012) where they formed a separate cluster. This group distinction is further supported by the requirement of specific *Rhizobium* strains (Young and Matthews 1982) due to *sym2* mutant recessive allele of nodulation factor (Lie *et al.* 1984). A biogeographical study of genus *Pisum* was conducted by Kosterin and Bogdanova (2008) using a combination of mitochondrial (*cox1*), chloroplast (*rbcL*) and nuclear (seed albumin SCA) genes. Analysis of 47 wild and 42 cultivated peas revealed that all accessions of *P. fulvum* and *P. abyssinicum* had combination A, the majority of cultivated forms of *P. sativum* had combination B while wild representatives of *P. sativum* subsp. *elatius* had both combinations A and B. Updated study with more accessions from central and western Mediterranean region by Kosterin *et al.* (2010) proposed following scenario for the evolution of wild pea. Combination A was the ancestral state of the genus and was inherited by the early *P. sativum* originating from the eastern Mediterranean, based on the present area of the lineage A in Israel. Here *P. sativum* grows sympatrically with *P. fulvum*, which also has combination A (**Figure 8**).

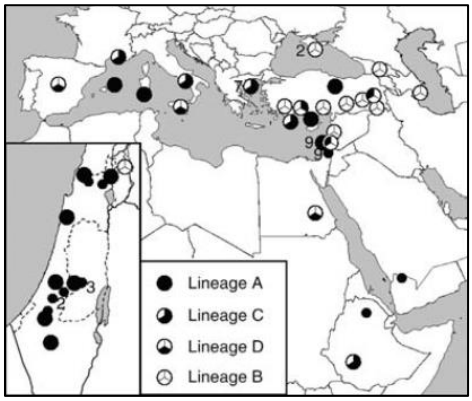


Figure 8

Phylogeography of Pisum sp. based on three molecular makers (cox1, rbcL and albumin protein) indicating A, B, C and D lineages based on marker allele combinations. Smýkal et al. 2011

P. abyssinicum, another species with exclusively combination A, occurs in Yemen and N Ethiopia, that is also in East Mediterranean sensu lato. From this area, lineage A spread within wild *P. sativum* to the west over the Mediterranean. The two accessions with combination A found on Sardinia and Menorca may represent island refugia of that early spread. During Pleistocene, the seal level oscillated and the islands were repeatedly isolated and merged again. It was proposed that the westward spread of line A occurred during the Pleistocene climate coolings, when the sea occupied less area. During this westward dispersal, a loss of restriction site in *coxI* was fixed, giving rise to lineage C (**Figure 8**) which spread over Central and West Mediterranean and north east Africa (Kosterin *et al.* 2010). However, these three genes cannot fully reflect the diversity pattern, as individual genes might have different evolutionary trajectories (Jing *et al.* 2007; Burstin *et al.* 2015). This is evident from our recent study where neither cpDNA nor ITS haplotypes match the genome-wide DArTseq pattern and/or gene-based SNP assay (Smýkal *et al. submitted*). The possible hybrid origin of *P. abyssinicum* was revealed by retrotransposon based diversity analysis (Ellis *et al.* 1998; Vershinin *et al.* 2003; **Jing *et al.* 2010**) and is clearly shown (**Figure 13**) in our genome-wide study (Smýkal *et al. submitted*). Interestingly, Zaytseva *et al.* (2015) have detected rare intragenic recombination events in histone H5 subtype between *P. fulvum* and *P. sativum* subsp. *elatius* in geographically limited samples. *P. abyssinicum* histone sequences were placed within one of the two *P. sativum* subsp. *elatius* groups in this study. Independent of the taxonomic status assigned, wild peas comprise a broad continuum of forms (**Jing *et al.* 2012**) with a variable degree of reproductive isolation among representatives of wild and cultivated peas (Ben-Ze'ev and Zohary 1973; Bogdanova and Berdnikov 2001; Bogdanova and Kosterin 2006; Yadrikhinskiy and Bogdanova 2011; Bogdanova *et al.* 2014).

1.3 Pea genome

*Commenting on **Smýkal P**, Aubert G, Burstin B et al. (2012) Pea (Pisum sativum L.) in the Genomic Era. Review. Agronomy 2: 74- 115.*

***Smýkal P,** Kalendar R, Ford R, Macas J, Griga M (2009) Evolutionary conserved lineage of Angela-like retrotransposons as a genome-wide microsatellite repeat dispersal agent. Heredity 103: 157–167.*

Despite their close phylogenetic relationships, crop legumes differ greatly in their genome size, base chromosome number, ploidy level, and reproductive biology. Nevertheless, early studies indicated that members of the Papilionoideae subfamily exhibit extensive genome conservation based on comparative genetic mapping (Kalo *et al.* 2004). To establish a unified genetic system for legumes, two legume species in the Galegoid clade, *Medicago truncatula* and *Lotus japonicus*, from Trifolieae and Loteae tribes,

respectively, were selected as model systems for studying legume genomics and biology (Cook 1999, Sato *et al.* 2008). Unlike many of the major crop legumes, *M. truncatula* and *L. japonicus* have small genome size, are amenable to forward and reverse genetic analyses, and are, therefore, well suited to biological inquiries important to crop legume species. For cultivated pea, nuclear genome size estimates have been produced for several accessions using different methods (Dolezel and Greilhuber 2010) and estimated to be 9.09 pg DNA/2C corresponding to the haploid genome size (1C) of 4.45 Gbp. The average GC content is 37.4% and approx. 30% C residues are methylated (Pradhan & Adams, 1995). Early studies of sequence composition of the pea genome employing DNA reassociation kinetics and melting behavior measurements indicated that its large part (75-97%) is made up of repetitive sequences, being confirmed recently by next generation sequencing approach (Macas *et al.* 2007). The *Ty3/gypsy* LTR-retrotransposons were identified as the main component of the pea repeats, with highly amplified group of Ogre elements alone representing 20-33% of the pea genome. Another interesting lineage of *copia*-type *Angela*-family retrotransposon, has been shown to be evolutionary conserved and involved in microsatellite repeat dispersal (**Smýkal *et al.* 2009**). Some of these elements were found useful as a source of molecular (**Smýkal 2006**) or cytogenetic markers allowing discrimination of individual chromosomes within the karyotype (Flavell *et al.* 2003; Jing *et al.* 2005; Neumann *et al.* 2002).

The standard pea karyotype comprises seven chromosomes: five acrocentric chromosomes and two (4 and 7) with a secondary constriction corresponding to the 45S rRNA gene cluster. However the numbering of pea chromosomes is unconventional. The largest chromosome is conventionally called 5 rather than the usual 1. There is no simple solution to this problem, in part because the two small , submetacentric chromosomes (1 and 2) are impossible to distinguish in terms of relative size and arm length ratios (Hall *et al.* 1997). A set of translocation stocks was generated by Lamm and Miravalle (1959) but there was considerable disagreement about which linkage groups and chromosomes were involved (Lamm 1977, 1983; Folkeson 1990). For this reason the chromosome numbers and linkage group numbers are referred to using Arabic and Roman numerals respectively (1 = VI, 2 = I, 3 = V , 4 = IV, 5 = III, 6 = II and VII = 7)(Fuchs *et al.* 1998, Neumann *et al.* 2002) **Figure 28**. The JI145, JI146 and JI148 lines with reconstructed karyotypes were used for flow sorting of individual pea chromosomes with over 95% purity suitable for PCR-based physical mapping in pea (Neumann *et al.* 2002). Therefore, the only mean to reliably distinguish between all chromosome types is to label the chromosomes with markers showing chromosome-specific FISH pattern. Pea centromeres exhibit a unique structure consisting of remarkably extended primary constrictions containing multiple CenH3 domains and designated as "meta-polycentric" (Neumann *et al.* 2012, 2016). Such chromosomes were also found in its sister genus *Lathyrus*, contrary to the closely related genera of *Vicia* and *Lens* which possess monocentric chromosomes (Neumann *et al.* 2015).

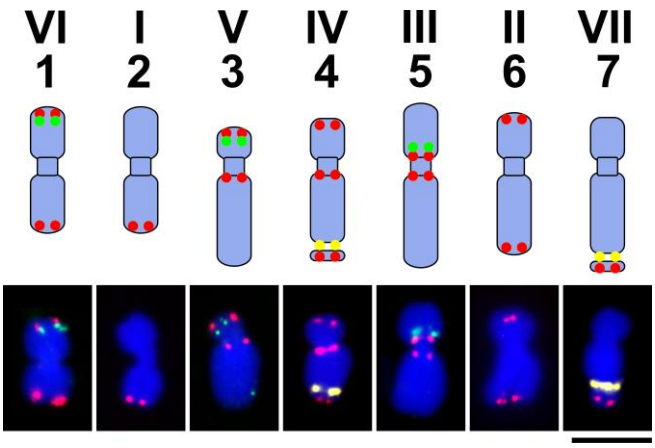


Figure 28

The pea karyotype. *Arabic and Roman numerals refer to chromosome type and linkage group, respectively, as assigned by Neumann et al. 2002 and Fuchs et al. 1998. The upper panel shows a scheme of the pea karyotype with the loci for PisTR-B (red), 5S rDNA (green), and 45S rDNA (yellow). The bottom panel shows the same loci detected by FISH on isolated metaphase chromosomes. Bar = 5 µm. Smýkal et al. 2012*

There is a long history of genetic mapping studies in pea. Different types of polymorphisms were successively used: morphological markers, isozymes, RFLP, RAPD, SSR, EST-based, PCR-based techniques and, more recently, high-throughput parallel genotyping (reviewed in **Smýkal and Konečná 2014**). Three different crosses (Terese/K586, Champagne/Terese, Shawnee/Bohatyr) were used to build a composite genetic map of 1430 cM (Haldane) comprising 239 microsatellite markers (Loridon *et al.* 2005). These markers are evenly distributed throughout the seven linkage groups of the map with 85% of intervals between the adjacent SSR markers being smaller than 10 cM. This map was used to localize numerous QTLs for disease resistance as well as quality and morphology traits. More recently, functional maps composed of genes of known function were developed (see **Smýkal *et al.* 2012** for review). The consensus map was published by Bordat *et al.* (2011), it includes 214 functional markers, representing genes from diverse functional classes and is linked to previous maps. The most recently, high density map using pea 13.2k SNP assay (Tayeh *et al.* 2015) was developed. This map is based on annotated genes and can be effectively used in homology and synteny based searches.

Translational genomics is also beginning to assist identification of candidate genes or saturating markers in a zone of interest of pea. For example, candidate genes responsible for two floral zygomorphy mutant loci in pea, *Keeled Wings* (*K*) and *Lobed Standard* 1 (*LST1*), were identified using genomic information from *L. japonicus* (Wang *et al.* 2008). Similarly, the flowering locus GIGAS was identified using a candidate gene approach in comparison with *M. truncatula* (Hecht *et al.* 2011). In order to support comparative legume biology, several databases were developed, integrating genetic and physical map data and enabling *in silico* analysis (see **Smýkal *et al.* 2012** for review). A candidate gene approach based on comparative genome analysis is being used to study genes controlling flowering time in pea. Wild *P. sativum* ssp. *elatius* and a subset of pea landraces and winter cultivars do not flower at all under short photoperiods, but this long-day requirement has been genetically relaxed in a majority of cultivated lines. Up to six loci contribute to 'natural' variation related to flowering in pea, with derived or cultivated alleles generally conferring earlier flowering and a reduction in photoperiod

response. In addition, numerous other loci have been identified through mutational studies (Weller *et al.* 2009). A "functional candidate" approach has also been used to clone the photoperiod response locus *Hr*. *Hr* was originally defined as a major locus controlling flowering time and showing Mendelian inheritance under controlled short photoperiod conditions, with recessive *hr* alleles causing reduction but not complete loss of the response to photoperiod (Weller *et al.* 2009). More recently, comparison of *Hr* and *hr* genotypes revealed defects in circadian rhythms, and comparative mapping of circadian clock genes has identified *Hr* as the pea ortholog of *Arabidopsis ELF3* (Weller *et al.* 2012).

1.4 Genomic analysis of pea

Although pea is amenable to genetic transformation, similarly to most of the legumes crops this remains a challenge and precludes systematic characterization of gene functions (Somers *et al.* 2003, **Švabová *et al.* 2005**). In spite of this, co-cultivation process was elaborated and several successful pea transformations were published (Švabová and Griga 2008; Atif *et al.* 2013). In addition to *Agrobacterium*-mediated, direct gene transfer methods such as electroporation of isolated pea protoplasts (Puonti-Kaerlas *et al.* 2000) and biolistic (Warkentin *et al.* 1992) Recent review of legumes transformation has summarized successful and published pea transgenosis. Despite the fact that pea transformation was reported over 20 years ago its efficiency remains low (in range of 0.1 to 6,5%) (Atif *et al.* 2013). Virus-induced gene silencing (VIGS) has become an important reverse genetics tool for functional genomics and VIGS vectors based on *Pea early browning virus*, *White clover mosaic virus, Bean pod mottle virus* are available for legume species and were successfully used to silence pea genes (Constantin *et al.* 2004; Luo *et al.* 2013; Ido *et al.* 2012; Meziadi *et al.* 2016). The genomics tools such as fast neutron and TILLING mutant populations were developed for reverse genetics approaches (Dalmais *et al.* 2008; Wang *et al.* 2008; Hofer *et al.* 2009). The TILLING method combines the induction of a high number of random point mutations with mutagens like ethyl methanesulfonate (EMS) and mutational screening systems to discover induced mutations in sequence DNA targets (McCallum *et al.* 2000). Two sufficiently large pea TILLING population were made in cv. Cameor and cv. Terese backgrounds (Dalmais *et al.* 2008, Triques *et al.* 2007). The population currently has 4702 M2 families for cv. Cameor, resulting in 4817 lines of which 1840 have been characterized for phenotype and 464 mutations have been identified. Once the pea genome sequence data is available, mutant identification can be substantially extended to any genomic region as in several other crops. The commercial pea variety Cameor was used also to develop BAC library, an essential tool for positional cloning and also for pea genome sequencing. A second BAC library was developed from PI 269818 and could used to introgress genetic diversity into the cultivated germplasm pool (Coyne *et al.* 2007). Moreover, the complete pea chloroplast genome

sequence is also available and may be useful for evolutionary as well as transgenic applications (Magee *et al.* 2010; Bogdanova *et al.* 2015). In spite of great effort and progress in the molecular resources, the use of molecular genetics data in pea breeding has been limited mainly due to high genotype x environment interactions on the expression of important quantitative traits, necessity to test polymorphism of respective molecular markers in different genetic backgrounds, often large (5-10 cM on average) genetic distances between markers and the genes controlling respective traits, inprecise phenotypic description of targeted traits, resulting in wrong association, and small size of mapping populations (50-200 individuals) resulting in limited genetic resolution (**Smýkal *et al.* 2012**; **Warkentin *et al.* 2015**).

Recently, the progress made in gene-based marker density (Bohra *et al.* 2014, Tayeh *et al.* 2015a) resulting in development and use of GenoPea 13.2K SNP Array. These markers were then used to predict phenotypes: the date of flowering, the number of seeds per plant and thousand seed weight traits on the panel of 372 pea accessions (Burstin *et al.* 2015; Tayeh *et al.* 2015b). With expected pea genome in hands, the use of genomic and germplasm resources should be greatly facilitated.

Transcriptome analysis has been a key area of biological investigation for decades. The development of expressed sequence tags (ESTs) from pea has provided a source for mining novel simple sequence repeats (SSR) markers, valuable resources for gene discovery, expression analysis, and genome annotation (Kaur *et al.* 2012; Gong *et al.* 2010). Pea 6k oligo-array (Ps6kOLI1) developed from diverse sources of genomic sequence, especially seed EST libraries, have been performed for several transcriptome analyses. Seed development processes and specific genes involved in primary metabolism or hormone deficiency were investigated (Weigelt *et al.* 2008, 2009; Riebeseel *et al.* 2010; Radchuk *et al.* 2010). Hydrogen peroxide has been shown to accumulate during seed germination. The effect of treatment of mature pea seeds with hydrogen peroxide on several oxidative features and the expression of genes known to be activated by hydrogen peroxide were monitored as well as metabolites and function of antioxidant enzymes during maturations of pea seed (Matamaros *et al.* 2009; Barba-Espín *et al.* 2011). The development of transcription quantitative PCR methods facilitated transcript detection, increased the experimental throughput, and reduced the required quantity of input RNA. Important evaluation of candidate reference genes in pea varieties subjected to various abiotic and biotic stresses was undertaken, resulting in identification of tubulin-3 and TIF genes as the most stably expressed (Saha and Vandemark 2012). Transcriptome variations in reaction to abiotic and biotic stresses were also analyzed using several transcriptomic approaches. Microarray studies were used to obtain a global view of gene expression and provide information about the possible mechanisms and pathways involved in the resistance (Fondevilla *et al.* 2011). Similarly, chilling and acclimation mechanisms in freeze-tolerant pea line were compared with a sensitive line

on transcriptome gene profiles and were associated with morphological measurements and histological observations (Lacau-Danila *et al.* 2012). Further development in the microarray field led to other transcriptomic applications, such as detection of non-coding RNAs, single nucleotide polymorphisms (SNPs), and alternative splicing events. Currently leading DNA sequencing approaches to transcriptome analysis have been dominating over microarray-based methods. RNA sequencing (RNA-seq) based on next-generating technologies enable comprehensive survey of transcriptome even in species wihout available genome sequence (Wang *et al.* 2009) and allow method for both mapping and quantifying transcriptome. RNA-seg was used to study changes in gene expresion in legumes, including *Medicago truncatula* (Benedito *et al.* 2008), *Medicago sativa* (Zhang *et al.* 2015), soybean (Severin *et al.* 2010; Patil *et al.* 2015), faba bean (Kaur *et al.* 2012), *Lotus japonicus* (Verdier *et al.* 2013b) and chickpea (Pradhan *et al.* 2014). In pea, transcriptome analysis studies have generated libraries from flowers, leaves, cotyledons, epi- and hypocotyls and seedlings (Fransessen *et al.* 2011), from different reproductive tissues (flowers, immature and mature pods same as seeds) of four field pea cultivars (Kaur *et al.* 2012), further from eight various cultivars – spring sown, winter sown and fodder (Duarte *et al.* 2014), from different tissues (reproductive, subterranean and vegetative tissues) from two cultivars (Sudheesh *et al.* 2015), from various developmental stages of seeds and pods for grain and vegatable pea cultivar (Liu *et al.* 2015), young pea nodules (Zhukov *et al.* 2015). Seed coat transcriptome of various pea cultivars was analyzed in relation to proanthocyanidin pahway (Ferraro *et al.* 2014) and seed ageing (Chen *et al.* 2013). Moreover, there is pea RNA-seq gene atlas for 20 cDNA libraries including different developmental stages and nutritive conditions (Alves-Carvalho *et al.* 2015).

Chapter 2

Pea diversity

2.1 Analysis of *Pisum* genus diversity

Commenting on ***Smýkal P,*** *Hradilová H, Rathore A, Trněný O, Bariotakis M, Das RR, Hanáček H, Bhattacharyya D, Varshney R, Kilian A, Coyne CJ, Pirintsos S (2016) Characterizing of Pisum genus genetic diversity: past, present, and future patterns with inference about pea domestication. Molecular Ecology (submitted)*

2.1.1 Chloroplast and ITS diversity of wild *Pisum sp.* and cultivated pea landraces

The 458 samples representing wild types of the *Pisum sativum* subsp. *elatius, P. sativum* subsp. *humile* (215) complex, *P. abyssinicum* (20), *P. fulvum* (149) and cultivated landraces (76) were subjected to sequencing analysis of cpDNA (*trnS-G*) region in length of 855 bp. I adopted taxonomical classification of Abbo and Ladizinsky (2015). The samples were retrieved from major germplasm collections based on search for accessions with sufficiently reliable passport information regarding the specimen origin, including GPS data (**Smýkal *et al.* 2013**). These were tested for possible duplication (by passport data) and for misidentification (by cultivation and morphological assessment). Moreover, some important herbaria were inspected and leaf samples for DNA extraction were taken from 109 vouchers. The *trnS-G* analysis identified 7 haplotypes, which differed in 5 SNPs at positions of 132, 154, 164, 420 and 765 bp and one 6 bp indel in position 407-413 bp. This polymorphism defined 6 haplotypes in *P. s.* subsp. *elatius/humile*, *P. abyssinicum* and one in *P. fulvum* samples (**Figure 9**).

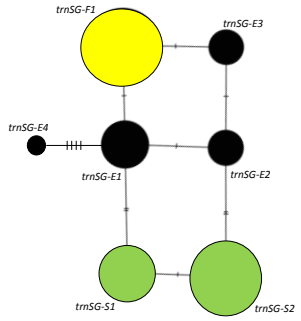


Figure 9

Haplotype trnSG network of 7 haplotypes differing in 5 SNPs and one 6 bp indel within the 855 bp region.

All analysed *P. abyssinicum* samples shared *trnSG-E1* haplotype with *P. s.* subsp. *elatius* (**Table 1**). Of 153 *P. fulvum* samples all but 6 had typical *P. fulvum trnSG-F1* haplotype. Of these six samples 5 had *trnSG-S2* (JI2510, JI2523, JI2539, WL2140, JI2539, JI1006, JI2521) haplotype of *P. s.* subsp. *humile* and IG112136 *trnSG-E3* haplotype of *P. s.* subsp. *elatius*. Wild *P. s.* subsp. *elatius/humile* samples had all together 6 *trnSG-E1-4, trnSG-S1-2* haplotypes. *TrnSG-E1* and *trnSG-E3* were separated by one SNP from *P. fulvum trnSG-F1* haplotype, *trnSG-E2* by further one SNP from *trnSG-E1*, while *trnSG-S1* and *trnSG-S2,* were by two SNPs from *trnSG-E1, E2* respectively (Fig. 1). *TrnSG-E4* was identical to *trnSG-E1* except of 6 bp (TACAAA) insertion. Thirty four samples had *trnSG-E1*, 29 *trnSG-E2,* 21 *trnSG-E3,* 5 *trnSG-E4,* 68 *trnSG-S1* and 50 *trnSG-S2*. Cultivated landraces were distributed among 4 haplotypes (4 *trnSG-E2*, 10 *trnSG-E3,* 2 *trnSG-S1* and 60 *trnSG-S2*). The Chinese origin group of cultivated pea represented by 15 samples was particularly diverse as they were distributed among the three *trnSG-E2, E3, S2* haplotypes.

	trnSG haplotypes							
	F1	E1	E2	E3	E4	S1	S2	Total
P. sativum subsp. elatius/humile		34	33	22	5	68	58	**220**
P. fulvum	143			1			5	**149**
P. abyssinicum		20						**20**
landraces			4	10		2	60	**76**

	ITS ribotypes																
	aby1	*aby2*	*f1*	*f2*	*f3*	*f4*	*ela1*	*ela2*	*ela3*	*ela4*	*ela5*	*ela6*	*ela7*	*ela8*	*ela9*	*ela10*	*ela11*
P. sativum subsp. elatius/humile	1	2					27	6	10	19	13	6	12	3	1	5	3
P. fulvum			73	4	23	49											
P. abyssinicum	8	7					2					1					
landraces	2						9			4	9	17			4	1	

	ela12	*ela13*	*ela14*	*ela15*	*ela16*	*ela17*	*ela18*	*ela19*	*ela20*	*ela21*	*ela22*	*ela23*	*ela24*	*unique*	*land1*	*land2*	Total
P. sativum subsp. elatius/humile	11	41	9	9	6	2	1	2	2	4	3	3	2	12			**215**
P. fulvum																	**149**
P. abyssinicum														2			**20**
landraces	2	5	2				1							4	9	4	**76**

Table 1

Summary of cpDNA haplotypes and ITS ribotypes detected in complete set of 458 samples.

The analysis of nuclear encoded ITS region resulted in 18 polymorphic sites within the region of 664 bp detecting all together 31 major and 19 unique ribotypes (represented by single sample). The alignment included 27 bp of 18S rDNA, 238 bp of ITS1, 164 bp of 5.8S rDNA, 213 bp of ITS2 and 22 bp of 26S rDNA, totalling 664 bp. Twenty *P. abyssinicum* samples had 2 (11 *its-aby1*, 9 *its-aby2*) major and 2 additional ribotypes represented by one sample (*its-ela1* for JI1974 and JI1457, *its-ela6* for JI2385). One *P. s.* subsp. *elatius/humile* (JI1090 from Mersina, Turkey) and two landraces (PI358608, JI1834 from Spain) had *its-aby1* haplotype, while two *P. s.* subsp. *elatius/humile* had *its-aby2*. 153 samples of *P. fulvum* had 73 *its-f1,* 4 *its-f2*, 23 *its-f3* and 49 *its-f4* being mutually separated by one mutation step, while from the closest *P. s.* subsp. *elatius/humile* samples (*its-ela11, its-ela15, its-ela20* haplotypes) by 11 mutation steps (**Figure 10**). Wild *P. s.* subsp. *elatius/humile* had 24 ribotypes represented by more than single sample (38

its-ela1, 6 *its-ela2,* 10 *its-ela3,* 23 *its-ela4,* 22 *its-ela5,* 24 *its-ela6,* 12 *its-ela7,* 3 *its-ela8,* 5 *its-ela9,* 6 *its-ela10,* 3 *its-ela11,* 13 *its-ela12,* 46 *its-ela13,* 11 *its-ela14,* 9 *its-ela15,* 6 *its-ela16,* 2 *its-ela17,* 2 *its-ela18,* 2 *its-ela19,* 2 *its-ela20,* 4 *its-ela21,* 3 *its-ela22,* 3 *its-ela23* and 2 *its-ela24*) and 12 unique single sample ribotypes (**Table 1**). Cultivated *P. sativum* shared 10 ribotypes (2 *its-aby1,* 9 *its-ela1,* 4 *its-ela4,* 9 *its-ela5,* 17 *its-ela6,* 4 *its-ela9,* 1 *its-ela10,* 2 *its-ela12,* 4 *its-ela13* and 2 *its-ela14*) with wild *P. sativum* subsp. *elatius/humile*, while two ribotypes were exclusive for landraces (9 *its-land1*, 4 *its-land2)* separated by 5 or 6 mutation steps from nearest wild *its-ela22* and *its-ela23* and 7 landraces had unique ribotype (**Table 1**). There were two major haplotype groups of wild *P. s.* subsp. *elatius/humile* 38 samples with *its-ela1* and 46 with *its-ela13*. As ancestral sequence for group of *P. s.* subsp. *elatius/humile* was identified of *its-ela1* ribotype.

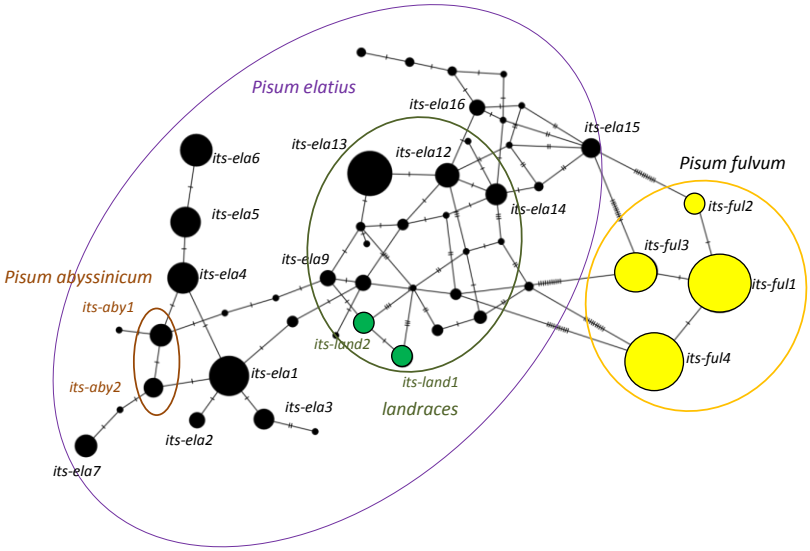


Figure 10

ITS ribotype network based on 664 bp ITS region with 18 polymorphic sites detecting 31 major and 19 unique ribotypes in 458 samples.

Bayesian analysis of ITS sequences of 458 samples (158 *P. fulvum*, 217 *P. s.* subsp. *elatius/humile*, 20 *P. abyssinicum* and 82 landraces) revealed structuring into K=6 groups. All *P. fulvum* samples were grouped together and distinguished from K=3 to 6, irrespective of 4 detected ribotypes. Wild *P. s.* subsp. *elatius/humile* samples were divided into two (at K=3) up to five (K=6) clusters (**Figure 11**). *P. abyssinicum* were incorporated into *P. s.* subsp. *elatius/humile* group at all K values. Several *P. abyssinicum* samples (PI358607, PI358610, PI358611, PI358613, PI358614, W808) at K=4 and K=6 showed admixture between two distinct *P. s.* subsp. *elatius/humile* clusters. Landraces were structured at two (K=3), three (K=4) and 4 clusters (K=5 and 6).

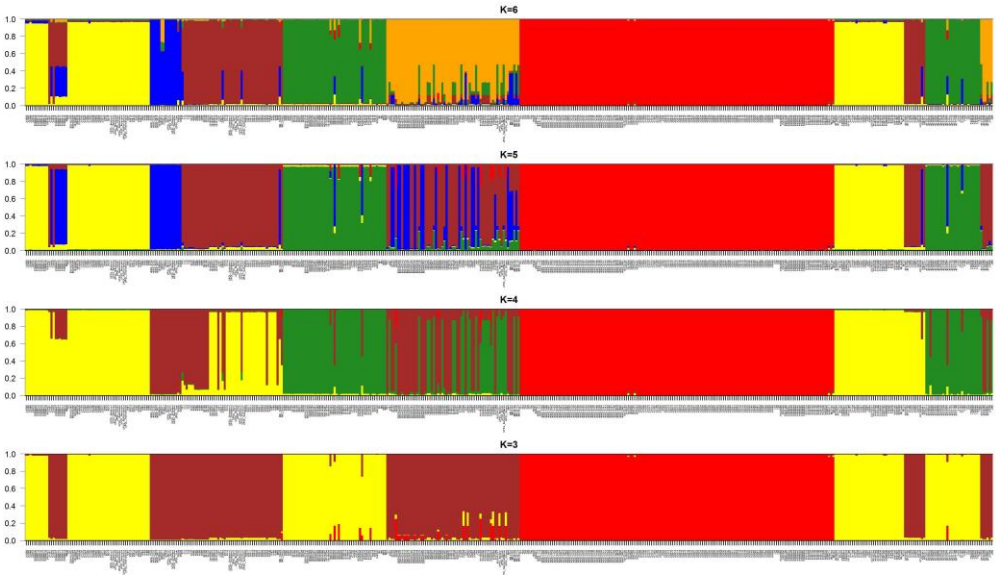


Figure 11

Bayesian analysis of ITS sequences of 458 samples (158 P. fulvum, 217 P. sativum subsp. elatius/humile, 20 P. abyssinicum and 82 landraces) at K= 4 to 6 groups based on 18 Single Nucleotide Polymorphism. Each genotype is represented by a vertical bar, which is partitioned into K-coloured components representing the ancestry fractions to given K value.

Taken together we have detected complex cpDNA and ITS haplotype network of wild pea samples, which is partly geographically structured with several cases of hybridization events.

2.1.2 Genome-wide analysis of *Pisum* diversity

The drawbacks of marker bias are largely overcome by next-generation sequencing (NGS) technologies allowing for an unbiased genome-wide view on thousands of fragments often representing genes as used in our DARTseq study. Isolated genomic DNA from 64 selected landraces (e.g. representing cultivated types), 52 wild *Pisum subsp. elatius/humile*, 20 *P. fulvum*, 10 *P. abyssinicum* and 4 *Vavilovia formosa* samples, was subjected to standardized DArTseq analysis at Diversity Arrays Technology Ltd. Canberra, Australia using proprietary methodology. DArTseq™ represents a combination of a DArT complexity reduction methods and next generation sequencing platforms (Kilian *et al.* 2012; Courtois *et al.* 2013; Cruz *et al.* 2013; Raman *et al.* 2014). Four methods of complexity reduction were tested in peas and the *PstI-MseI* method was selected. DNA samples were processed in digestion/ligation reactions principally as per Kilian *et al.* (2012). After PCR equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to c-Bot (Illumina) bridge PCR followed by single read sequencing on Illumina Hiseq2500. Sequences

generated from each lane were processed using proprietary DArT analytical pipelines. DArTseq analysis resulted in more than 35,500 sequenced fragments in SNP-based output per accession.

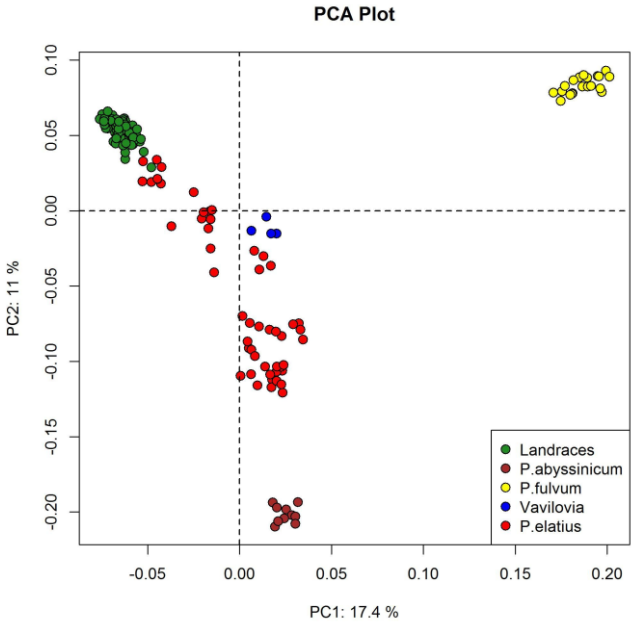


Figure 12

Principal Component Analysis (PCA) of pairwise individual genetic distances among wild P. sativum subsp. elatius/humile, P. fulvum, Vavilovia formosa and domesticated pea landraces based on DARTseq dataset.

The unweighted neighbour joining tree for 147 accessions (**Figure 13**) showed two major clusters. One cluster contained all 64 domesticated pea samples. The second cluster was structured into three sub-clusters, one with *P. fulvum* and adjacent *Vavilovia formosa* (4), a second positioned between *P. fulvum* (20) and a third the part of "*elatius/humile"* (51) containing all *P. abyssinicum* (10) samples. Moreover *P. abyssinicum* were genetically very homogenous, indicating a a restricted genetic basis and little differentiation in this group. At the base of *P. abyssinicum* there was one *P. sativum* subsp. *elatius* (IG52520) from Turkey. The second part of "*elatius/humile"*contained 14 samples of various geographical origins, including 4 samples from the Caucasus. There were 15 wild *P. s.* subsp. *elatius/humile* samples found in the domesticated genepool (E1-4 in **Figure 13**) either being incorporated within landraces or forming separate subcluster E4. Closer inspection of domesticated samples (landraces) either within or without the context of wild pea revealed further structuring, with a partial geographical pattern. The most diverse group was sub-cluster E1 with 34 samples of broad assembly of landraces, while sub-cluster E2 contained mainly so called "*transcaucasicum*" and "*asiaticum*" types from Afghanistan or Georgia. Sub-cluster E3 contained samples from Nepal, Afghanistan, Pakistan, China, Tibet as well as from southern Turkey (IG52518), Syria (IG52514) and Israel (JI1853).

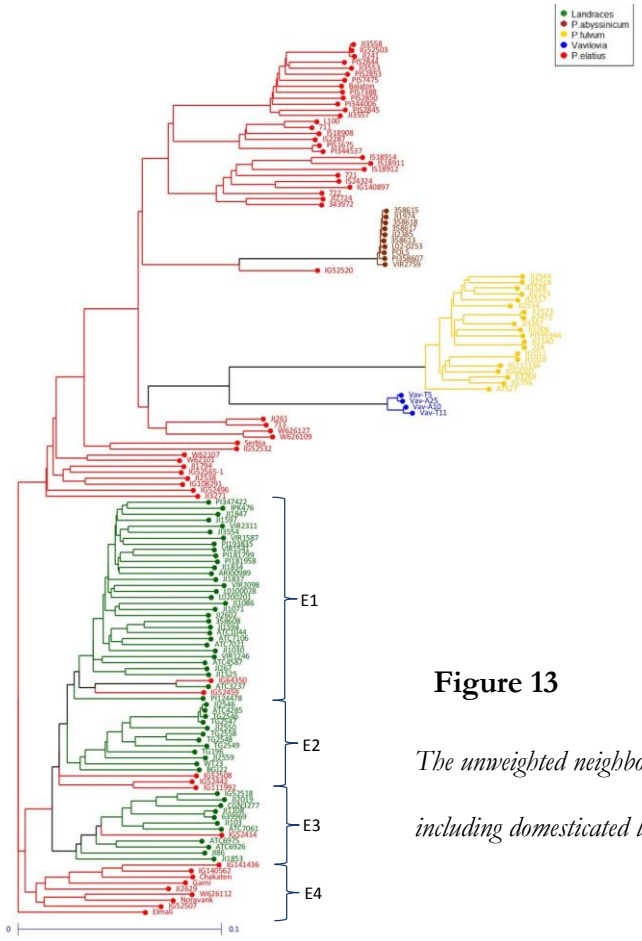


Figure 13

The unweighted neighbour joining tree for 147 accessions including domesticated landraces based on DARTseq data.

An unweighted neighbor joining tree of only wild samples (85) confirmed results of the whole set, but also clarified the subdivision of "*P. elatius/humile"* into two main clusters. One cluster (A) consisted of 29 samples, predominantly of European origin (13), Israel (10), one from Armenia and two from Turkey. Geographical assignment of 20 *P. fulvum* samples showed separation of northern (northern Syria and southeast Turkey) and southern (southern Israel and Jordan) samples (not shown) while there was no clear correspondence between ITS ribotypes and DARTseq based integer Neighbour-joining network. Next to *P. fulvum*, were the 4 analyzed *Vavilovia formosa* samples. Finally, at the base of the *P. fulvum* and *Vavilovia* branch, there were 4 "*elatius/humile"* accessions from Turkey, Georgia and Israel with *trnSG_E1, E2* haplotype and *its-ela3, its-ela10* ribotypes. The ten *P. abyssinicum* samples were genetically very close, nearly identical and formed a distinct group (between *P. fulvum* and *P. s.* subsp. *elatius/humile* with IG52520 from south-east Turkey, at the base. Calculation of molecular diversity within the five genepools (*P. s.* subsp. *elatius/humile, P. abyssinicum, P. fulvum, Vavilovia* and pea landraces) showed differences in genetic distances with the lowest value (0.0053) for *P. abyssinicum*, followed by *Vavilovia* (0.0134) and *P. fulvum* (0.0678). Cultivated pea (landraces) had slightly lower (0.1343) genetic distances than wild "*elatius/humile"* (0.1784) samples. Similarly Polymorphic Information Content (PIC) was lowest for *P. abyssinicum* (0.006) and *Vavilovia* (0.007), followed by *P. fulvum* (0.053), cultivated

landraces (0.112) and wild "*elatius/humile"* (0.147). Evaluation of population structure by STRUCTURE indicated the best probabilities for K=3, followed by K=6 and 7. At K=3, there were three clusters, one of *P. fulvum*, a second of wild *P. s.* subsp. *elatius/humile* containing all *P. abyssinicum* as well as *Vavilovia formosa* samples (not shown). The latter display about 0.3 proportion of *P. fulvum* alleles. The third cluster formed the landraces and eight wild samples. These were of various origins, which had 0.6-0.7 proportion of domesticated genepool alleles while the remaining portion was shared with the *P. s.* subsp. *elatius/humile* genepool. Similarly three "*asiaticum*" samples showed a portion of *P. fulvum, P. s.* subsp. *elatius/humile* and cultivated *P. sativum*. The lowest level of admixture was observed among the landraces at any K value. At K=4 (**Figure 14**), a newly separated cluster consisted of all *P. abyssinicum* samples (10). IG52520 from Turkey classified as wild "*elatius/humile"* had about 0.5 portion of *P. abyssinicum*, 0.4 of "*elatius/humile"* and 0.1 of *P. sativum* (landrace) alleles. Four *Vavilovia* samples at K=4 were assigned to *P. s.* subsp. *elatius/humile*, but showed also a proportion of *P. fulvum* (0.3) and *P. abyssinicum* (0.1) alleles. IG52532 (Turkey) and SRB (Serbia) samples displayed almost an equal proportion of wild and cultivated genepools. In the cluster of wild "*elatius/humile"*, there were samples with only portion of wild genepool (0.2 to 0.4). These samples originated from Armenia, Georgia, south-eastern Turkey, Israel, Jordan, Tunisia, Italy-Sardinia and from some "asiaticum" samples, all labelled as wild by passport data. Conversely, in cluster of landraces (**Figure 14**), there were 7 samples of wild "*elatius/humile"* by passport as well as morphological data. These were from south-eastern Turkey and Syria, Morocco (IG111992) and Algeria (IG64350). These showed only small proportion of wild pea alleles (up to 0.1 in STRUCTURE analysis) in all analysed K values. At K=5, the wild pea set was divided into three clusters. One cluster already distinguished at K=4 comprised Armenian, Georgian samples having a significant proportion of cultivated genepool alleles (0.6 to 0.8). One (red) of two newly identified clusters consisted of 19 mainly from Israel, Turkey and PIS1675, PI344537 from Italy. A second (blue) cluster included 15 samples of largely European origin but also from Turkey and Israel (JI241). Serbian (SRB) and Turkish (IG52532) samples consistently display an admixture with about 0.4-0.6 proportion of cultivated genepool at all analyzed K values. The PI358608, by passport data from Ethiopia, is in USDA-GRIN designated as *P. abyssinicum* but should be treated as a *P. sativum* landrace. Similarly, there are two misidentifications of *P. fulvum* samples: JI2538 and IG52496 which are, in fact, "*elatius/humile"* by phenotype and DARTseq data. At K=6, there were a few samples from Armenia which showed an additional cluster admixture at a very low proportion (0.1-0.15). At K=6 wild "*elatius/humile"* was divided into three clusters, as at K=5. One cluster corresponded to A, the second cluster to C of distance-based clusters (**Figure 13**). *Vavilovia* samples showed a 0.2 proportion of *P. fulvum* and 0.8 of "*elatius/humile"*alleles. Finally at K=7, seven identified clusters matched distance-based groups. One cluster contained all landraces, a second *P. fulvum*, a third *P.*

abyssinicum. Wild "*elatius/humile"* were separated into four clusters, one containing *Vavilovia* samples. A newly identified cluster (**Figure 14**, violet) contained samples from Armenia, Turkey, Israel and JI2724 from the Balearic Islands. These *P. elatius* clusters do not correspond to cpDNA haplotypes or ITS ribotypes as both "sativum" and "elatius" groups are equally present in all. Similarly there was no clear geographical partitioning, although two clusters are enriched with Israeli samples (7/10 and 6/9), while one identified already at K= 5 has samples predominantly of European (11/13) origin. The samples also contained Ben-Zeév and Zohary´s (1973) accessions such as 711, 712 (L100), 713 (*humile* types), 721, 722 (*elatius* types) which at this K value were grouped into two different clusters. Two *P. fulvum* by passport data (IG52496 and JI2538) samples were separated from rest of *P. fulvum* already at K=3 and display 0.4 to 0.5 proportion of wild and landraces genepools respectively. JI2538 has been identified to be misclassified and it is *P. s.* subsp. *elatius* instead. Domesticated landraces form a coherent cluster at all K (3 to 7) without any further structuring.

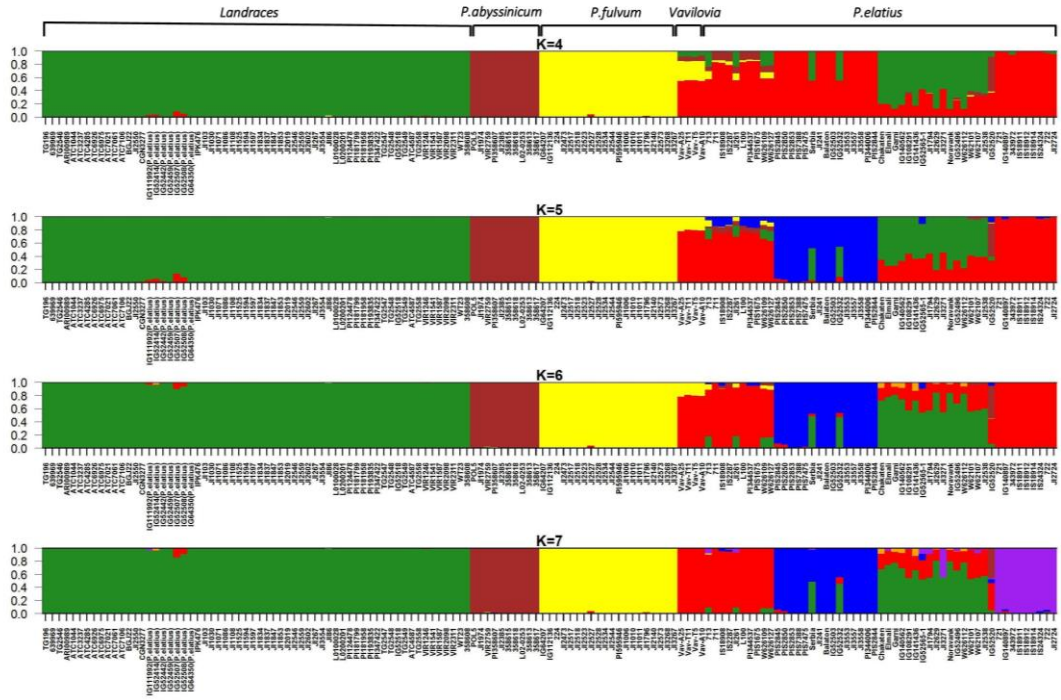


Figure 14

Estimated population structure by STRUCTURE software based on complete DARTseq data for 147 samples. K values 4 to 7 are shown and assignment to respective taxonomical groups is given. Each genotype is represented by a vertical bar, which is partitioned into K-coloured components representing the ancestry fractions to given K value.

When ITS ribotypes were assigned to DARTseq analysed 147 samples, landraces had only 3 ribotypes (10 *its-ela13*, 9 *its-ela18* and 12 *its-ela22*), *P. fulvum* samples contained 3 ribotypes (11 *its-f1*, 2 *its-*

f3 and 7 *its-f4*) while wild "*elatius/humile"* samples had 18 major and 5 unique ribotypes. *Vavilovia* had distinct both *trnSG* and *ITS* haplotypes. By *trnSG* haplotypes, landraces had predominantly *trnSG-S2* haplotype (46), one of each *trnSG-E1, E4* and *S1,* 4 of *trnSG-E2* and 5 of *trnSG-E3*. In wild "*elatius/humile"*samples, *trnSG-E2* (16), *trnSG-S1* (12) and *trnSG-S2* (10) were most abundant. Interestingly *V. formosa* was not separated by STRUCTURE analysis, but was included into group of "*elatius/humile"*. Distance based analysis of DARTseq data has positioned *V. formosa* in proximity of *P. fulvum*, detected of about 3,000 SNPs and showed very low heterozygosity, between 0.005 and 0.01. The STRUCTURE analysis showed clear separation of cultivated pea and wild *P. sativum* subsp. *elatius* and *P. fulvum*. *P. abyssinicum* were separated from *P. sativum* subsp. *elatius* at K=4.

2.1.3 The past, present, and future of genetic diversity patterns in the genus *Pisum*

Using the GPS data for 409 *P. s.* subsp. *elatius/humile* and 106 *P. fulvum* accessions, the potential niches of the species were modeled using Maxent version 3.3.3k (Philips *et al.* 2006) and this part of the work is currently under revision in submitted journal. The modeling was done by Prof. Stergios Pirintsos and his team at University of Crete, Heraclion, Greece. The environmental predictors that were used in the models were the 19 bioclimatic variables (Hijmans *et al.* 2005) extracted from www.worldclim.org. Each of these models was evaluated via a leave-one-out procedure, using the Area Under the receiver operating characteristic Curve (AUC) as a measure of performance. AUC values range from 0 to 1. Models with an AUC C 0.7 are considered acceptable, with an AUC C 0.8 are considered excellent and models with an AUC C 0.9 are considered outstanding (Hosmer and Lemeshow 2000). The potential niches of the species were then projected in past and future climatic conditions, in particular during the Last Glacial Maximum (about 22 thousand years ago) and during the year 2070 respectively, following in the latter case the Representative Concentration Pathway (RCP) 6.0 scenario using bioclimatic data created by the Global Climate Model CCSM (Community Climate System Model) 4.0. The three steps of niche analysis, namely niche modelling, niche similarity tests and niche diversity, were also carried out for the different genotypic groups of cpDNA analyses, as they resulted from the molecular analyses. According to model evaluations, modelling accuracy for the species *P. fulvum* and *P. s.* subsp. *elatius/humile* was excellent, with all AUC values being above 0.9. Predictions of the potential distribution of the three species, as they result from niche modelling, are displayed in **Figure 15**. These predictions are generally in accordance with the distribution of the occurrence points, with *P. fulvum* showing a much narrower potential distribution from the other two species.

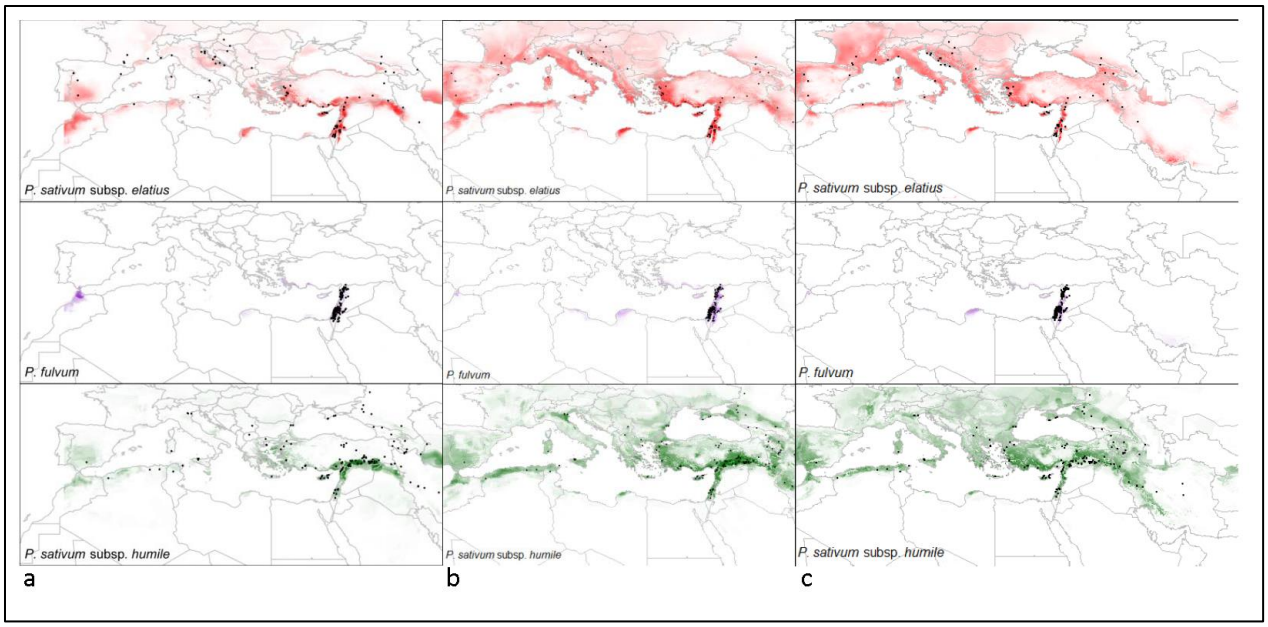


Figure 15

*Results of ecological niche models for the species P. fulvum and P. s. subsp. elatius/humile (**a.** LGM projection, **b.** current prediction, **c.** future projection). Lighter colours correspond to lower probabilities of occurrence, while more saturated colours correspond to higher probabilities of occurrence. Black dots represent the occurrence points that were used in the models.*

In the case of *P. fulvum*, BIO7 (Temperature Annual Range = Max Temperature of Warmest Month – Min Temperature of Coldest Month) had the highest contribution in the bioclimatic model (38%) followed by BIO19 (Precipitation of Coldest Quarter) (13.6%) and BIO12 (Annual Precipitation) (13.3%). For the *P. sativum* subsp. *elatius* BIO6 (Min Temperature of Coldest Month) had the highest contribution (41.4%), followed by BIO15 (Precipitation Seasonality) (17%) and BIO5 (Max Temperature of Warmest Month) (9.6%), while for the *P. sativum* subsp. *humile* BIO12 (Annual Precipitation) contributed the maximum with 28%, BIO4 (Temperature Seasonality) followed with 10.6% and BIO2 (Mean Diurnal Range = Mean of monthly (max temp - min temp) with 10.2% (**Figure 16**). Additionally, predictions reveal new areas of potential distribution for the three species. Most prevalent was the statistically significant niche similarity of *P. s.* subsp. *elatius* and subsp. *humile*, with three out of four relevant tests (both directions for metric D and one direction for metric I) showing statistically significant similarity, and the fourth yielding statistically non significant results.

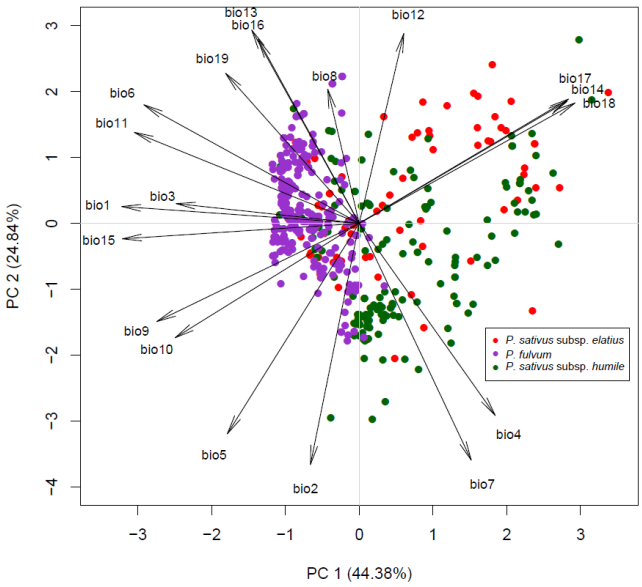


Figure 16

Ordination plane of PCA analysis. Dots and arrows represent samples and bioclimatic variables respectively. Arrows slightly elongated for visual optimisation.

There is a clear geographical pattern with high wild pea diversity at the Southern-Eastern part of the Mediterranean Basin (Northern Africa, Near East, Cyprus, Southern-Western Mediterranean coasts of Turkey and Southern Aegean islands). Out of the seven cpDNA haplotypes that were identified, six occurred in enough locations and could thus be modelled using Maxent: *trnSG*-E1, E2, E3, F, S1 and S2. Model evaluation again showed high predictive performance with AUC ranging from 0.897 ± 0.222 to 0.994 ± 0.005 (mean ± standard deviation). The spatial patterns of certain pairs, such as *trnSG*-E1 - E2 seem to be following similar patterns, while some haplotypes, such as F, seem to follow a more distinct pattern. As with the species, there was no definite case of divergence, while the pairs E1-E2 and E3-F were found statistically significantly similar for all four tests (two metrics and two directions). The pattern of Shannon's diversity index, showed discrepancy between this spatial pattern and the spatial diversity of the niche patterns for the three wild pea mainly concerns the high haplotype diversity which is predicted for the Southern part of the Western Mediterranean Basin and the Western and Balkan expansion of the Northern part of Eastern Mediterranean Basin. The discrepancy which concerns the high haplotype diversity at the two continents (Africa and Europe) across the strait of Gibraltar is also noteworthy. Raven and Polhill (1981) hypothesized that legumes differentiated some time before the end of Cretaceous in Africa, while recent phylogenetic results together with fossil evidence suggest that dispersal and vicariance, putatively linked to the Tethys seaway, is more likely to explain present legume distributions (Schrire *et al.* 2005; Sprent *et al.* 2013). The tribe Fabeae, following the classical scheme of Schaeffer *et al.* (2012), originated and evolved in the Eastern Mediterranean in the middle Miocene (23-16 Mya) and from the Mediterranean the tribe expanded its range into Central and Western Europe, to Asia and tropical Africa and to the New Word. Concerning the Atlantic crossing, the stepping-stone hypothesis has been proposed (Axelrod 1975), where shallowly submerged

seamounts would have been emergent during the extreme glacial sea-level minima, forming an array of stepping stones distributed between the present-day islands and the Iberian Peninsula and North Africa (see also Fernández-Palacios 2011). However, the biogeographical analysis of Fabeae of Schaeffer *et al.* (2012) revealed that the middle Atlantic islands did not serve as stepping-stones for lineages colonizing the New Word. On the other side the analysis revealed that long-distance dispersal events are relative common in Fabeae. This characterizes also the *Pisum* genus, and specifically the wild pea which spread from its center of origin, the Middle East, eastwards to the Caucasus, Iran and Afghanistan, and westwards to the Mediterranean (Smýkal *et al.* 2011). For the circum-Mediterranean area both the visualization of the predicted geographical pattern and the statistically significant niche similarity of *P. s.* subsp. *elatius* and subsp. *humile* support the here suggested re-classification of *Pisum*, which for wild pea identifies *Pisum elatius* in broader sense and *Pisum fulvum* as two distinct taxa, as well the hypothesis of ecological shift after the earlier major divergence of the genus (Zaytseva *et al.* 2015). Moreover, the spatial diversity of the niche patterns for the three reported wild taxa, as indicated by Shannon's Index reveal not only the species diversity center of the Near East, but also a predicted center of Northern Africa in the Eastern Mediterranean Basin and a predicted northern one incorporating the Southern-Western Mediterranean coasts of Turkey, Cyprus, and the Southern Aegean islands. These findings reveal the road of Northern Africa as an extra road for the westward expansion of the wild pea which has been rather neglected. This is supported by study of Zaytseva *et al.* (2015), which estimated divergence of *Pisum* genus between 1.7 to 1.3 MYA and subsequent radiation (Kosterin *et al.* 2010; Zaytseva *et al.* 2015).

2.1.4 Intra-population diversity of wild *Pisum sativum* subsp. *elatius* and its reproduction strategy

In contrast to domesticated legume crops (Smýkal *et al.* 2015), genetic pattern in natural populations was rarely studied. Published studies include endangered *Lathyrus pannonicus* (Schlee *et al.* 2011), *Vicia cracca* (Eliášová *et al.* 2014) and recently pea related glacial relict *Vavilovia formosa* (Smýkal *et al.* under review). In *Lathyrus pannonicus* (Schlee *et al.* 2011) ecological rather than geographical differentiation was found. So far the only published study using wild pea collected in nature is of Zaytseva *et al.* (2015) conducted on Histone H1 genes on one small population from Portugal. This study showed no variation in studied sequences among 19 individuals. However there is long standing question of genetic diversity and pollination system in natural conditions (Ellis 2011). Similarly, there are very limited studies performed on intra-population diversity of crop wild relatives (CRW) collected in nature, mostly done on wild cereals, such as exploring natural populations of *Hordeum spontaneum*

(Ozkan *et al.* 2005; Hubner *et al.* 2009, 2012, Jakob *et al.* 2014, Bedada *et al.* 2014), that is, using materials collected directly from the wild without long-term *ex situ* storage. Recently study of barley landraces (Poets *et al.* 2015) showed a broad contribution of wild progenitor populations to the landraces, indicating the existence of gene flow. Other studies used again germplasm preserved and *ex situ* multiplied samples, including wild pea (Jing *et al.* 2010, 2012, **Smýkal *et al.* 2012** and submitted). Such analysis is important both from botanical perspective to estimate intra-population diversity and gene-flow associated with open pollination as well as practical aspects related to germplasm conservation and potential use in breeding.

We have conducted so far unpublished study on material collected in natural conditions of south-eastern Turkey and eastern Europe to address the issue of intra-population diversity using both dominant AFLP and co-dominant microsatellite markers. We have studied all together 17 populations represented by 262 plants (Chaloupská 2015). The number of sampled plants varied according to population size which was estimated by habitat survey, accordingly we sampled about every 5-10th plants per side (**Table 2**).

Description of habitats and size of studied populations

Number of individuals per habitat varied largely (**Table 2**) from few solitary plants to several hundred plants over the area of 100 to $4x10^6$ sqm. In eastern Turkey typical habitat was ungrazzed or little grazed grassland among the shrubs (*Quercus sp., Pistacia terebinthus, P. lentiscus, Ceratonia siliqua, Pyrus elaeagnifolia*) and often close to some stony walls or rocky deposits. Accompanied by *Medicago sp., Lens orientalis, L. odemensis, Hordeum spontaneum*, *Aegilops sp.*

Population	Origin	Province	lat_dd	lon_dd	elevation	Population size estimates	Sample size	Population area estimation(m^2)	Genetic variation (AFLP)
A1	TUR	Antalya	36º 52'818''N	30º 24'376''E	990	20	12	1000	0.27
A2	TUR	Antalya	36º 53'472''N	30º 22'337''E	1180	150	15	200	
A3	TUR	Antalay	36º 53'140''N	30º 22'145''E	1278	50	15	400	
T1	TUR	Thermessos	36º 59'062''N	30º 27'791''E	986	100	25	$2x10^6$	0.23
KM	TUR	Kahraman Maras	37º 36'991''N	37º 04'780''E	1200	500	5	5000	n.t.
XS	TUR	XaniaSor	37º 34'372''N	39º 49'121''E	1430	13	13	$4x10^6$	0.25
SA	TUR	Savur	37º 32'091''N	40º 53'736''E	930	20	5	5000	n.t.
PINA	TUR	Pinadere	37º 28'796''N	40º 48'924''E	980	10	5	100	n.t.
HI	TUR	Hisar	37º 38'016''N	40º 53'347''E	730	200	40	$1x10^6$	0.20
BAG	TUR	Bagyaki	37º 31'586''N	40º 42'778''E	845	10	17	200	0.00
KEB	TUR	Kebapcik	37º 32'153''N	40º 31'718''E	900	100	20	$1x10^6$	0.03
SUL	TUR	Sultankoy	37º 26'070''N	40º 38'225''E	992	15	10	100	n.t.
DE	TUR	Derik	37°23'34.96"N	40°17'31.19"E	900	2	2		n.t.
YES	TUR	Yesilkoy	37°35'54.00"N	40°29'5.94"E	900	50	19	1000	0.16
GUR	TUR	Gurbuz	37º 38'440''N	41º 25'697''E	825	200	42	$2x10^6$	0.05
GAZ	TUR	Gaziantep	37° 1'48.30"N	37°13'48.78"E	720	10	5	100	n.t.

n.t. = not tested

Table 2: *List of eastern Turkey collection sites with indicated samples, population sizes and AFLP detected variation.*

	KM	XS	SA	PINA	HI	BAG	KEB	SUL	DE	YES	GUR	GAZ
KM	0.00											
XS	241.40	0.00										
SA	336.40	95.03	0.00									
PINA	329.70	88.50	9.34	0.00								
HI	335.50	94.54	11.00	18.28	0.00							
BAG	320.30	79.01	16.13	10.41	19.57	0.00						
KEB	304.10	62.72	32.36	26.05	33.57	16.29	0.00					
SUL	314.40	73.81	25.39	16.53	31.37	12.22	14.79	0.00				
DE	283.00	41.72	53.48	47.55	52.91	37.64	21.50	34.62	0.00			
YES	300.00	58.78	36.87	31.97	35.81	21.63	7.93	22.62	17.10	0.00		
GUR	382.90	142.00	48.39	56.90	47.48	64.30	80.12	73.43	100.30	83.21	0.00	
GAZ	66.55	236.80	329.10	321.20	330.40	313.00	297.20	304.90	277.70	294.70	377.30	0.00
												Km

Great-circle distance between two points based on a spherical earth. In fact earch is slightly ellipsoid.

Table 3

Pairwise geographical distances between 12 eastern Turkey populations, ranging from 10 to 383 km.

Locality Villany, southern Hungary

Village Villany is located 30 km south-east of Pécz, southern Hungary. Locality with wild pea distribution is on Szarsomlyo mountain (45°51´18.25´´/18°25´8.25´´, 442 m n. m.), declared since 1944 as natural reservation. Along with locality at Mt. St. Georgyu, Balaton,Hungary, this is the most northern distribution of *P. sativum* subsp. *elatius*. Xerothermic southern slopes are composed of shrubs (some species) *Inulo spiraeifoliae-Quercetum pubescentis* association and rocky slopes with *Sedo sopianae-Festucetum dalmaticae* (Erdős *et al.* 2012, 2013).

Locality Sv. Prohor, Pcinja valley, southern Serbia

The lowest, gorge-like, part of the Pčinja valley is characterized by the presence of a significant number of Mediterranean elements of flora and vegetation that comprise different plant communities (**Zlatkovič *et al.* 2010**). *P. sativum* subsp. *elatius* occures at foothill of northerly exposed slopes of Mount Kozjak. The slopes are covered by thermophilous submediterranean forests and scattered scrub vegetation consisted of pubescent oak (*Quercus pubescens*) and juniper (*Juniperus oxycedrus*) overgrowing siliceous rocky ground. Two small groups of the individuals were recorded in the edges of the forests, in open spaces between the trees and shrubs.

This is ongoing study, the very first of genetic analysis of nature collected wild pea samples at individual plant level. Since AFLP provides dominant markers, we have been genotyping this material also by co-dominant microsatellite markers. The dataset is not yet final however it shows already now that there is both intra-population diversity (from zero to 40%) as well as heterozygozity. The later is

likely underestimated, and varies from zero to 83%. AFLP analysis showed zero variability within population sampled at Bagyaki (0% polymorphism, 0.004 He), while highest was at Antalya (77% polymorphism, 0.268 He) **Figure 17**. This indicates the existence of significant open-pollination which was so far only speculated (Ellis 2011) and which is bellow 1% in case of cultivated pea crop in European conditions (Dostálová, Seidenglanz, Griga 2005; **Griga *et al.* 2008**). Interestingly, there is variation in *ELF3* gene, involved in photoperiodic regulation of flowering (Weller *et al.* 2012) and as well as nuclear-cytoplasmic incompatibility genes (Bogdanova *et al.* 2015). We speculate that variation of onset of flowering and possibly also incompatibility genes greatly contribute to the maitenance of intra-population diversity.

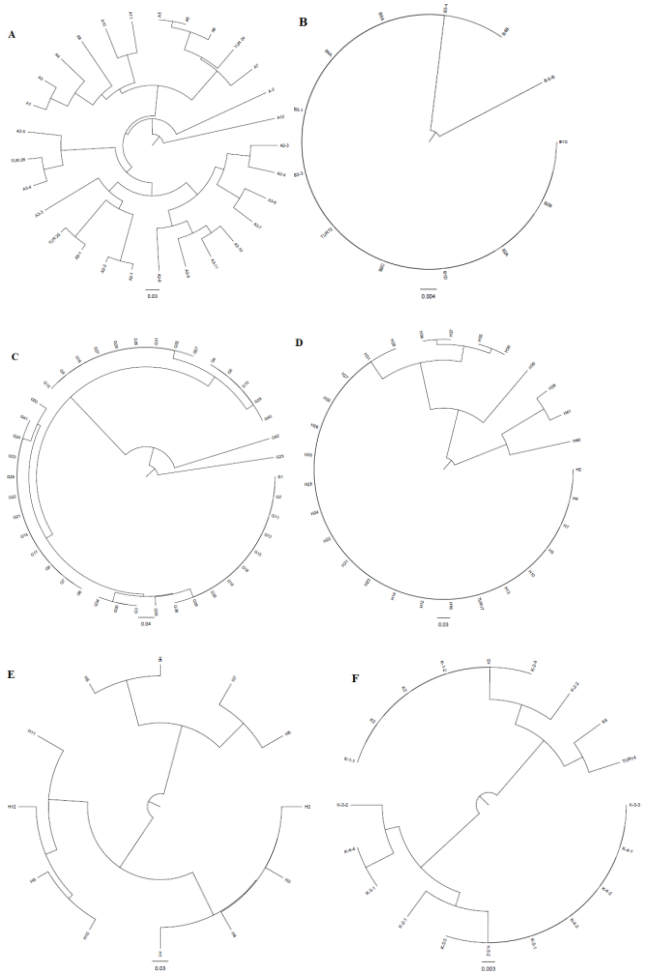


Figure 17

UPGMA representation of genetic diversity of studied Pisum sativum subsp. elatius populations studied by AFLP (A-K: Antalya, Bagyaki, Gurbuz, Hisar, Hungary-Balaton, Kebapčik, Hungary -Villany, Srbsko, Thermessos, Xaniasor, Yesilkoy). adapted from Chaloupská 2015

2.2 Assessment and conservation of pea diversity

Commenting on: Jing R, Ambrose MA, Knox MR, ***Smykal P*** *and 12 co-authors (2012) Genetic Diversity in European Pisum Germplasm Collections. Theoretical and Applied Genetics 125: 367-380.*

Cieslarová J, Hýbl M, Griga M, Fialová E, ***Smýkal P*** *(2012) Molecular Analysis of Temporal Genetic Structuring in Pea (Pisum sativum L.) Cultivars Bred in the Czech Republic and in Former Czechoslovakia Since the Mid-20th Century. Czech J. Genet. Plant Breed. 48: 61-73.*

De Beukelaer H, ***Smykal P,*** *Davenport G, Fack V (2012) Core Hunter II: fast core subset selection based on multiple genetic diversity measures using Mixed Replica search. BMC Bioinformatics 13: 312.*

Upadhyaya HD, Dwivedi SL, Ambrose M, Ellis N, Berger J, ***Smýkal P****, Debouck D, Duc G, Dumet D, Flavell A, Sharma SK, Mallikarjuna N, Gowda CL (2011) Legume genetic resources: management, diversity assessment, and utilization in crop improvement. Euphytica 180: 27-47.*

Cieslarová J, ***Smýkal P****, Dočkalová Z, Hanáček P, Procházka S, Hýbl M, Griga M (2011) Molecular evidence of genetic diversity changes in pea (Pisum sativum L.) germplasm after long-term maintenance. Genet Resour Crop Evol. 58: 439-451.*

Jing R, Vershinin A, Grzebyta J, Shaw P, ***Smýkal P****, Marshall D, Ambrose MJ, Ellis THN, Flavell AJ (2010) The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. BMC Evolutionary Biology 10: 44.*

Smýkal P*, Hýbl M, Corander J, Jarkovský J, Flavell A, Griga M (2008) Genetic diversity and population structure of pea (Pisum sativum L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. Theoretical and Applied Genetics 117: 413-424.*

2.2.1 Pea germplasm collections

The demand for high productivity and homogeneity in crops has resulted in a limited number of standard, high-yielding varieties, at the expense of heterogenous traditional local varieties (landraces), a process known as genetic erosion (Pistorius 1997). Traditional local varieties and landraces preserve much of the diversity within a species and comprise the genetic resources for breeding new crop varieties to cope with environmental and demographic changes. Although modern agriculture feeds more people on less land than ever before, it also results in high genetic uniformity by planting large areas of the same species with genetically similar cultivars, making entire crops highly vulnerable to pest

and diseases and for abiotic stresses. Thus uniform high-yielding cultivars are displacing traditional local cultivars, a process known as genetic erosion (Breese 1989). There are two approaches for conservation of plant genetic resources, namely in situ and ex situ. While in situ conservation involves the maintenance at natural habitats, ex situ involves conservation outside, like seed bank or field bank and botanical gardens. The danger of landrace diversity vanishing from cultivation was recognized very early upon scientific breeding (von Proskowetz 1890; Schindler 1890). To avoid such genotype extinction and enable long term ex situ conservation, the germplasm collection concept was proposed by Baur (1914) and made a reality by N.I. Vavilov in 1920-1940. Currently, about 7.4 M accessions of plant genetic resources are maintained globally, while 25-30% of total holdings are unique (2nd Report on the State of the World Plant Genetic Resources for Food and Agriculture, 2009, referred hereafter as SWPGRFA 2009). Legumes constitute the second largest group (~15% of all the accessions) after cereals. Collectively, ~1 M samples of grain legume genetic resources are preserved in ex-situ genebanks globally. Managing and utilizing such large diversity in germplasm collections are great challenges to germplasm curators and crop breeders (**Upadyaya *et al.* 2011**).

Landraces and wild populations are usually genetically heterogeneous and therefore have complex genetic structures, even when the degree of self-pollination is virtually complete. Furthermore, especially in case of wild species, features like seed dormancy, seed shattering, and high variability in flowering time and seed production play important role in the relative frequency of alleles as a result of changes in population genetics. We have investigated the changes in genetic integrity (**Table 4**) of pea collection in process of germplasm maintenance (**Smýkal *et al.* 2008c**; **Cieslarová *et al.* 2011**) and temporal diversity changes over the 70 years of pea breeding (**Cieslarová *et al.* 2012**). In both cases the shifts in genetic diversity was observed. This brings important implications for germplasm maitenance as well as breeding.

Sample/Accesion *(20 individuals/year)*	Number of alleles	Observed heterozygozity	Number of detected lines	Number of polymorphic loci	Shannon's diversity index	Nei´s gene diversity
Arvika						
1963-Mean	14	0	3	4	**0.15**	**0.093**
St.Dev.					0.21	0.13
2004-Mean	15	0	5	5	**0.23**	**0.16**
St.Dev.					0.29	0.20
Bohatýr						
1982-Mean	11	0	2	1	**0.15**	**0.094**
St.Dev.					0.21	0.11
2004-Mean	13	0.019	3	3	**0.17**	**0.11**
St.Dev.					0.28	0.12
Český_Banán						
1961-Mean	18	0	2	9	**0.53***	**0.37***
St.Dev.					0.26	0.18
2004-Mean	12	0	2	2	**0.04***	**0.017***
St.Dev.					0.08	0.038
Hanák						
1963-Mean	10	0	1	0	**0.00**	**0.00**
St.Dev.			1	0	0.00	0.00
2004-Mean	10	0	1	0	**0.00**	**0.00**
St.Dev.			1	0	0.00	0.00
Klatovský_zelený						
1963-Mean	15	0.015	8	5	**0.19**	**0.12**
St.Dev.					0.22	0.10
2004-Mean	15	0.010	8	5	**0.19**	**0.12**
St.Dev.					0.26	0.12
Moravský_Hrotovický						
1965-Mean	20	0	9	9	**0.53***	**0.37***
St.Dev.					0.29	0.19
2004-Mean	13	0	2	3	**0.13***	**0.090***

Table 4

Genetic erosion during the 40 years of germplasm regeneration process studied on 6 pea accessions by 10 microsatellite loci. Number of alleles, observed heterozygosity, number of detected lines per sample, number of polymorphic loci and Shannon's and Nei's indexes of gene diversity. Adapted from Cieslarová et al. 2010

To prevent the extinction of such genotypes, *ex situ* conservation of germplasm resources was pioneered by Vavilov (1926) and nowadays, germplasm collections hold over 7 million crop plant accessions world-wide (**Upadhyaya *et al.* 2011**). World-wide there are approximately 98,000 pea accessions stored in various genebanks, of which around 58,000 might be unique (**Smýkal *et al.* 2013**). Currently, no international center conducts pea breeding and genetic conservation and no single collection predominates in size and diversity. Important genetic diversity collections of Pisum with over 2000 accessions are found in national genebanks in at least 15 countries, with many other smaller collections worldwide (**Smýkal *et al.* 2009, 2012, 2013**). There are 25 larger collections preserving pea diversity, holding together around 72,000 accessions. These include the John Innes Centre (JIC), UK (3,557 accessions); the Nordic Gene Bank (NGB), Sweden (2,724); the United States Department of Agriculture (USDA), USA (3,710); the International Center for Agricultural Research in the Dry Areas (ICARDA), Syria (6,105); Instituto del Germoplasma, Bari, Italy (4,297); Leibnitz Institute of Plant Genetics and Crop Plant Research (IPK), Germany (5,336); the Australian Temperate Field Crops Collection (ATFC), Australia (6,567); the Vavilov Institute of Plant Breeding (VIR), Russia (6,790); and the National Genebank of China (NGC), China (3,837). Also Czech National Pea Collection holds currently 2200 accesions, largely of European origin. Although this is a large number of accessions, it should be noted that the bulk of these are of cultivated peas, wild forms are sampled poorly and there is also significant duplication of collections between institutes. The remaining 27,000 accessions are distributed over 146 collections worldwide (**Smýkal *et al.* 2013**). Only 1876 (2%) of these are wild pea relatives, approximately one-quarter (24,000) each are commercial varieties, 8500 landraces, while 600 and 6000 represent breeding and recombinant inbred lines or mutant stocks, respectively. In the case of true wild Pisum species, there are only 706 *P. fulvum*, 624 *P. subsp. elatius*, 1562 *P. sativum subsp. sativum*

and 540 *P. abyssinicum* accessions (**Figure 18**) preserved *ex situ* in collections. Moreover, when passport data on geographical origin are summarized, there is a large bias (17%) towards Western and Central European accessions, as these regions represent modern pea breeding activities. Substantially less well represented are Mediterranean (2.5%), Balkan (2%) regions, Caucasus (0.8%) and Central Asia (2%) centres of pea crop domestication and diversity (**Smýkal *et al.* 2013**).

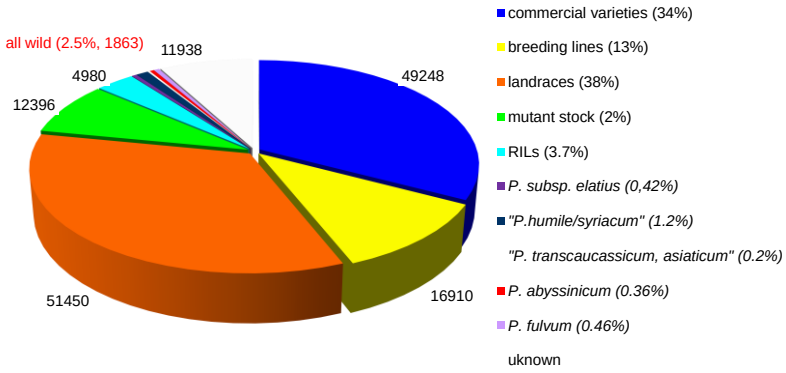


Figure 18

Stratification of pea germplasm collections based on the Smýkal et al. 2013 inventory by species, subspecies and breeding status, with inicated numbers and percentage of total.

There are gaps in the collections, particularly of wild and locally adapted materials (landraces), that need to be collected before these genetic resources are lost forever (Maxted *et al.* 2008). With the anticipated climate change affecting agricultural production, collecting pre-adapted germplasm as well as pest and disease resistance genotypes is a priority (Etterson and Shaw 2001). Climate change can be expected to exacerbate climate unpredictability and to result in unprecedented levels of heat and drought stress during the reproductive phase in agricultural areas of the temperate, sub-tropical zones worldwide, especially in the sub-Sahara and northcentral India (Coyne *et al.* 2011). Targeted utilization of selected landraces and wild relatives for adaptation to climate change will almost certainly be an urgent priority during this century. Pea as a major food legume has the capacity for enhanced nitrogen fixation and CO_2 capture, which may partially offset growth reduction associated with higher temperature, shorter growing season, and periods of drought (Craufurd and Wheeler 2009). There is an urgent need to systematically sample the genetic diversity in wild relatives that was only partially captured in the domestication of pea (Ellis 2011; **Smykal *et al.* 2011, 2013**), since natural habitats are being lost due to increased human population, increased grazing pressure, conversion of marginal areas to agriculture and ecological threats due to future climate change (Keiša, Maxted & Ford-Lloyd 2007; **Smýkal *et al.* 2013**). The target areas for comprehensive collection of wild relatives of peas include the habitat from the Mediterranean through the Middle East and Central Asia, as these are likely to contain genetic diversity for abiotic stress tolerances (Coyne *et al.* 2011, Smýkal *et al. submitted*). Pea has also a large number of mutant lines, either spontaneous or induced. It has been used as a model plant species for experimental morphology and physiology in mutagenic studies. Numerous morphologically well-

described mutants exist, many of them being used in genetic mapping. The earliest collection lists 21 pairs of cultivated pea lines for contrasting characters covering plant form, foliage, flowers, pods and seeds, which were the subject of genetic investigation, held within a collection of 550 cultivars (Vilmorin 1913). Later, Blixt (1972) made a list and linkage group positions of 169 genes (loci) which occurred spontaneously or were induced. Induced mutagenesis has become widespread for the creation of novel genetic variation for selection and genetic studies (Blixt, 1972; Lamm, 1951; Lamprecht, 1964) with mutants in traits for physiology, chlorophyll, seed, root, shoot, foliage, inflorescence, flowers and pods. These genetic analyses contributed to Pisum genus classification. The mutant collections have been largely preserved in John Innes Centre (JIC) (585 accessions) and Nord Genebank (**Smýkal *et al.* 2013**). In addition Murfet and Reid (1993) have developed and maintain developmental mutants in Tasmania. There is a pea population of 4817 lines newly established by the technique of targeting induced local lesions in genomes (TILLING) at Institut National de la Recherche Agronomique (INRA). In addition, fast neutron-generated deletion mutant resources (around 3000 lines) are available for pea, which have been useful in identifying several developmental genes (Hellens *et al.* 2010; Hofer *et al.* 2009; Wang *et al.* 2008).

2.2.2 Assessment of pea germplasm diversity

Molecular analysis of pea diversity preserved in germplasm collections, was carried out using various methods, including Amplified Fragment Length Polymorphism (AFLP) (Ellis *et al.* 1998; Tar'an *et al.* 2005), Retrotransposon-based Insertion Polymorphism (RBIP) (**Jing *et al.* 2010, 2012, Smýkal *et al.* 2006, 2008, 2010, 2011**), Sequence-Specific Amplification Polymorphisms (SSAP) (Pearce *et al.* 2000, Majeed *et al.* 2012; Vershinin *et al.* 2003), microsatellites (Ford *et al.* 2002; Baranger *et al.* 2004; **Smýkal *et al.* 2008c**; Zong *et al.* 2009a,b; Nasiri *et al.* 2010; Kwon *et al.* 2012), and gene sequences (Jing *et al.* 2007; Zaytseva *et al.* 2015; Sindhu *et al.* 2014; Burstin *et al.* 2015; Tayeh *et al.* 2015a,b). For the analysis of pea diversity, Simple Sequence Repeats (SSRs or microsatellites) have become popular because of their high polymorphism and information content, co-dominance and reproducibility (Baranger *et al.* 2004, Loridon *et al.* 2005, **Smýkal *et al.* 2008c**). A potentially existing but largely neglected problem in using SSRs for characterizing highly diverse germplasm, is homoplasy (**Cieslarová *et al.* 2010)** and the possibility of back-mutations exhibited by this marker type. Alternately, marker systems based on retrotransposon insertion polymorphism (RBIP) have been extensively used for phylogeny and genetic relationship studies in pea, providing a highly specific, reproducible and easily scorable method (Jing *et al.* 2007, **Smýkal *et al.* 2008**). Using these markers, several major world pea germplasm collections have been analyzed and core collections formed (Jing *et al.* 2010, 2012; Kwon *et al.* 2012; **Smýkal *et al.***

2008a,c, 2009, 2011). Smýkal *et al.* (2008) study has shown, that both SSRs and RBIPs have similarly high information content. This was an important finding, as SSRs in spite of multiple alleles detection, are more difficult to transfer between labs, while essentially binary RBIPs are simpler. Although SSR and RBIP marker types are widespread now, their potential is at its limits. With advances in model legume sequencing and genomic knowledge, there is a switch to gene-based markers in pea (Jing *et al.* 2007). However even the largest analysed sample sets of Jing *et al.* (2010) and Smýkal *et al.* (2011) were dominated by cultivated types (3,020 and 4,429 respectively) with only 140 wild types (**Figure 19, 20**).

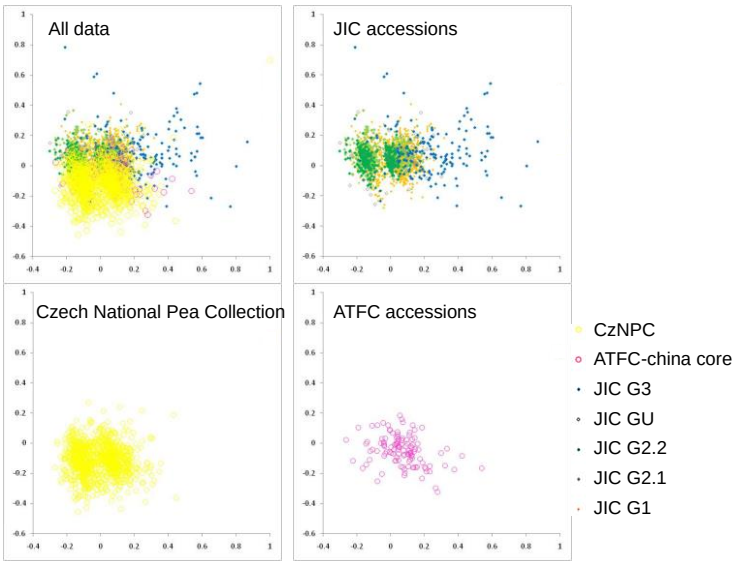


Figure 19

Multivariate analysis of composed dataset for the entire dataset the fraction of shared alleles for all pair-wise combinations of samples was analysed by multidimensional scaling. The output for the first two dimensions are shown. All points are plotted with each sample is colour-coded according to germplasm assignment: (A) composed dataset ,(B) JIC-TEGERM dataset (as in Jing et al. 2010), (C) Czech National Pea collection, L accessions, (D) chinese origin ATFC core collection

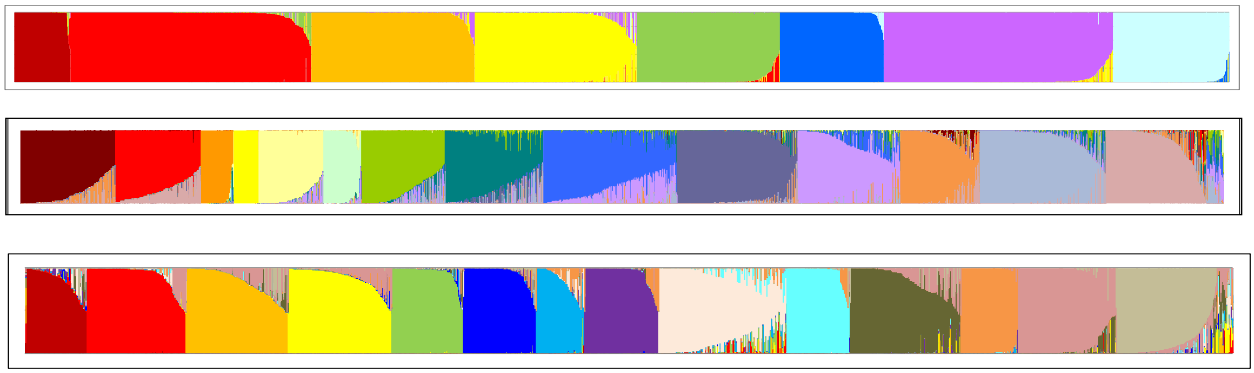


Figure 20

Bayesian Analysis of Population structure (BAPS) analysis partitioning. A) BAPS at K=8 of 2,120 accesions of ATFC collection (Zong et al. 2009a) genotyped by 21 SSR loci. ***B****) BAPS at K=14 of 3,029 accesions of JIC collection (TEGERM dataset, Jing et al. 2010) genotyped by 45 RBIP loci.* ***C****) Of 4,429 accessions of the combined set analyzed by 17 selected RBIP loci. adapted from Smýkal et al. 2011*

These studies showed that although *Pisum* is comparably small genus with two or three species, it is very diverse and diversity is structured, showing a range of degrees of relatedness that partially reflect

taxonomic identifiers, eco-geography and breeding gene pools (Ellis 2011; **Smýkal *et al.* 2011**; **Jing *et al.* 2012**).

2.2.3 Establishment of pea core collections

As the result of germplasm characterization over the past decades, there was attempts to make representative core collections to capture sufficient diversity into managable number of accessions. An important reason for the underutilization of germplasm in crop improvement programs is the lack of information on the performance of large number of accessions, particularly for traits of economic importance which display a great deal of genotype x environment interaction and require multilocation evaluation. The study of genetic diversity for both germplasm management and breeding has received much attention, especially following the introduction of the **core collection concept** by Frankel and Brown (1984). Also for legumes, core collections have been defined using various strategies (Erskine and Muehlbauer 1991; Van Hintum 1999) from passport to the use of evolutionary, agroecological and molecular data (Tohme *et al.* 1995; Baranger *et al.* 2004; **Smýkal *et al.* 2008a,c, 2009**; **Upadhyaya *et al.* 2011**; **Jing *et al.* 2012**). Landraces and older crop varieties preserved in germplasm collections still contain this unused diversity and comprise the genetic resources for breeding new crop varieties to cope with environmental and demographic changes (Esquinas-Alcazar 2005). Upon diversity analysis several core collections were formed (**Smýkal *et al.* 2008**; **Jing *et al.* 2012**; Zong *et al.* 2009a,b; Kwon *et al.* 2012), as well as trait focussed cores (**Upadhyaya *et al.* 2011**) to represent genus diversity (**Figure 21, 22**). Improvements in marker methods have been accompanied by refinements in computational methods to convert original raw data into useful representation of diversity and genetic structure. Distance-based methods (Reif *et al.* 2005) have been challenged by model-based Bayesian approaches. The incorporation of probability, measures of support, ability to accommodate complex model and various data types (Beauomont and Rannala 2004; Corander and Tang 2007) make them more attractive and powerful. Core subset selection is the problem of sampling such core collections which retain as much of the diversity of the original collection as possible, according to some diversity measure. These measures can be based on a variety of criteria including phenotypic traits or genetic marker data, or a combination of both. Over the years, many algorithms for core set selection have been proposed, including stratified sampling techniques. Other non stratied methods have also been developed, such as genetic distance sampling (Jansen, van Hintum 2007) which constructs core sets where no two accessions are closer to each other than a given threshold, according to some genetic distance measure. This avoids the need to specify a desired core size, but introduces the threshold as a new input parameter. All of the previously mentioned methods assume that the desired core size is

determined in advance and given as input to the sampling strategy, which then tries to create a good core set of the desired size according to the specific objective used. However, a related problem is that of finding the smallest possible core set which retains all unique alleles from the original collection (**De Beukelaer *et al.* 2012)**.

Recently the CoreHunter was developed as a new, very exible framework for selecting core collectionswere developed and applied on germplasm collections including pea (Thachuk *et al.* 2010; **Jing *et al.* 2012**). We have further improved this method by switching to a new advanced search technique. We have introduced a new advanced search algorithm - Mixed Replica Search (MixRep) which uses heterogeneous replicas, an approach based on results of a comparison of several algorithms' performance and showed that this new method improves on the results of the original REMC algorithm (**De Beukelaer *et al.* 2012**). It is clear that cores based largely on anonymous DNA markers can not properly reflect specific diversity, such as resistances to biotic and abiotic stresses. The potential improvement in screening efficiency offered by the core collection concept is applicable to modern allele mining efforts (Reeves *et al.* 2012) to recover useful adaptations from gene banks. However, exploring rare novel alleles in a large germplasm collection is still challenging task, inspite of some high-throughput method of identifying novel alleles from a large collection such as high resolution melting PCR approach (Hofinger *et al.* 2009), Eco-TILLING or genotyping by sequencing.

Figure 21

STRUCTURE assignments of selected subsets of accessions of Jing et al. (2010) is shown in 'd', using their colour codes and accession order. Panel 'e' shows the assignment of accessions to three STRUCTURE Groups identified here (Blue QB, Red QR and Green QG). Panels 'c' and 'g' show the samples selected by Core Hunter at 5% representation (black) with the seven accessions also in the minimum core highlighted in red. The 10%, 20% and 30% Core Hunter selections are in panels 'l' , 'm' and 'n' (MFA) and h, i, and j (STRUCTURE) respectively. Panels 'o' and 'k' show the 10% representation reselected from the 30% selection. In red circle are indicated wild pea (P. fulvum, P. sativum subsp. elatius) samples, representing the largest diversity. adapted from Jing et al. (2012)

A recent gap analysis study of *Cicer*, *Lathyrus*, *Lens*, *Medicago*, *Pisum* and *Vicia* species (Maxted *et al.* 2010; Ramirez-Villegas *et al.* 2010) indicated that the diversity held *ex situ* does not fully reflect natural *in situ* diversity and that there is a need for further collection. Many studies have been conducted on *Pisum* germplasm collections to investigate genetic and trait diversity (see **Smýkal *et al.* 2015** for review). The information on genetic diversity is important also for breeders to know the genetic basis of cultivars, especially whether they have become so narrow in diversity as to render the crop vulnerable to disease or pests. In other words accessions most genetically distinct from others are likely to contain the greatest number of novel alleles.

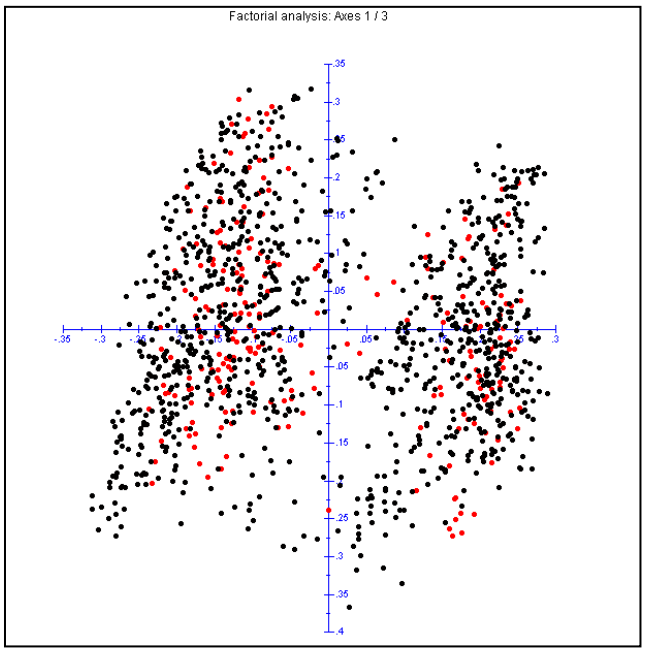


Figure 22

CoreHunter II selection of representative pea core set.

PCA plots of cores sampled from fingerprinting of 1283 bulks of 10 pea individuals each using 19 RBIP markers, with 4 diferent possible states for each bulk at each maker locus:(i) presence of allele in each individual,(ii) absence in each individual, (iii) mixed state having both individuals with presence and absence in the same bulk, and finally (iv) the zero state which means no data is available. The sampling intensity was set to 0.2. Core samples are in Red. (adapted from De Beukelaer et al. 2012)

Chapter 3

Pea crop origin

3.1 Pea domestication

*Commenting on **Smýkal P** and 8 co-authors (2014) A comparative study of ancient DNA isolated from charred pea (Pisum sativum L.) seeds from an Early Iron Age settlement in southeast Serbia: inference for pea domestication. Genet Resour Crop Evol. 61: 1533–1544.*

***Smýkal P,** Vernoud V, Blair MW, Soukup A and Thompson RD (2014) The role of the testa during development and in establishment of dormancy of the legume seed. Frontiers in Plant Sciences 5: 351.*

About ten thousand years ago, during crop and animal domestication, man had started the process resulting in the flora and fauna we see today. Without much of exaggeration we might consider domestication as one of the largest changes made in human history. The change from food gathering toward crop cultivation and animal husbandry can be definitely viewed as one of the largest milestones leading to the origin of modern agriculture and likely allowing the establishment of civilizations. Thus, thousands of years prior to great discoveries made by Charles Darwin or J.G. Mendel, successful gene combinations were chosen, many of which remain integral to crop production today. The domestication process affected not only plants, but also humankind. However, there were also penalties, as permanent settlement and population density led often to epidemics of diseases both on human as well as crop plants. Not to forget negative impact on environment and nature diversity. The process of plant domestication had consequences on plant genetic makeup. More than 250,000 higher plant species are described at the moment, but only about 500 have been cultivated at one time or another (Hammer 2003). Moreover, the current world population of 6 billion people relies on approximately 30 species, of which 4 grass species (rice, maize, wheat and sorhum) comprise over 50% of our food supply! Such reduction in biological diversity can be seen not only at the level of species, but also within single species. Crop genetic diversity, being crucial for feeding humanity as well as for the environment, continues to be reduced (Hammer 2003).

There is a debate on the origin of plant cultivation and domestication especially in the Near East. One party suggests that their origin was singular, rather fast (hundreds of years), and took place in the so-called "core area" about 250 × 150 km situated at the sources of the Euphrates and Tigris in Turkey (Dijarbakyr and Mardin regions); this was followed by further evolution of domesticated crops improving their quality (Abbo *et al.* 2010a, 2011a, 2012, 2013). Second group supports the so-called protracted model of plant domestication and argue that both cultivation and domestication had multiple origins over the Fertile Crescent, went on slowly and in parallel, with pre-domestication cultivation for 1–1.5 thousand years preceded domestication (as a genetical phenomenon) which was gradual, lasted for not less than 3,000 years and was crowned by fully domesticated crops (Willcox 2005; Tanno, Wilcox 2006b; Weiss *et al.* 2006; Fuller 2007; Allaby *et al.* 2008; Fuller *et al.* 2010, 2011, 2012a,b).

The European/West Asian civilisation is based on the onset of plant cultivation in the Near East centre which is characterised by a number of founder crops: eight "traditional" ones from three families: einkorn wheat (*Triticum monococcum* L.), emmer wheat (*T. dicoccum* (Schrank) Schuebl), barley (*Hordeum vulgare* L.), lentil (*Lens culinaris* Medic), pea (*Pisum sativum* L.), chickpea (*Cicer arietinum* L.), bitter vetch (*Vicia ervilia* (L.) Willd.) and flax (*Linum usitatissimum* L.) (Lev-Yadun *et al.* 2000; Zohary, Hopf, 2000; Abbo *et al.* 2010a, b; Zohary *et al.* 2011), with addition of broad bean (*V. faba*) (Abbo *et al.* 2013), plus possibly some lost crops (Melamed *et al.* 2008; Fuller *et al.* 2011; Abbo *et al.* 2013).

Members of the Fabaceae were domesticated as grain legumes in parallel with cereal domestications (De Candolle 1884; Vavilov 1951; Hopf 1986; Smartt 1990; Zohary and Hopf 2000; Abbo *et al.* 2012). The Legume family is arguably one of the most abundantly domesticated crop plant families. There are 13 genera (in six legume tribes) that constitute major legume crops (Lewis *et al.* 2005; **Smýkal *et al.* 2015**). Among the first legumes to be domesticated were members of the galegoid tribe such as pea, faba bean, lentil, grass pea and chickpea which arose in the Fertile Crescent of Mesopotamian agriculture. These grain legumes (pulse legumes) accompanied cereal production and formed important dietary components of early civilizations in the Near East and the Mediterranean regions. Archaeological evidence dates for pea is from 10,000 BP in the Near East (Baldev 1988; Zohary and Hopf 2000; Helbaek 1959, 1970; Fairbairn *et al.* 2005) and Central Asia (Riehl *et al.* 2013). In Europe, pea has been cultivated since the Stone and Bronze Ages (Zohary and Hopf 2000). Cultivation of pea spread from the Fertile Crescent to today's Russia, and westwards through Balkan (Kroll 1991; Borojević 2006) and along the Danube valley into Europe and/or to ancient Greece and Rome, which further facilitated its spread to northern and western Europe. In parallel, pea cultivation moved southward to Egypt and todays Sudan (7000 BP) eastward to Persia, India (4000 BP) and China

(Makesheva 1979; Chimwamurombe and Khulbe 2011; Zohary and Hopf 2000; Hancock 2012). The remains of pea and other pulses occur at high frequencies from sites dated from the 10th and 9th millennia BC (Willcox *et al.* 2008) and suggest that the domestication of grain legumes accompanied or could even predate that of cereals (Baldev 1988; Kislev and Bar-Yosef 1988; Weiss *et al.* 2006). Rich etymological evidence also supports the status of pea as one of the most ancient Eurasian crops (Mikić 2012). Domestication is often described as a quality of plants in which morphological (and genetic) changes are found amongst cultivated in comparison to wild populations. These domestication triggered changes represent adaptations to cultivation and human harvesting, accompanied by genetic changes (Lenser and Theissen 2013). Common set of traits have been recorded for unrelated crops, named domestication syndrome (Hammer 1984; Zohary and Hopf 2000). These include loss of germination inhibition and increase of seed size, linked to successful early growth of planted seeds. An increase in seed size of domesticates compared to their wild relatives is seen for almost all of the grain legumes and even for forage legumes. One theory suggests that greater planting depth in agricultural systems led to selection of larger legume seeds which produced more vigorous seedlings, but it seems likely that early farmers already selected for a higher proportion of starch, oil and protein as well. As greater seed size was selected and seeds were stored from one season to another, the potential for absorption of water and germination during storage made it necessary to select for seed dormancy. Moreover seed imbibition has crucial role in cooking ability of most grain legumes (**Smýkal *et al.* 2014**). Hence, reducing seed coat thickness led to a concurrent reduction of seed coat impermeability during the domestication. Seed shattering was selected against to avoid the natural explosive seed pod opening mechanism of wild legumes (**Figure 23, 25**). Experimental growing of wild peas and lentil, have demonstrated that both seed dormancy and pod dehiscence cause poor crop establishment via reduced germination as well as dramatic yield losses via seed shattering (Abbo *et al.* 2011). The results were inconsistent with models suggesting protracted domestication of Near Eastern grain legumes (Abbo *et al.* 2013).

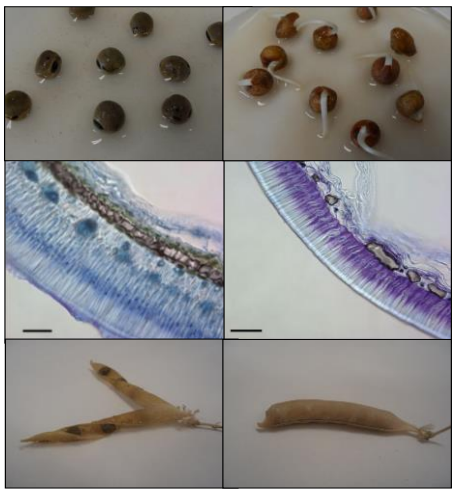


Figure 23

Comparison of wild (JI64) and cultivated pea (JI92).

Seed imbibition and germination (A, B)

seed coat thickness (C, D)

and pod dehiscence (E, F)

The loss of seed shattering has been a fundamental characteristic under selection in most legume grain crops to facilitate seed harvesting, while in wild plants, shattering is a fundamental trait to assure seed dispersal. The evolution of non-shattering would have occurred automatically as a result of harvesting that favoured non-shattering mutants in harvested populations which were then sown. Central to the ballistic mechanisms of seed dispersal in pea is the dehiscent pod (single carpel fused along its edges) where the central pod suture undergoes an explosive rupturing along a dehiscence zone (Ambrose *et al.* 2008). In domesticated species, this is removed or delayed. Breeding experiments have shown that the genetic control of seed shattering is often governed by a single locus. Single locus control of pod dehiscence was found in lentil (Erskine 1985), while there are two loci in pea (Weeden 2007). The second most important domestication trait in grain legumes relates to seed dormancy, often called hard-seededness due to the physical barrier of testa water permeability. Moreover seed imbibition has a crucial role in cooking ability of most grain legumes. Hence, reducing seed coat thickness led to a concurrent reduction of seed coat impermeability during the domestication. This was largely overcome in all domesticated grain legumes (Werker *et al.* 1979; Smartt 1990; Weeden 2007). A single recessive locus has been reported in lentil (Ladizinski 1985), while Weeden (2007) has identified two to three loci involved in pea seed dormancy, mediated by testa thickness and structure of testa surface.

Despite of the crucial position of legumes as protein crops in human diet, comparably little is known about their domestication. The loss of fruit shattering has been under selection in most seed crops, to facilitate seed harvesting (Purugganan and Fuller 2009), while in wild plants, shattering is a fundamental trait to assure seed dispersal. The evolution of non-shattering would have occurred as a result of particular methods of harvesting that favoured non-shattering mutants in harvested populations which were then sown. Breeding experiments have shown that the genetic control of seed shattering is often governed by a single locus. Orthologous genes and functions were found to be conserved for seed shattering mechanisms between mono and dicotyledonous plants (Konishi *et al.* 2006) but none yet in legumes. Seed dispersal in wild legumes, is normally by pod dehiscence. Central to the ballistic mechanisms of seed dispersal in *Pisum* is the dehiscent pod (single carpel fused along its edges) where the central pod suture undergoes an explosive rupturing along a dehiscence zone (Ambrose *et al.* 2008). The domestication syndrome (Hammer 1984) in case of pulses applies to increases in seed size, reduction or elimination of pod shattering, and loss of germination inhibition, shoot basal branching and seed toxins and anti-metabolites (Smartt 1990; Zohary and Hopf 2000; Weeden 2007).

The domestication process is only recently beginning to be revealed. One of the best studied legume genera is *Phaseolus*, with *P. vulgaris* widely distributed from northern Mexico to north-western Argentina, and it is characterized by two major eco-geographical gene pools: those of Mesoamerica and the Andes.

These two gene pools show parallel wild and domesticated geographical structure and suggest two independent domestication events (Nanni *et al.* 2011; Bitocchi *et al.* 2013; Schmutz *et al.* 2014; Rodriguez *et al.* 2015) contributing to the modern common bean crop (reviewed in Bellucci *et al.* 2014). In contrary, a recent comparative study of the chickpea cultigen and its progenitor (van Oss *et al.* 2015) revealed a monophyletic origin of the cultigen and provides evidence on gene flow which identifies introgression of the wild into the cultivated genepool. Resequencing of 302 wild, landrace and improved soybean accessions detected 230 selective sweeps and 162 selected copy number variants of which 96 correlated with reported oil QTLs and 21 contained fatty acid biosynthesis genes (Zhou *et al.* 2015). Moreover, some traits and loci were associated with China geographical regions in relation to latitude and indicated differential introgression between groups (Zhou *et al.* 2015).

The domestication of pea has been experimentally tested, both in order to determine the genetic basis which led to the cultivated crop (Weeden 2007), as well as wild pea harvesting (Abbo *et al.* 2008). Similarly to other legumes, also in pea, explosive pod dehiscence and seed dormancy (hard seededness) were barriers to domestication (Smartt 1990; Abbo *et al.* 2014). Based on these morphological and genetic studies, *P. humile/syriacum, P. elatius* and *P. fulvum* were identified as 'wild' germplasm in that they display traits such as dehiscent pods and seed dormancy (thick testa), that are necessary for survival in the wild and undesirable in a domesticated annual crop. In contrast, *P. sativum* including var. *arvense, transcaucasicum* and *asiaticum* generally display indehiscent pods and little seed dormancy, and could be considered domesticated. *P. abyssinicum* is early flowering, with indehiscent pods, moderately large seeds, and lacks seed dormancy. An independent domestication of the Ethiopian (*P. abyssinicum*) pea has been proposed by several authors (Vershinin *et al.* 2003; Jing *et al.* 2010; Polans and Moreno 2009; Ellis 2011) and is supported by chromosomal translocation (Ben-Ze'ev and Zohary 1973; Conicella and Errico 1990). Since pea was never found in Ethiopia and is unlikely to be a native plant, it was most likely introduced via human migration and eco-geographic adaptation, similarly to barley (Pourkheirandish *et al.* 2015). It is known by human genetics demography that southern Arabia (Yemen) and Ethiopia was a melting pot between Africa and Eurasia. The major episodes in the peopling of Arabia took place from north to south in the Late Glacial and, to a lesser extent, the immediate post-glacial/Neolithic (Fernandes *et al.* 2015). Genetic connection between contemporary Ethiopians and Anatolia people as well as archaeological evidence dates the arrival of Near Eastern crop domesticates to the same time period (circa 3,000 years ago), suggests that the direct descendants of the farmers that earlier brought agriculture into Europe may have also played a role in the development of agriculture in the Horn of Africa (Gallego Lorrente *et al.* 2015) including the origin of *P. abyssinicum*. Based on its phenotype, it has been identified as partially and possibly independently domesticated (Vershinin *et al.* 2003; Ellis 2011). Investigation of this question using segregating populations derived from *P. s.* subsp. *sativum* x *P. s.*

subsp. *abyssinicum* crosses revealed no characters in which different loci had been selected during the two domestication processes (Weeden 2007). All of the progeny displayed indehiscent pods, thin testa (non-dormant seeds), medium to large seeds, and relatively early flowering. The recent studies hypothesize that the origins of cultivation of wild cereals and pulses considerably predate their domestication (Hopf 1986; Tanno and Willcox 2006; Willcox *et al.* 2008; Abbo *et al.* 2010). There are several major differences between the wild progenitors of grain legumes and cereals. One concerns the low germination rate imposed by the hard seed coat of legumes (Ladizinsky 1998; Abbo *et al.* 2009; Weeden 2007). While sowing wild wheat or barley with their approx. 50 % germination and profuse tillering may easily produce an agronomic-like stand capable of competing with weeds, the legumes with low (5–20 %) germination rates would result in poor stands likely to be overtaken by aggressive competitors. Second relates to relatively uniform, synchronous spike-ripening in cereal stands in contrast to the prolonged ripening of legumes (Abbo *et al.* 2009).

3.1.1 Analysis of ancient DNA isolated from charred pea seeds from an Early Iron Age

In last two decades with the progress in molecular methods, there is growing body of studies analyzing ancient DNA extracted from excavated material. This preserved DNA has been shown to be useful source of data about past agricultural systems. This allows the analysis of genetic information preserved in archaeological samples, revealing details not possible by any other means. In the 1970´s nucleic acid fragments were detected in emmer from prehistoric Fayum (6,400 years old) and Tutankhamen's tomb (3,300 years old). Ancient DNA was used in numeral studies devoted to domestication of animals including origin of humans, to much less extent to analysis of plant material (Jones and Brown 2000,). It has been recovered from charred wheat grains from and Iron Age (Allaby *et al.* 1994), neolitic dwellings (Schulmbaum *et al.* 1998), dessicated Egyptian barley (Palmer *et al.* 2009) or olive stones (Elbaum *et al.* 2005) to name some. One of the main reasons of such disproportional is the nature of remains, while DNA in bones is relatively well preserved, it is less so in plant material more prone to decay, with only exception of seeds. Until now, the crop domestication and plant archaeological studies focused largely on cereals. Although genes underlying domestication traits in legumes (such as pod dehiscence, seed dormancy, for review see Weeden 2007) have not been identified yet, the analysis of plastid encoded genes can make inference on origin of plant material. With the growth of molecular genetics, phylogenetic variation within species brought a new level of detail to the analysis of domestication. The early studies did not paid sufficient attention to the problems of contamination, which would nowadays be considered a critical issue (Jones and Brown 2000). To recognize genuine aDNA several criteria as used, on the main being mutation frequency increase over time due to

chemical demage (Schlumbaum *et al.* 2008). We have been able to analyze archeological find of 2,572 pea seeds in Hissar settlement in South Eastern Europe (**Smýkal *et al.* 2014**). Pea from Hissar was a distinct crop, stored separately from others. Archaeobotanically, the bulk of the peas recovered at 11th cent. B.C. settlement Hissar belongs to the cultivated pea. Several morphological characteristics indicate this: smooth surface of the seed coat, "coffee-bean-shaped" hilum, broad ellipsoid seed shape, small size range difference between seeds and high 1,000-grain weight of charred seeds. However since wild or semi-wild pea species can be found till today in the area, it could be assumed that seeds could be gathered in wild rather than cultivated. The cultivation of wild pea is highly unprofitable due to low germination rates and pod dehiscence (Abbo *et al.* 2011). However even when collected in sufficient quantities, seed testa permeability will prevent proper imbibition of wild peas resulting in impaired cooking ability, palatability and digestability, so called hard-seededness, occuring even today in some primitive lentil and soybean landraces (**Smýkal *et al.* 2014**). Prolonged time is required to cook less domesticated legumes to a point at which they are palatable, render protein and starch digestible and detoxify anti-nutrients. To confirm our finding, we subjected two samples to molecular analysis. We succeded to amplify and sequence overlapping fragments of *trnLF, trnSG, matK* and *rbcLA* cpDNA regions in total length of 2900 bp. There were 4 informative single polymorphism sites (SNPs) and one 6 bp indel in *matK* sequence, 4 SNPs in *trnSG*, 2 SNPs in *rbcLA* and one SNP in *trnFL*. In addition to these there were also additional substitutions likely attributed to damage of DNA (**Smýkal *et al.* 2014**). We sequenced 6 clones from each of cloned PCR products and in addition to mentioned phylogenetically informative SNPs we found variable single nucleotide differences, which may be the result of polymerase errors or post mortem damage. Since the majority of these 16 substitutions in 2,900 bp sequence are of type 2 transitions (C to T, G to A), it supports the evidence of amplification of truly ancient and not modern pea DNA. These results from deamination of cytosine (and 5-methyl cytosine) to uracil (and thymine), as shown to be associated with postemortem damage (Binladen *et al.* 2006; Ho *et al.* 2007). Deduced from *matK, rbcL* and *trnSG* chloroplast DNA sequenced, sample showed a full homology to cultivated *P. sativum* subsp. *sativum* and partially to wild *Pisum sativum* subsp. *elatius*. The level of the detected mutations in the DNA chain proved that genuine ancient DNA, non-contaminated with the modern pea DNA, was analyzed. Based on molecular analysis of recovered ancient DNA, we assume that material of our study was not wild pea, rather than it represents early pea domesticates (**Smýkal *et al.* 2014**). We speculate that Iron Age pea would be of coloured flower and pigmented testa (**Figure 24**), similar to today´s fodder pea (*Pisum sativum* subsp. *sativum* var. *arvense*), possibly of winter type (**Smýkal *et al.* 2014**). Although charred pea seeds were found in many archaeological sites previously, this is the first report of ancient pea seeds DNA analysis (Mikič 2015).

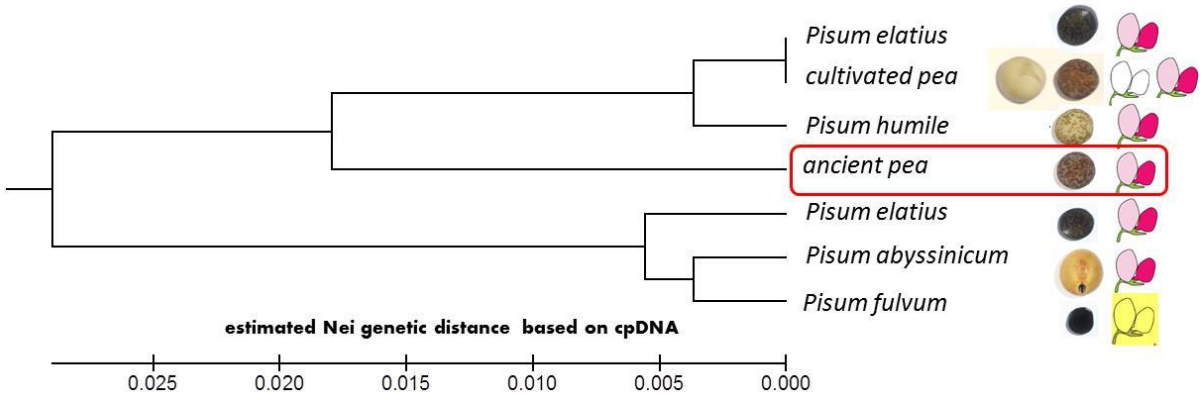


Figure 24

Maximum Likelihood tree for composed trnSG, trnK, matK and rbcL data, showing that analyzed ancient sample could be "an ancestor" of modern cultivated pea. adapted from Smýkal et al. 2014

In our recent study we have detected several wild pea samples with a close affinity to the cultivated pea crop from Armenia and eastern Turkey (Smýkal *et al. submitted*). This might parallel barley domestication where selection for a non-brittle spike trait (Pourkheirandish *et al.* 2015) happened, independently, at least twice in the Levant and Central Asia. Being domesticated in the Fertile Crescent, pea has moved to Europe and in parallel eastward to Iran (then Persia), India and China (Makasheva, 1979; Chimwamurombe and Khulbe, 2011). This might explain our findings and those of Zong *et al.* (2009a,b) of some novel diversity of Chinese origin pea samples. It was proposed that the distinct differentiation of the Chinese *P. sativum* genotypes may in part reflect the historic isolation of agriculture in eastern Asia from that in southern Asia, Europe and northern Africa (Zong *et al.* 2009a,b). Three relatively distinct gene pools of Chinese pea landraces have been differentiated and formed under natural and artificial selections. Gene Pool I is typically represented with resources in Inner Mongolia and Shaanxi in the north central cropping area boundary of China. Gene Pool II comprises landraces from Henan, which is the most northerly and coldest irrigated area of winter sowing. Gene Pool III includes the majority of Chinese landraces. The study of Li *et al.* (2013) suggested that natural selection throughout the Chinese habitat range acted for more than 2 millennia. Similarly analysis of pea potyvirus resistance gene has revealed geographically acting evolution (**Konečná *et al.* 2014**). Our ITS analysis has shown that Chinese samples have 5 different and many display unique ribotypes. Moreover, six out of eight have *trnSG-E2* and *trnSG-E3* chloroplast haplotypes rather than "sativum S1 and *trnSG*-S2 as found in other studied landraces. ITS region pointed out to broader diversity and possible multiple origin of pea crop, as studied landraces were distributed in 8 its haplotype groups shared with wild *P. elatius* and two haplotypes (*its-land1* and *2*) were

exclusive for landraces (**Table1**). This suggests introgression of wild *P. elatius* which might be revealed further by deeper genome-wide sequencing once a reference pea genome is available. Wild *Pisum* in its native range displays a typical winter habit in which plants germinate in the autumn, over-winter in the vegetative state and flower in response to increasing day length in spring (Weller *et al.* 2009; Abbo *et al.* 2013). The obligate or near-obligate requirement for long days suits pea to a winter cropping cycle and has been retained in some forage cultivars. However, most of the cultivated pea accessions from higher latitudes have a quantitative long-day response and are grown as a spring crop (Weller *et al.* 2009). One hypothesis is that pea was domesticated more than once, or that ancient stocks involved more than one cytotype (Abbo *et al.* 2013). Several studies (Palmer *et al.* 1985; Hoey *et al.* 1996; Kosterin and Bogdanova 2008) have proposed so called northern populations of *P. humile* to be putative progenitor of pea crop, however all these studies used limited set of samples. Finally, we can not entirely exclude the possibility of sample misidentification during germplasm maintenance.

3.2 Wild pea as model to study legume domestication process

The origin of the agriculture was one of key points in human history, and a central part of this was the evolution of new plant forms, domesticated crops. The process of crop domestication began 10,000 years ago in the transition of early humans from hunter/gatherers to pastoralists/farmers. The transformation of wild plants into crop plants can be viewed as an accelerated evolution, the result of human and natural selection. Domestication is often described as a quality of plants in which morphological (and genetic) changes are found amongst cultivated in comparison to wild populations. These domestication triggered changes represent adaptations to cultivation and human harvesting, accompanied by genetic changes. Common set of traits have been recorded for unrelated crops, named **domestication syndrome** (Hammer 1984; Zohary and Hopf 2000) which is viewed as result of convergent evolution (Lenser and Theissen 2013). These include loss of germination inhibition and increase of seed size, linked to successful early growth of planted seeds. However the loss of natural seed dispersal is considered the single most important domestication trait, because it makes a species dependent upon the human farmer and in turn allows farmer to harvest. Recently, the identity of some of the responsible genes has been revealed (reviewed in Hancock 2012) by analysis of crop to wild crosses and recently by association mapping. In species with sufficiently known and mainly small genomes, re-sequencing of both wild and domesticated species allows genome wide searches for genes related to domestication (Gross and Olsen 2010; Olsen and Wendel 2013).

Despite of crucial position of legumes, as protein crops, in human diet, comparably little is known on their domestication. Also in legumes, domestication-related traits include changes in plant architecture,

gigantism, reduced seed dispersal and loss of seed dormancy. Experimental growing of wild peas and lentil, have demonstrated that both seed dormancy and pod dehiscence cause poor crop establishment via reduced germination as well as dramatic yield losses via seed shattering (Abbo *et al.* 2011). Genetic analysis has been performed in mungbean (Isemura *et al.* 2010, 2012; Kongjaimun *et al.* 2012) resulting in identification of QTLs for 38 domestication related traits. Extensive analysis by single nucleotide polymorphisms derived from 670 genes revealed evolutionary history of cultivated pigeonpea (Kassa *et al.* 2012). The domestication process is only recently beginning to be revealed. One of the best studied genera is *Phaseolus*, with *P. vulgaris* widely distributed from northern Mexico to north-western Argentina, and it is characterized by two major eco-geographical gene pools: those of Mesoamerica and the Andes. These two gene pools show parallel wild and domesticated geographical structure and suggest two independent domestication events (Nanni *et al.* 2011; Bitocchi *et al.* 2013; Schmutz *et al.* 2014; Rodriguez *et al.* 2015) contributing to the modern common bean crop (reviewed in Bellucci *et al.* 2014). A recent comparative study of the chickpea cultigen and its progenitor (van Oss *et al.* 2015) revealed a monophyletic origin of the cultigen and provides evidence on gene flow which identifies introgression of the wild into the cultivated genepool. On the other hand two types of cultivated chickpea, *desi* and *kabuli*, display large genomic differences which support a two origin scenario (Varshney *et al.* 2013; Parween *et al.* 2015). Resequencing of 302 wild, landrace and improved soybean accessions detected 230 selective sweeps and 162 selected copy number variants of which 96 correlated with reported oil QTLs and 21 contained fatty acid biosynthesis genes (Zhou *et al.* 2015). Moreover, some traits and loci were associated with China geographical regions in relation to latitude and indicated differential introgression between groups (Zhou *et al.* 2015). Comparably smaller number and mostly anonymous markers were so far used for mapping of pea domestication traits (Weeden 2007).

3.2.1 Pod dehiscence

The loss of fruit shattering has been under selection in most seed crops, to facilitate seed harvesting (Purugganan and Fuller 2009), while in wild plants, shattering is a fundamental trait to assure seed dispersal. The evolution of non-shattering would have occurred as a result of particular methods of harvesting that favoured non-shattering mutants in harvested populations which were then sown. Breeding experiments have shown that the genetic control of seed shattering is often governed by a single locus. Orthologous genes and functions were found to be conserved for seed shattering mechanisms between mono and dicotyledonous plants (Konishi *et al.* 2006) but none yet in legumes. Seed dispersal in wild legumes, is normally by pod dehiscence. Central to the ballistic mechanisms of seed dispersal in *Pisum* is the dehiscent pod (single carpel fused along its edges) where the central pod suture undergoes an explosive rupturing along a dehiscence zone (Ambrose *et al.* 2008). As the pod

matures and the cells dry, the pod walls shrink in opposite directions generating stresses that are released when the sutures rupture. In domesticated species, this is removed or delayed. Single locus control of pod dehiscence was found in lentil (Ladizinski 1998), while two loci in mungbean (Isemura *et al.* 2012), yardlong bean (Kongjaimun *et al.* 2012), one controlling the number of twists along the length of the shattered pod, and second the percentage of shattered pods, similarly to two loci found in pea (Weeden *et al.* 2002, Weeden 2007), and common bean. In pea, the trait is semi-dominant and monogenic (*Dpo1* locus) with linkage to LGIII (Weeden 2007), while additional (*Dpo2)* and *Gp* loci (yellow pod) have been detected only in certain crosses (Weeden *et al.* 2002). Recently, common bean SHATTERPROOF and INDEHISCENCE homologous genes have been identified (Nanni *et al.* 2011, Gioia *et al.* 2013), although not yet shown to be directly involved in pod shattering. Both genes belong to the family of MADS-box genes, with wide range of functions, such as in the formation of flowers, control of flowering time and were shown to be involved in fruit shattering in *Arabidopsis* (Liljegren *et al.* 2000). Recently, two different genes have been identified to be involved in pod dehiscence in soybean. One of them is the dirigent-like protein, designated as *Pdh1,* promoting pod dehiscence by increasing the torsion of dried pod walls, which serves as a driving force for pod dehiscence under low humidity (Funatsuki *et al.* 2014). The functional gene *Pdh1*, was highly expressed in the lignin-rich inner sclerenchyma of pod walls, especially at the stage of initiation in lignin deposition. A survey of soybean germplasm revealed that *pdh1* was frequently detected in landraces from semiarid regions and has been extensively used for breeding in North America (Funatsuki *et al.* 2014). Yet, another NAC family gene SHATTERING1-5 was identified (Dong *et al.* 2014) in soybean. This gene functionally activates secondary wall biosynthesis and promotes the significant thickening of fibre cap cells of the pod ventral suture secondary walls. The differences between wild and cultivated soybean is within promoter region and subsequently expression level (Dong *et al.* 2014). Using the recombinant inbred lines (RILs) of wild JI64 cross to cultivated JI92 we have been working on mapping of *Dpo1* locus as well as comparative transcriptomical analysis of pod wall. This resulted in identification of differentially expressed candidate genes (DEGs) one of them correctly placed at LG III. This is currently under further study.

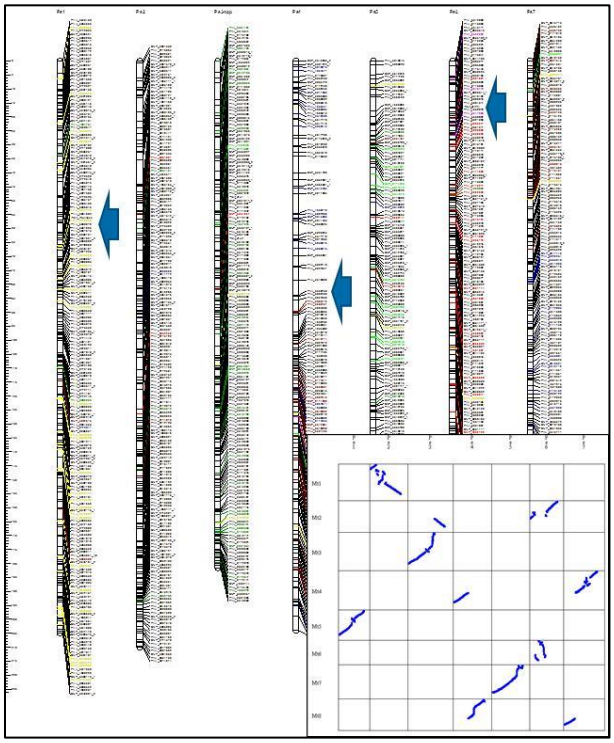
Figure 25

Macrograph of mature indehiscent pod of cultivated pea cv. Cameor (E) and dehiscent pod of wild pea P. elatius (JI64) (F) pod section stained with fluoroglucinol to visualize lignin deposition at pod suture of cultivated (G) and wild (H) peas.

3.2.2 Seed dormancy and seed coat impermeability – key mechanism of physical dormancy

Seed dormancy had played a significant role in evolution and adaptation of plants, since germination is one of the key moment in plant's life. It determines the outset new generation, where and when plant grow (Nonogaki *et al.* 2014; **Smýkal *et al.* 2014**). In wild there are many dormant seeds, which may take several months or even years before germination (Bewley 1997; Baskin *et al.* 2000; Finch-Savage *et al.* 2006). Several dormancy classes were defined among plant species, which can be divided into morphological, physiological, morpho-physiological, physical, and combinational dormancy (Nikolaeva 1969; Baskin 2003; Finch-Savage *et al.* 2006). Physical seed dormancy is caused by one or more water-impermeable layers cell in seed coat (Baskin *et al.* 2000). Seed coat (testa) functions as a physical barrier, as a pathway of water intake and it also protects against microbial attacks. On the other hand, reducing thickness of seed coat leads to current decrease seed coat impermeability during domestication of crop legumes (**Smýkal *et al.* 2014**) and brings advantage in food processing and crop production. The permeability of the testa, being the part of the seed that comes into contact with the ambient water, plays a central role in water uptake (Koizumi *et al.* 2008; **Smýkal *et al.* 2014**; Weitbrecht *et al.* 2011). Histological analysis of the seed coat in *Medicago truncatula* revealed changes in cell wall thickness in the outer integuments throughout seed development (Verdier *et al.* 2013a). In *Arabidopsis* and *Melilotus* (legume), seed permeability was enhanced by mutations affecting suberin biosynthesis (Verdier *et al.* 2012). Similarly, in *M. truncatula* and pea, cells of the outer integument showed abundant accumulation of polyphenolic compounds; which upon oxidation may impact seed permeability (Moise *et al.* 2005; Mayer 1974; Werker *et al.* 1979). Recently, the seed coat structure of the model legume *Medicago truncatula* has been characterized (Wang and Grusak 2005). Comparably more is known on *Arabidopsis* seed development, with numerous *transparent testa (tt)* and *tannin deficient seed (tds)* mutants (Appelhagen *et al.* 2014). Many of these indicating the important role of proanthocyanidins and flavonoid pigments in testa development (Graebner *et al.* 2012) including *Medicago* (Liu *et al.* 2014). Although hardseededness was largely overcome in all domesticated grain legumes (Werker *et al.* 1979; Smartt 1990; Weeden 2007), it appears in lentil or soybean depending on the cultivation conditions. Moreover hardseededness often remains in fodder legumes, where various physical and chemical methods are used to overcome it. A wide range of germination rates was reported amongst wild accessions of lablab, which suggest that domestication drew upon existing genetic variation (Maass and Usongo 2007). We have obtained similar results while testing wild *Pisum* accesions (*unpublished*, Juračková 2012). Development of pea and model legume *Medicago truncatula* seeds have been well characterized (Hedley *et al.* 1994; Thompson *et al.* 2008), also on transcriptomic and proteomic levels (Gallardo *et al.* 2007).

Seed dormancy was identified as monogenic trait in mungbean (Isemura *et al.* 2012); while six QTLs were detected in yardlong and rice bean (Kongjaimun *et al.* 2012). In pea, Weeden (2007) has identified two to three loci involved in seed dormancy, via testa thickness and structure of testa surface. Two genes involved in seed coat water permeability were recently identified in soybean. One of them, *GmHs1-1*, encodes a calcineurin-like metallophosphoesterase transmembrane protein, which is primarily expressed in the Malpighian layer of the seed coat and is associated with calcium content. The transition from impermeability to permeability in domesticated soybean was caused by artificial selection of a point mutation in *GmHs1-1* (Sun *et al.* 2015). Independently of this, *qHS1*, a quantitative trait locus for hard seededness in soybean, was identified as endo-1,4-β-glucanase (Jang *et al.* 2015). This genes seems to be involved in the accumulation of β-1,4-glucan derivatives such as xyloglucan and/or β-(1,3)(1,4)-glucan that reinforce the impermeability of seed coats in soybean. Using the above mentioned 126 recombinant inbred lines (JI64xJI92) and genome wide DARTseq analysis we have produced ultra-high density map (**Figure 26**) placing seed testa thickness (measured by micrometer or scanning electron photography) and germination (as test of seed dormancy) traits onto LG I and LG III. This genome-wide analysis has in contrast to Weeden (2007) resulted in quantitative rather than qualitative determination of seed dormancy trait.



Trait	Chr	Position (cM)	LOD
Testa thickness	Ps1	62.3	7.602
Testa thickness	Ps2	30.6	2.689
Testa thickness	Ps3(botto	4.322	5.421
Testa thickness	Ps3(top)	16.06	4.366
Testa thickness	Ps4	84.5	3.664
Testa thickness	Ps6	10.89	6.599
percent_germ	Ps1	39	3.259
percent_germ	Ps4	62.93	2.357
mean_germ_time	Ps7	17.72	2.701
arcsin	Ps1	39.2	3.295
logit	Ps1	28.12	2.361
coefficient_velocity	Ps2	210.43	3.601
coefficient_velocity	Ps7	20.14	2.479

Figure 26

Ultra-high density map based on 126 RILs (JI64xJI92) 7109 SNPs, 9259 PAVs produced from DARTseq data using Multipoint software and aligned to Tayeh et al. 2015 map, indicating testa thickness and percent germination traits on LG I, IV and VI (blue arrows).

3.2.3 Legume seed coat morphology

Fundamental to the study of seed coat developmental genetics is its morphology. Physical seed dormancy is generally caused by the presence of water-impermeable layers of palisade cells in the seed coat, identified early on by anatomical studies (Spury 1963, 1964; Finch-Savage and Leubner-Metzger 2006). Histological analysis of the seed coat in *Medicago truncatula* revealed changes in cell wall thickness in the outer integuments throughout seed development (Verdier *et al.* 2013a). In *Arabidopsis* and *Melilotus* (legume), seed permeability was enhanced by mutations affecting suberin biosynthesis (in Verdier *et al.* 2012). Similarly, in *M. truncatula*, cells of the outer integument showed abundant accumulation of polyphenolic compounds; which upon oxidation may impact seed permeability (Moise *et al.* 2005). As morphological features of the legume seed coat are relatively insensitive to environmental conditions, there are used for taxonomy. Yet, information on the differentiation and development of seed coats is generally lacking. Recently, the seed coat structure of the model legume *Medicago truncatula* has been characterized (Wang and Grusak 2005). Comparative anatomical and histochemical analysis of wild and domesticated pea seed coat revealed differences in testa thickness, surface structure and chemical composition (**Smýkal *et al.* 2014**). Comparably more is known on *Arabidopsis* seed development, with *transparent testa (tt)* and *tannin deficient seed (tds)* mutants. Many of these indicating the important role of proanthocyanidins and flavonoid pigments in testa development (Graebner *et al.* 2012).

Within our running project, metachromatic staining with toluidine blue revealed that most of the non-dormant pea genotypes exhibit high level of polyanionic cell wall components, most likely carried out by pectins with free carboxyl which are not methylated nor linked with other bonds. The non-dormant, but well pigmented JI92 showed similar color patterns as the dormant type genotypes with neutral sugars being the most abundant. To find out more specifically which types of pectins and other polysaccharides are localized in cell walls of wild and non-dormant pea testas, a set of immunodetection reactions were performed. Most of the non-dormant-domesticated (Cameor, JI92) as well as dormant-wild (JI64, VIR320) pea genotypes excluding Terno (domestricated) and L100 (wild) are positive in counterpalisade cells for partially methylated (JIM5) and highly methylated (JIM7) homogalacturonans (HGs). All genotypes showed signal for aniline blue in the light line and macrosclereids, especially in their outer part composed in majority by the secondary cell walls. The strongest signal was observed in VIR320 genotype. The signal pattern of aniline in the lens and the side area of the seed is very similar to that observed in the hilum (except there is no counterpalisade layer) with the highest intensity of signal in the light line of dormant type pea genotypes and JI92. There is some supposition for callose deposition in the light line area, but aniline blue is not specific just for callose detection, but might interact also with other cell wall compounds such as various xylans.

3.2.4 Role of phenolic compounds in seed testa impermeability

Positive correlation in content of phenolics, the requirement of oxidation and the activity of catechol oxidase in relation to seed dormancy (germination) in wild versus domesticated pea seeds have been shown by Marbach and Mayer (1974) and Werker *et al.* (1979). Recently, epicatechin, cyanidin 3-O-glucoside, and delphinidin 3-O-glucoside were isolated in wild compare to cultivated soybean seed coats (Zhang *et al.* 2011) with epicatechin in significant positive correlation with hardseededness. Proanthocyanidins (PAs), also known as condensed tannins, are oligomeric and polymeric flavonoids, a large group of plant phenolic secondary metabolites. The presence of PAs in seed coats can be assessed by the appearance of brownish coloration, which is the result of PA oxidation by polyphenol oxidase (Marles *et al.* 2008). The insoluble PAs are the result of oxidative cross-linking with other cell components. Variation in PA content in the pea seeds has been reported (Troszynska *et al.* 2002; Jin *et al.* 2012) but not in relation wild versus cultivated peas. These compounds play also important roles in defence to pathogens, as well as affect quality of products, and because of the health benefits are of industry and medicine interest. Mutations in either structural or regulatory genes lead to a loss of pigmentation. Among them, MYB, basic helix-loop-helix (bHLH) transcription factors or WD40 proteins are known regulators of anthocyanin biosynthesis. MYB-transcription factor was identified as key regulator of PAs biosynthesis in *M. truncatula* seed coat (Verdier *et al.* 2012). Similarly, UDP-glycose:flavonoid-3-O-glycosyltransferase and anthocyanindin reductase were identified to be involved in seed testa colour in soybean (Kovinich *et al.* 2012) and *Medicago* (Pang *et al.* 2007) by metabolomics and transcriptomic analysis. Interestingly, Mendel A gene conferring flower colour and testa pigmentation has been identified as bHLH transcription factor (Hellens *et al.* 2010). While, B gene of pea encodes a defective flavonoid 3´, 5 ´-hydroxylase, and confers pink flower colour, by control of hydroxylation of flavonoid precursors (Moreau *et al.* 2012). However none of these mutations result in alteration of seed dormancy.

In collaboration with group of Dr. P. Bednář (Department of Analytical Chemistry, UP Olomouc) we have conducted comprehensive analysis of extracts of pea seed coats by means of ultra-performance liquid chromatography coupled with electrospray ionization high resolution tandem mass spectrometry and ion mobility mass spectrometry (UPLC/ESI-IMS-MS) and direct analysis using laser desorption/ionization high resolution tandem mass spectrometry and ion mobility (LDI-IMS-MS). Several markers of dormancy were identified belonging to flavonoids (quercetin and polymers of gallocateching) were found and further analysis is in progress using MALDI imaging. Results from LDI-IMS-MS in negative ion mode pointed out the significantly higher signals of anions of particular hydroxylated fatty acids, i.e. hydroxyhexacosanoate, dihydroxyhexacosanoate, dihydroxyheptacosanoate and dihydroxyoctacosanoate in dormant JI64 compared to non-dormant JI92. Targeted analysis of

composition of separated parts of seed coat would allow to understand the process of imbibition in deep. For such experiments, however, a method allowing separation of very small parts of seed coat tissue typically in dimensions from tens to few hundreds of micrometers is needed. Two step dissection involving a cross cryosection of seed coats in first stage followed by a laser microdissection of selected parts cell lines from the cross sections in second stage was done in cooperation with group of Dr. A. Soukup (PřF UK, Praha) at the Department of Plant Developmental Genetics, Institute of Biophysics, CAS (Dr. R. Hobza and W. Jesionek). The micro-dissected samples (dissected cell walls prepared from cuticle and inner parts of seed coat. A more comprehensive study of differences in sugar profile in microdissected parts of seed coats is currently underway.

3.2.5 Genetic relationship between cultivated pea and its wild progenitor

Being domesticated in the Fertile Crescent, pea has moved to Europe and in parallel eastward to Iran (then Persia), India and China (Makasheva 1979; Chimwamurombe and Khulbe 2011). This might explain our findings (**Smýkal *et al.* 2011**) and those of Zong *et al.* (2009a,b) of some novel diversity of Chinese origin pea samples. The study of Li *et al.* (2013) suggested that natural selection throughout the Chinese habitat range acted for more than 2 millennia. Peas are grown widely throughout China, separated into a winter-sown region in southern China, and a spring-sown region in northern China. Similarly analysis of pea potyvirus resistance gene has revealed geographically acting evolution (**Konečná *et al.* 2014**). This might parallel barley domestication because selection for a non-brittle spike trait (Pourkheirandish *et al.* 2015) happened, independently, at least twice in the Levant and Central Asia. This suggests introgression of wild *P. elatius* which might be revealed further by deeper genome-wide sequencing once a reference pea genome is available. Wild *Pisum* in its native range displays a typical winter habit in which plants germinate in the autumn, over-winter in the vegetative state and flower in response to increasing day length in spring (Weller *et al.* 2009; Abbo *et al.* 2013). The obligate or near-obligate requirement for long days suits pea to a winter cropping cycle and has been retained in some forage cultivars. However, most of the cultivated pea accessions from higher latitudes have a quantitative long-day response and are grown as a spring crop (Weller *et al.* 2009). The genome-wide DARTseq analysis identified a group of wild *P. elatius* from regions of Armenia, Georgia, Crimea, Morocco, Algeria and mainly south-eastern Turkey-Syria, with the closest proximity to the cultivated pea genepool. These showed variable proportion of wild and domesticated alleles, ranging from less than 0.1 to 0.5 of wild (**Figure 14**). There are two main possible scenarios. Either these are the most likely progenitors of domesticated pea, or they represent early escapes from cultivation with a reversion to the wild type. The first option is more plausible since all these accessions display a strong pod dehiscence and seed dormancy phenotypes which are signatures of wild pea origin. As both traits are

dominant (Weeden 2007) and their recessiveness is loss of function, it is less likely to reverse as gain of function mutation. Once the genes underlying any of these domestication traits are identified, these samples would offer interesting material to test domestication origins. In this context, Ladizinsky and Abbo (2015) based its occurrence in disturbed habitats suggested a feral origin for *P. sativum* subsp. *humile* var. *humile ,* i.e. it is proposed to be a reversal from a past domesticated pea crop (Abbo *et al.* 2013). Yet another scenario might involve gene-flow between wild and domesticated crop, as shown in *Phaseolus* (Papa and Gepts 2003) and recently chickpea (van Oss *et al.* 2015). One hypothesis is that pea was domesticated more than once, or that ancient stocks involved more than one cytotype (Abbo *et al.* 2013). Several studies (Palmer *et al.* 1985; Hoey *et al.* 1996; Kosterin and Bogdanova, 2008) have proposed so called northern populations of *P. humile* to be putative progenitor of pea crop, however all these studies used limited set of samples. Finally, we can not entirely exclude the possibility of sample misidentification during germplasm maintenance.

Besides the common crop (*Pisum sativum* L.) there is the pea of Ethiopian origin (*P. abyssinicum*) which has never been found in the wild outside of the cultivation in Ethiopia and in Yemen. *P. abyssinicum* has all the domestication traits (non-dehiscent pod, non-dormant seeds) but with a distinct allelic composition and karyotype (Vershinin *et al.* 2003; Jing *et al.* 2007). It has a very low morphological and genetic diversity suggesting an extreme bottleneck supported by our genome-wide analysis. Furthermore our analysis has shown *P. abyssinicum* to have a closer affinity to the wild *P. elatius* genepool rather than to cultivated *P. sativum*, in agreement with a retrotransposone insertion based analysis and gene sequencing (Jing *et al.* 2007, 2010). Of note, one sample of *P. elatius* (IG52520) from the Antakya region of Turkey, close to Syria and Lebanon, and occurring sympatrically with *P. fulvum* and *P. elatius* displays a proximity to *P. abyssinicum*. Since none of the wild *Pisum* species have ever been found in Ethiopia or the Yemen, we speculate that an early flowering, less photoperiod sensitive, hybrid suited to drought conditions (Weller *et al.* 2012) was moved southwards along trading routes, as suggested by Vershinin *et al.* (2003) and Ellis (2011).

Comparison of datasets for the wild (53) and cultivated (64) pea samples using threshold P= $5x10^{-8}$ and 2,421 DART fragments (SNPs) with identified homologues of *Medicago truncatula* v4.0 genome resulted in detection of a total of 234 SNPs. Of these 107 were chloroplast genome fragments, particularly of the *rps12* and *rps7* intergenic spacer and photosystem II reaction center protein D (44 and 60 SNPs respectively) which was diverse in the wild (*P. s.* subsp. *elatius/humile*) genepool, but fixed in cultivated germplasm. The remaining 127 SNPs most of them at different genes were spread across all *Medicago* chromosomes: 24 on Mt-chr.1, 18 Mt-chr.2, 22 Mt-chr. 3, 14 on Mt-chr. 4, 22 on Mt-chr. 5, only 1 on Mt-chr.6, 16 on Mt-chr.7 and 9 on Mt-chr. 8 (**Figure 27A**). Both distance and Bayesian based

analysis of diversity structure have indicated that samples from Armenia (Chakaten, Garni, Noravank, IG140562, IG141436), Georgia (Caucasus), Tunisia (IG108291) and eastern Turkey (W6_26112, W6_2101, W6_2107), south-eastern Turkey (Elmali, IG52507, IG52508, IG52414, IG52442, IG52520, IG52496), Israel (JI1794), Syria (JI2538 originally designated as *P. fulvum*), Crimea (JI2629), Morocco (IG111992), Algeria (IG64350) and Italy-Sardinia (JI3271) are the closest relatives to the domesticated genepool, sharing between 0.5 to 0.8 alleles. PCA analysis based on SNP variation showed two major components jointly explaining 28.4% of allelic diversity and supporting the continuum between wild and domesticated genepools (**Figure 12**), while *P. fulvum* and interestingly also *P. abyssinicum* were clearly separated. Comparative analysis of wild and domesticated samples of 13.2k SNP array dataset has identified 35 SNPs in various genes (**Figure 27B**), positioned at all linkage groups: 3 on Ps LGI, 1 on LGII, 8 on LGIII, 5 on LGIV, 8 on LGV, 6 on LGVI and 4 on LGVII. Using annotation of Tayeh *et al.* (2015a) various gene types were identified including flowering locus *FTc* on LG V (Hecht *et al.* 2011).

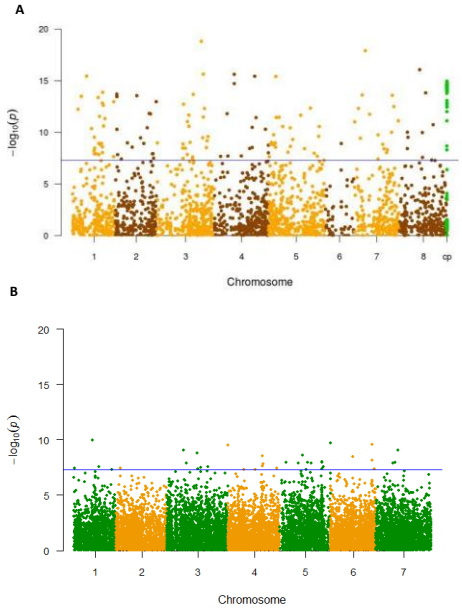


Figure 27

A) *Manhattan plot of the wild (53) and cultivated (64) pea samples using threshold P= 5x10-8 and 2,421 DART fragments (SNPs) with identified homologues of Medicago truncatula v4.0 genome resulting in 234 SNPs. Of these 107 were chloroplast genome fragments.*

B) *Comparative analysis of wild and domesticated samples of 13.2k SNP array dataset resulting in 35 differential genes.*

Although neither DARTseq nor SNP data allow for proper analysis of domestication genomic-wide selection as compared to re-sequencing based analysis as performed in soybean (Zhou *et al.* 2015; Han *et al.* 2016; Wang *et al.* 2016) and chickpea (Bajaj *et al.* 2015; Kujur *et al.* 2015), comparison of allelic diversity between cultivated and wild *P. elatius* genepools has identified several genes subjected to selection. One of them, flowering locus *FT* was previously identified (Hecht *et al.* 2011) to be implicated in photoperiod response. Second, transcription factor GTE6 homolog at LGVI is transcription activator that plays a role in the promotion of seed germination by both negatively and positively regulating the abscisic acid and phytochrome A transduction pathways, respectively (Duque

and Chua, 2003). Finally, a homolog of glucan endo-1,3-ß-glucosidase identified on LGVII might play a role in testa rupture, a process that involves release of coat-enhanced dormancy (Leubner-Metzger, 2005) as recently shown in soybean (Jang *et al.* 2015) although no such trait was mapped to LGVII in pea (Weeden, 2007). This analysis will be further refined once pea genome data as well as mapped domestication trait candidate genes are available, allowing comparative re-sequencing of selected wild pea samples.

Chapter 4

Use of pea diversity for breeding

4.1 Example of wider germplasm use – the case of potyvirus resistance gene

Commenting on: Konečná E, Šafářová D, Navrátil M, Hanáček P, Coyne C, Flavell A, Vishnyakova M, Ambrose M, Redden R and ***Smýkal P*** *(2014) Geographical gradient of the eIF4E alleles conferring resistance to potyviruses in pea (Pisum) germplasm. PLOS One 9, 3: e90394.*

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

The use of genetic resistance is considered to be the most effective and sustainable strategy to control plant pathogens in agricultural practice (Wang and Krishnaswamy 2012). Domestication of wild plants, led to crop distribution away from their original centres (Vavilov 1926, 1992; Lovisolo *et al.* 2003; Abbo *et al.* 2012) followed also by pathogens. Long before plants were domesticated and grown as monocultures, plant pathogens were co-evolving with wild plants growing in mixed species communities. Evolution has continued to occur within domesticated plants growing as selected genotypes in denser populations than in the wild. Furthermore, domestication of wild plants, led to crop distribution away from their original centres (Zohary and Hopf 1993) followed also by the eventual introduction of pathogens. This co-evolutionary process shaped both plants and their pathogens, including viruses (Lovisolo *et al.* 2003; Le Gall *et al.* 2011). During early domestication,

cultivated plants would have been invaded by viruses from their wild ancestors grown in abundance nearby. Further virus evolution occurred within domesticated plants growing as selected genotypes in denser populations than in wild. Beside abiotic stresses, plant pathogens are a major constraint to agriculture and threaten global food security. After fungi, viruses cause the most devastating diseases worldwide causing economically significant losses of crop yield and quality (Scholthof *et al.* 2011). Moreover, on-going climate change could accelerate temporal and spatial diseases spread and their severity. Detailed knowledge about natural and cultural pathosystems is necessary to avoid and/or limit negative impact on crop production by development of resistant/tolerant cultivars and application of appropriate agro-technical measurement and agricultural zoning. Among the biggest and the most important viral genera are the Begomovirus (Geminiviridae, 192 species) and Potyvirus (Potyviridae, 146 species). *Pea seed borne mosaic virus* (PSbMV) is a member of *Potyvirus* genus. Legumes are its natural hosts, including economically important host plants are pea, lentil, faba bean, and chickpea (Latham and Jones 2001). Discovered in Czechoslovakia (Musil 1966) it has been recently reported worldwide, causing serious yield losses. Due to its seed borne transmission PSbMV represents serious phytosanitary risk both for germplasm maintenance and seed production. The virus causes various symptoms depending on the host and virus isolate/pathotype, such as downward rolling of leaflets, the transient clearing and swelling of leaf veins, chlorotic mosaics, stunting, and delayed flowering. PSbMV is transmitted between plants in a non-persistent manner by the aphids and then infect seeds (Hampton and Mink 1975). The full genome sequences of three isolates are known and the P3-6K1 and VPg (virus-genome linked protein) proteins have been identified as viral determinants (Hjulsager *et al.* 2006), responsible for physical interaction with host *eIF4E* or *eIF(iso)4E* proteins and are critical for viral infection. The studies suggest that potyviruses may selectively use either *eIF4E* and/or *eIF(iso)4E* to achieve infection (Ruffel *et al.* 2006) although the exact mechanism is yet to be elucidated. Recessive resistance of pea to PSbMV correspond with the matching-allele model, proposed as the common system explaining the interaction between potyviruses and plant hosts (Fraile and Garcia-Arenal 2010). The *eIF4E* allelic diversity has been systematically screened in various crop collections, such as pepper (Rubio *et al.* 2009; Ibiza *et al.* 2010; Jeong *et al.* 2012), melon (Nieto *et al.* 2007), tomato (Rigola *et al.* 2009), barley (Hofinger *et al.* 2011) and pea (**Smýkal *et al.* 2010**; **Konečná *et al.* 2014**). A possible link between potyviruses radiation and origin of agriculture was suggested by Gibbs *et al.* (2008) and showed human mediated spread of crop hosts, followed by further diversification of viruses. We have combined our knowledge of pea germplasm diversity with that of the *eIF4E* gene for virus resistance and screened 2,803 accessions with known geographical origin and 149 accessions of wild *Pisum sp.* (**Konečná *et al.* 2014**). Using length polymorphism of *eIF4E* intron 3 we detected four alleles ($eIF4E^{A-B-C}$ and $eIF4E^{S}$). Sequencing resulted in the detection of 35 haplotypes with 156 polymorphic sites

(**Figure 28**). Original *eIF4E*^{A} allele conferring resistance to P1 PSbMV was found in 53 accessions (1.9%), of which 28 were modern varieties with single source of resistance, while 15 were landraces from India, Afghanistan and Nepal and 7 were from Ethiopia (**Figure 29**). A newly discovered allele *eIF4E*^{B} was present in 328 accessions (11.7%) originating from Ethiopia (29%), Afghanistan (23%), India (20%), Israel (25%) and China (39%). The deduced amino acid sequence of *eIF4E*^{B} allele was identical to susceptible *eIF4E*^{S} allele except of two amino acid exchanges. The *eIF4E*^{C} allele was detected in 91 accessions (3.2%) from India (20%), Afghanistan (33%), Iberian Peninsula (22%) and Balkan (9.3%). Haplotype network analysis indicated different evolution of *eIF4E* alleles (**Figure 30**). This suggests that substantial diversity in *eIF4E* gene can be anticipated in available pea germplasm, which would only be fully revealed by sequencing. It is notable that none of the 146 tested wild *Pisum sp.* samples had either *eIF4E*^{A} or *eIF4E*^{B} alleles, but at the amino acid level all corresponded to the susceptible *eIF4E*^{S} allele, despite substantial polymorphism both in introns and exons, suggesting selection for functionality. An intermediate state was found in JI261 *P. sativum* subsp. *elatius* from Turkey, which showed a sensitive allele on the amino acid level, while the nucleotide level showed transitory stage to *eIF4E*^{B} allele. The finding of intermediate allele is in agreement with published data obtained in *Arabidopsis thaliana* (Le Galle *et al.* 2011) suggesting a distinct mode of evolution of resistance in wild species in comparison to crops and supporting the scenario of potyviruses spread upon beginning of agriculture which brought plants into dense monocultures (Gibbs *et al.* 2008). Moreover, we do not currently know the natural reservoir for potyviruses in wild legumes found in abundance in natural habitats.

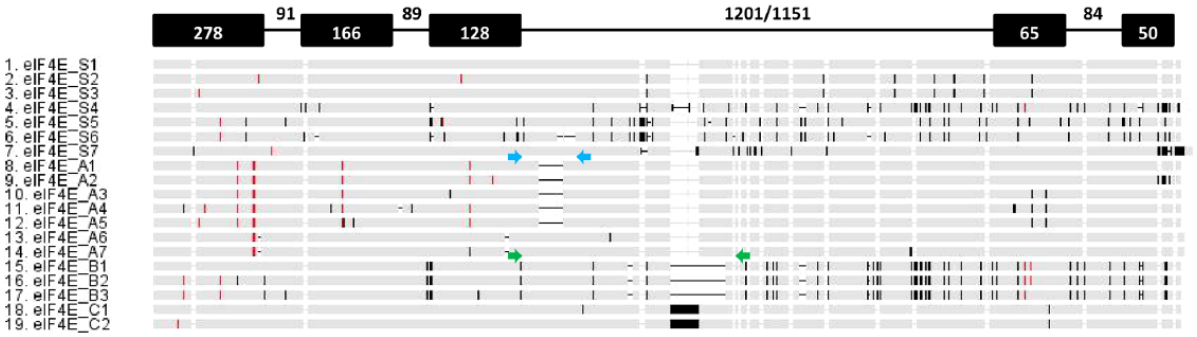


Figure 28 *Schematic representation of sequence alignment of the all 19 identified protein eIF4E alleles. Four principal variants identified by intron 3 polymorphism are designated A,B,C and S, numbers indicate the respective allelic variant. Black bars indicate polymorphism nucleotides within both exons and introns, red bars indicate polymorphism leading to amino acid exchanges. Horizontal lines indicate insertions/deletions. The heading line indicates nucleotide numbers and exon (solid black boxes)-intron (lines) positions and sizes. Blue arrows indicate primer A combination (Ps-eIF4E-750F and Ps-eIF4E-586gR) and green arrows indicate primer B combination (Ps-eIF4E-750F and Ps-eIF4E-1270R). taken from Konečná et al. 2014*

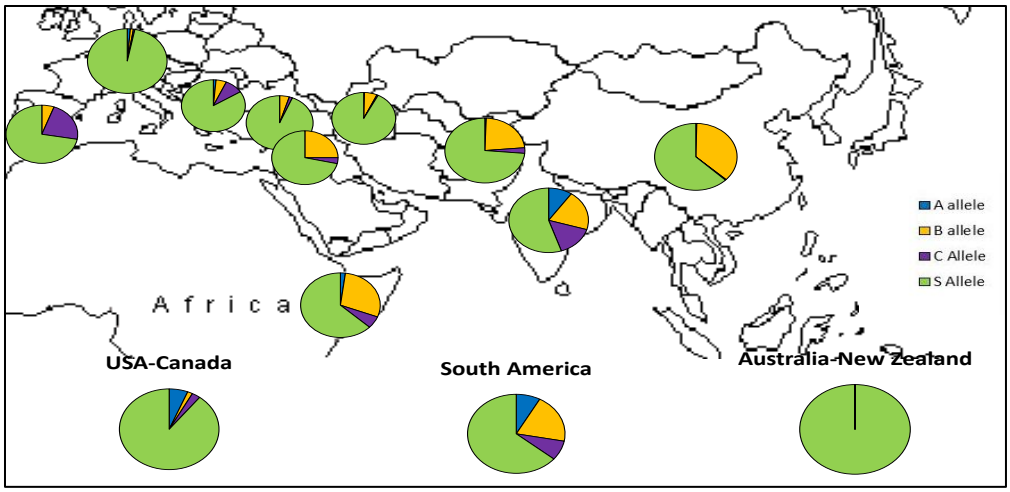


Figure 29

Geographical distribution of four eIF4E alleles expressed as percentage of total. taken from Konečná et al. 2014

These data highlight the importance of Ethiopian, Central Asia and Chinese regions, as secondary centers of pea diversity, corresponding with the diversity of the pathogen. With current use of the single resistance *eIF4E*^{A} allele in pea breeding programmes and anticipated virus evolution to overcome it, these alleles offer potential alternative protection. The existence of resistance alleles only in domesticated pea genepool leads us to speculate on the mutation origin during the early steps of cultivation. The series of alleles identified in this study provide the basis for the testing of various potyviruses and pathotypes to reveal possible co-evolution of potyviruses and its pea host (**Konečná *et al.* 2014**).

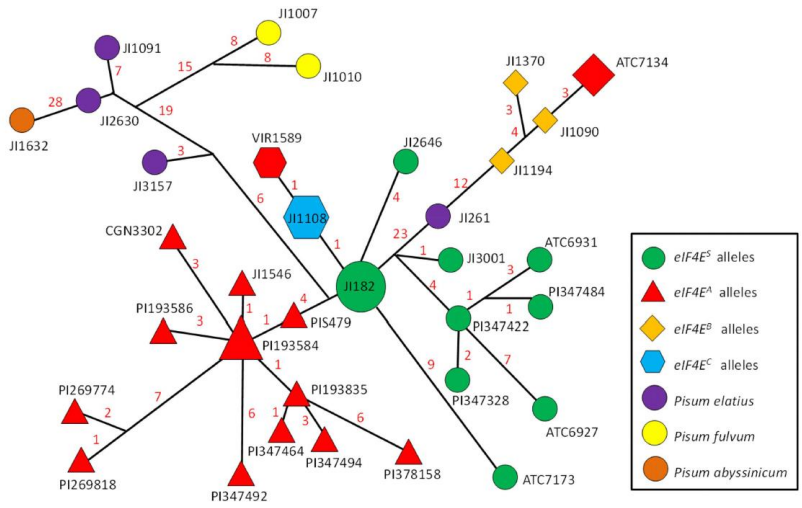


Figure 30

Haplotype network of 34 eIF4E haplotypes, using median-joining network algorithm, implemented in NETWORK, based on total of 156 SNP characters, excluding 50, 56 bp indels in intron 3 or 4, with indicated number of mutated positions (in red numbers above lines). In red colour are highlighted accessions which were resistant to P1 PSbMV. Size of symbols is proportional to number of accessions with given haplotype. taken from Konečná et al. 2014

4.2 Use of crop wild relatives – pea introgression lines

Plant evolution under domestication has led to increased productivity, but at the same time it has narrowed the genetic basis of the crop. As in other crops, also in pea, the demand for productivity and homogeneity has resulted in a limited number of standard, high-yielding varieties, at the price of the loss of heterogeneous but adaptable traditional local varieties (landraces), a process known as genetic erosion. It was shown that beside wild crop relatives, mainly landraces and older crop varieties preserve much of lost diversity and comprise the genetic resources for breeding new crop varieties to help cope with environmental and demographic changes (Esquinas-Alcazar 2005). This applies also to pea, as demonstrated on Czech pea varieties (**Cieslarova *et al.* 2011, 2012**). Genetic diversity is highly relevant for improvement of crop traits by breeding. The widespread use of genetically uniform varieties provides an ideal genetic environment for disease epidemics. Crop genetic resources are the reservois of often yet undiscovered allelic variants and thus provide an opportunity for genetic improvement of cultivated species (Warschefsky *et al.* 2014). In the past, a large number of agronomically important genes, including disease resistance genes, were introgressed from wild relatives and landraces into the cultivated species (Bullar *et al.* 2010). As the result of change to self-pollination (Dempewolf *et al.* 2012) fertility barriers between wild and cultivated populations facilitated fixation of the desired genotype (Zohary and Hopf, 2000,). Consequently, domestication bottleneck has resulted in high degree of relatedness between varieties, which was further pronounced in modern breeding programs, leading to narrower genetic base of cultivated germplasm, prone to pests and diseases (Harlan 1975; McCouch 2004; Zamir 2001). The study of genetic diversity captured in pea collections showed that although wide diversity is present among cultivated material (Ellis 2011; Jing *et al.* 2007, 1010, 2012; **Smýkal *et al.* 2011**), wild material provides much of the *Pisum* genus diversity, only partially captured by the domestication of pea (Ellis 2011; **Smýkal *et al.* 2011**). To reverse domestication bottleneck, genetic improvement of many crop plants has benefited from incorporation of traits from related wild species and other exotic germplasm sources (Zamir 2001; Gur and Zamir 2004; McCouch 2004). Highly variable germplasm is found in the secondary and tertiary pools of crop plants, including pea (**Smýkal *et al.* 2011, 2015**). This exotic material collected worldwide has largely remained uncharacterized and underutilized. Genetic improvement of many crop plants has benefited from incorporation of traits from related wild species and other exotic germplasm sources. The development of prebred lines has long been advocated as a means of facilitationg the transfer of genes from wild species. Vast pea germplasm collections (approximately 98 thousands accessions) are available but their use for crop improvement is limited as accessing genetic diversity is still a challenge (**Upadhyaya *et al.* 2011**; **Smýkal *et al.* 2013**). Unfortunately pea suffers largely from lack of international support, as compare to other grain legumes (**Smýkal *et al.* 2009, 2012**). Efficient extraction and exploitation of the adaptive

variation and valuable traits maintained in gene banks has yet to be fully achieved, though it remains a high priority of gene bank managers (**Upadhyaya *et al.* 2011**). Only small part of the enormous potential has been exploited in breeding of biotic and abiotic stresses or novel agronomical traits.

Plant breeders have tried to use interspecific crosses also in the *Leguminosae* to increase the size and diversity of the gene pool. Wide intergeneric legume hybrids have been critically reviewed in McComb (1975), with conclusion of insufficient evidence for most of the reported crosses and very often misplaced generic boundaries (**Smýkal *et al.* 2015**). Ochatt *et al.* (2004) confirmed the strong cross-incompatibility existing between *P. sativum* and *Lathyrus sativus* as first described by Campbell (1997), while successful although low fertility hybrids were obtained between *P. sativum* and *P. fulvum* (Errico *et al.* 1996, De Martino *et al.* 2000). Wild *Cajanus* species (*C. scarabaeoides, C. cajanifoloius* and *C. acutifolius*) have also been exploited to develop cytoplasmic male sterility (Saxena *et al.* 2005; Mallikarjuna and Saxena 2005), which have been used to develop commercial hybrids (Saxena *et al.* 2010). Resistance to legume pod borer (*Helicoverpa armigera*) in pigeonpea (*C. cajan*) has been introgressed from *C. acutifolius* and *C. scarabaeoides* (Mallikarjuna *et al.* 2007). Wild relative of chickpea, *Cicer reticulaum* have been used to reduce days to flowering, maturity, increase seed weight, seed yield and harvest index in cultivated chickpea (Upadhyaya *et al.* 2008). However, the transfer of genes from wild species, is often accompanied by inevitable genetic drag of undesirable wild material related traits and this has prevented its broader use (Zamir 2001). To avoid this, the synthesis of exotic libraries, such as introgression lines (ILs), near isogenic lines (NILs) and chromosome substitution lines (CSSLs), containing molecularly defined chromosome segments from wild species in a constant genetic background of the cultivated species has been applied to make the use of alien genomes more precise and efficient (Tanksley and McCouch 1997; Gur and Zamir 2004; Zamir 2001; McCouch 2004). Introgression lines are effectively used in discovering hidden genetic variation, identifying favourable genes, evaluating the action or interaction of QTLs in multiple environments and providing favourable experimental materials for plant breeding and genetics research. Set of ILs represents the genome of a donor parent through single lines each carrying one or few introgressed donor segments in the same genetic background of the recurrent parent. This is achieved by several rounds of backcrossing to the recurrent parent followed by marker-assisted selection (Zamir 2001). Use of such defined permanent libraries provide powerful tool for the identification of novel genes (Eshed *et al.* 1996). Owing to our current knowledge of two model legumes genomes, reinforced with prooved existence of syntheny and colinearity, together with large set of mapped markers, we have in hands sufficient molecular tools to reach required level of resolution in order to assess recombination between wild and cultivated parents.

Unlike cereals, the progress in developing of introgression lines in legumes lag behind, perhaps due to the greater difficulty in generating interspecific crosses and lack of DNA marker technology (both markers and high throughput assay) to monitor genomic coverage in progenies. With knowledge of two model legumes genomes, reinforced with prooved existence of syntheny and colinearity (Kalo *et al.* 2004), together with large set of pea specific mapped markers (Loridon *et al.* 2005; Aubert *et al.* 2006; Deulvot *et al.* 2010; Bordat *et al.* 2011) we have in hands sufficient molecular tools to reach required level of resolution in order to assess recombination between wild and cultivated parents.

In genus *Pisum*, *P. sativum* subsp. *elatius/humile* and *P. abyssinicum* are within primary gene pool, *P. fulvum* in secondary gene pool and *Vavilovia formosa* in tertiary gene pool, all being diploid with 2n = 14 and cross compatible (**Smýkal *et al.* 2015**). Recent molecular studies has confirmed this and showed the largest phylogenetic distance of *P. fulvum* (beside Vavilovia as sister species) to cultivated pea (**Jing *et al.* 2010**; Ellis 2011; **Smýkal *et al.* 2011**; Schaefer *et al.* 2012). Pioneering work of Ben-Ze´ev and Zohary (1973) on crosses among different *Pisum* species and subspecies, has not only contributed to taxonomy but also can be considered as first attempt for wider hybridization. It indicated some hybridization barriers, and as the result semi-fertility of F1 hybrids together with a reduction in chiasmata formation. The difference of *P. fulvum* karyotype from the other taxa (Ben-Ze´ev and Zohary, 1973) with observed distorted segregation and low pollen fertility of hybrids (De Martino *et al.* 2000) likely due to reciprocal translocations (Errico *et al.* 1991). After the resistance to pea weevil was identified in *P. fulvum* (Hardie *et al.* 1995), with a pod and seed resistance mechanism being implicated (Clement, Hardie and Elberson, 2002), it was attempted to introduce it into cultivated pea. Crosses were used to transfer the powdery mildew (Fondevilla et al. 2007) and bruchid (Byrne *et al.* 2008, Aryamanesh *et al.* 2014) resistances from *Pisum fulvum* into cultivated pea as well as incorporation of PSbMV and Fusarium resistances from primitive landraces (McPhee *et al.* 1999; Provvidenti 1990). The value of wild crop relatives has been illustrated by novel *Er3* gene, conferring dominant resistance to *E. pisi*, identified in *Pisum fulvum* (Fondevilla *et al.* 2008). Crosses were used to transfer the powdery mildew (Fondevilla *et al.* 2007), *Mycosphaerella pinodes* and *Orobanche crenata* (Rubiales *et al.* 2005; Fondevilla *et al.* 2005) and bruchids (Clement *et al.* 2009; Byrne *et al.* 2008) resistances from *Pisum fulvum* into cultivated pea as well as incorporation of PSbMV virus and *Fusarium* resistances from primitive *P. sativum* landraces (Provvidenti 1990; Provvidenti and Alconero, 1988; McPhee *et al.* 1999). As shown by Byrne (2005) two backcrosses were sufficient to restore much of the seed and plant architecture (pod, branching, flowering time) characters, while maintaining desired introgressed trait. However all mentioned results with wild material were achieved with dedicated crosses and specific selection, thus they need to be made in trait-by-trait manner, a time and money consuming process. Establishment of permanent population introgression library with characterized genomic fragments of wild species in

defined genetic background allows phenotypic characterization of unlimited number of target traits (Eshed *et al.* 1996; Chetelat, Meglic and Cisneros 2000), which coupled together with on going advance of available molecular tools provide mean for final gene identification and subsequent incorporation, pyramiding in desired genotypes ultimately leading to better performing commercial pea varieties.

We report the development of introgression lines containing chromosome segments of wild pea (*Pisum fulvum*) genome in cultivated pea (*Pisum sativum*) genetic background defined by molecular markers. Seventy two F_2 plants (originating from 6 F_1 individuals) of cross between WL1238 (*P. sativum)* x WL2140 (*Pisum fulvum*) and 37 F_2 plants (originating from 4 F_1 individuals) of reciprocal cross WL2140 x WL1238 (**Figure 31**) were used and as recurrent parent modern afila type dry-seed pea variety Terno (05L0100989, http://genbank.vurv.cz/genetic/resources) was used for three backcrosses (BC_3). After this, single-seed descent method through four self-pollinations (BC_3S_5) was performed. No intentional selection was applied during this process, except of plant fertility. The WL1238 (= JI73) *P. sativum* parent is a tester line with several mapped morphological markers. *P. fulvum* parent (WL2140 = JI224=PI560061) originates from Israel, valley of Cross (Kosterin and Bogdanova 2008). Variety Terno (pedigree: LU-134 x Rustic) is modern Czech origin variety, yellow seeded, diamond shape seeds, semi-leafless (*afila*) type, 102 cm long stem, intermediate growing period (111 days), thousand seeds weight 314 g, intermediate tolerance to complex of root diseases, ascochyta blight, *Mycosphaerella pinodes, Botrytis cinerea* and downy mildew.

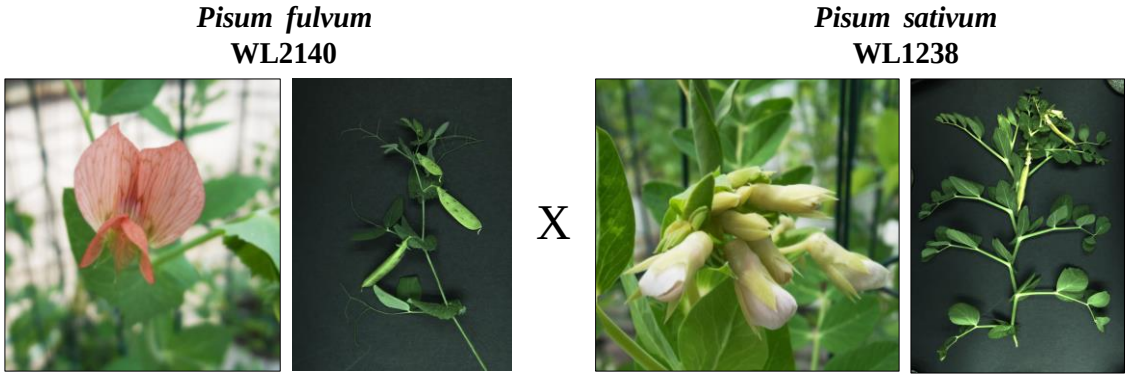


Figure 31 *Photos of two parental lines used to create introgression lines.*

To monitor breeding success and recombination process, based on the available genetic maps we selected easily scorable and locus specific markers, such as microsatellites and gene-based (**Figure 32**). Thirty nine microsatellite markers were tested for length polymorphism among parents, of these, 28 showed polymorphism scorable on agarose gel. Of thirty three gene-specific mapped from Konovalov (2006), twenty showed CAPS polymorphism. Further twenty one were successfully amplified from 39 markers selected from Aubert *et al.* (2006), 12 of Brauner *et al.* (2002) and further 15 were polymorphic

from Deulvot *et al.* (2010). Selected polymorphic 28 microsatellite and 44 gene-specific markers covered all seven linkage groups of pea 4.45 Gbp genome at 2 to 82 cM spacing, with mean of 15.4 cM. The resolution of genotyping was given by number of markers per linkage group (LG). There were 6 gene-specific per each of the linkage group except of LGIII with 8 markers and 4 microsatellite markers per LGI, II, III, IV, V, VI and VII, respectively. All together these markers delimited 78 chromosomal regions.

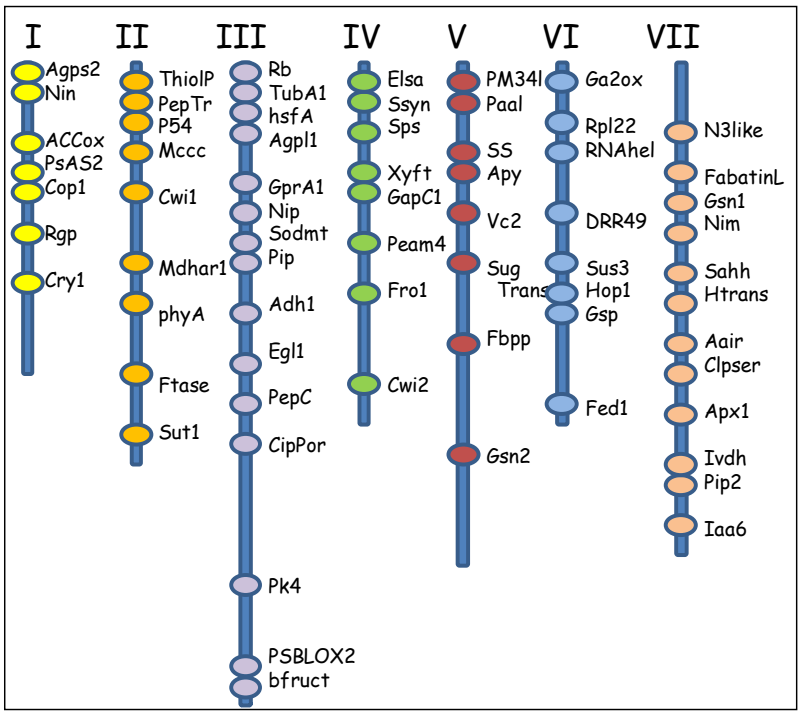


Figure 32

Schematic positions of gene-based markers per linkage groups used for initial genotyping of introgression lines.

Theoretical values of reccurrent parent genotype after two backcrosses is 87.5%, while donor genotype of wild *Pisum fulvum* is 12.5%. The observed numbers at BC_2S_3 generation as detected by molecular markers, showed heterozygozity in 533 (8%) cases while recurrent *Pisum sativum* parent in 4552 (69%) and introgressed segments of *P. fulvum* in 1551 (23%) of 30 cM in average (**Figure 33**). The differences between expected and observed values are likely due to selection of lines used for backcrosses, based on plant vigour and fertility. Also the selection of seeds as well as relatively lower number of propagated lines have contributed to this.

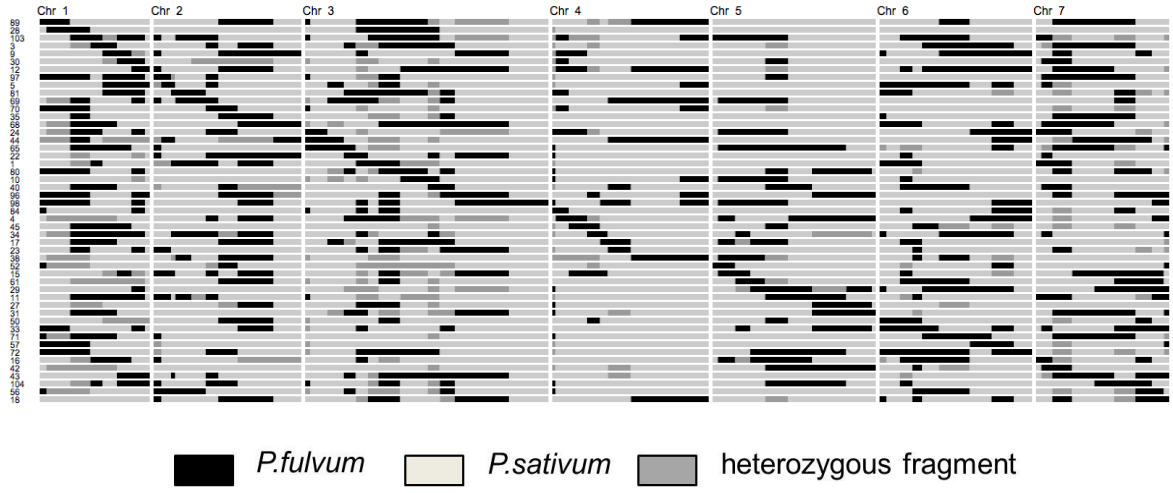


Figure 33 *CSSL Finder has selected 49 lines (of 105 analyzed) as representative subset.*

Taking in account the detection limit achieved with 64 markers for 7 LG and spacing total length of 1389 cM of *Pisum sativum* genome with average spacing 15 cM, the average size of introgressed *Pisum fulvum* fragments is 40 cM. The were 1.2, 1.3, 2.3, 1.3, 0.8, 1.1 and 1.5 segments per linkage group. This allowed estimates of *Pisum fulvum* introgressed genome proportion, between 20 – 70% of given linkage group, with 33% in average. This is higher than theoretical value of 12.5%, likely due to selection and number of analysed lines. There were 4, 16, 0, 13, 34, 21 and 7 lines per respective linkage group without introgressed *Pisum fulvum* chromosomal segment. Currently (spring-summer selected 50 lines from BC_3F_5 generation are under field trials (**Figure 34**) and in parallel being genotyped using 13.2k Pea SNP Illumina assay (Tayeh *et al.* 2015) and DARTseq genome wide approach. In addition, another introgression lines are being made using *P. elatius* (L100 line) x cv. Cameor (**Smýkal *et al.* 2014**).

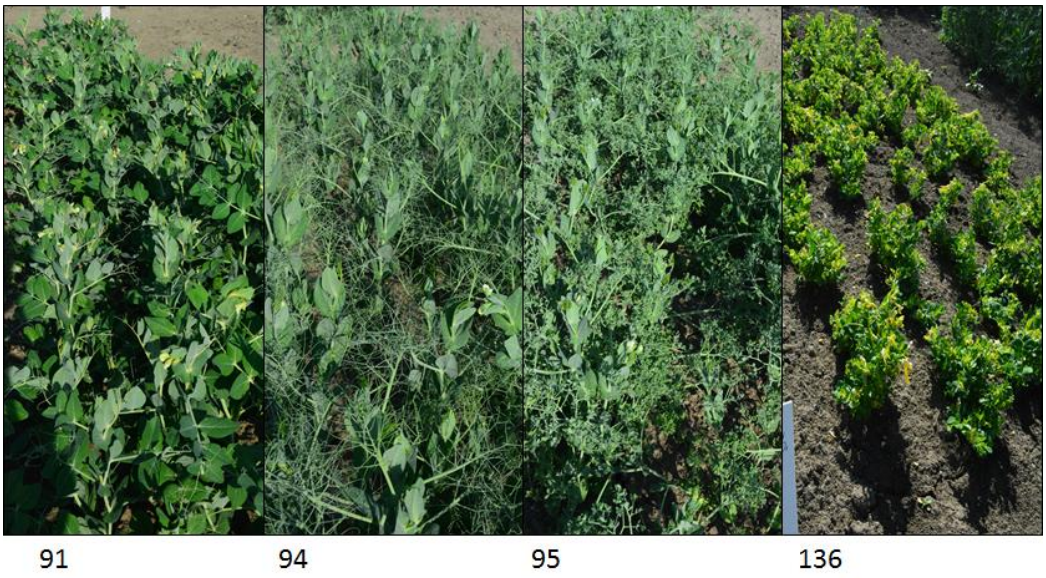


Figure 34: *Photos of 4 selected introgression lines from 2016 field trials*

4.3 Pea crop and breeding perspectives

Commenting on ***Smýkal P,*** *Aubert G, Burstin J et al. (2012) Pea (Pisum sativum L.) in the Genomic Era. Review. Agronomy 2: 74- 115.*

There are several records of garden peas in the writing of the old Greeks and Romans, as well as in the herbal references of several centuries ago. There is discussion on cultivation of pea in ancient India and Egypt (De Candolle 1882), indicated by both linguistic and archaeological evidence. Theofrastus of Greece (died 287 BC) records the use of *orobos* for the vetch, *erebinthos*, for the chickpea and *pisos* for the pea. Subsequently the transfer of Greeks *pisos* to Rome, become *Pisum*, a name passed to the English as peason, then pease or peasse, which after the drop of s became the universal name among English-speaking people (Mikič 2012). This interesting paleolinguistics study shows roots directly related to traditional Eurasian pulse crops. We are not certain when pea cultivation was taken up by Romans, as neither Cato (149 BC) nor Varro (27BC) name *pisum*, but use more general terms such as pulses or legumes, which are known to include lentils and chickpea (Cubero, Perez de la Varga, & Fratini, 2009). In the first century bc pea was mentioned by the Romans Collumela, Pliny and Virgil.

Dry pea currently ranks third after common bean (*Phaseolus vulgaris*) and chickpea (*Cicer arietinum*) as the most widely grown grain legume in the world with primary production in temperate regions and global production of 11.4 M tonnes plus additional 17.2 M tonnes of green vegetable pea in 2014 (FAOSTAT 2015). Pea is important temperate region pulse, with feed, fodder and vegetable uses. Pea seeds are rich in protein (23-25%), slowly digestible starch (50%), soluble sugars (5%), fibre, minerals and vitamins (Bastianelli *et al.* 1998) as well as in secondary metabolites such as isoflavonoids with anticancer and other health-promoting activities. On a worldwide basis, legumes contribute about one-third of humankind's direct protein intake, while also serving as an important source of fodder and forage for animals and of edible and industrial oils. Dry peas are grown in temperate zones and FAOSTAT registered 94 pea growing countries during the period from 2000 – 2015 and cultivated area of dry pea ranged from 6 to 6.5 million hectares. Dry pea production in Europe declined while increased production was recorded in Canada, USA, China and Russian Federation. The reasons for these changes include economic, biological, physical, sociological and technical factors. Canada accounts for close to one-third of world pea production and well over half of world pea exports. Countries with production area greater than 100,000 hectares, and yield less than 1000 kilograms per hectare included Pakistan and Ethiopia. The highest yields of 4000 – 5000 kilograms per hectare were traditionally achieved in Europe (Netherlands, France, Belgium). The worldwide average yield was

about 1700 kilograms per hectare and yields less than 500 kilograms per hectare were recorded in parts of Africa (**Smýkal *et al.* 2012**). Unfortunately, improvement in legume crop yields have not kept pace with those of cereals. In part, this difference is due to the unfavorable environmental conditions under which many legume species are grown. Legumes are often grown after corn or rice and are seeded toward the end of the growing season. They may have short growing seasons and may be subject to intermittent or terminal drought. Global annual field pea production has been relatively steady over the past 50 years, however, the key centres of production have shifted quite dramatically during that period (FAOSTAT, 2015). In the 1960s to 1980s Eastern Europe (primarily Russia and Ukraine) was the key production region. In the 1980s and 1990s Western European (primarily France) production increased substantially. In the 1990s and 2000s, production from the Americas (primarily Canada) became dominant, with some production from Australia as well. Production in Asia and Africa has been relatively steady at moderate and low levels, respectively. In North America the key field pea breeding activities are conducted at four public institutions, i.e., University of Saskatchewan, Saskatoon, SK, Agriculture and Agri-Food Canada, Lacombe, AB, USDA-ARS, Pullman, WA, and North Dakota State University, Fargo, ND. The key breeding objectives involve increasing yield potential by improving biotic and abiotic stress resistances, and enhancing quality for diverse food markets. Quality includes improved appearance of the seeds as well as improved nutritional value, cooking quality and flavour. Field pea breeding, and plant breeding in general, is primarily conducted by private companies in Europe, with public institutions conducting supportive basic and applied research. In the 1980s, approximately 20 small to medium sized companies were involved in field pea breeding in Europe. This has dropped to less than 10 at present, in concert with the gradual decline in pea production in Europe over the past 15 years, being replaced by winter wheat and winter canola production. In addition, public funding supporting pea research, and research on grain legumes in general, has declined in western Europe over the past decade. In Australia, a single national field pea breeding program is located at Horsham, Victoria with extensive evaluation conducted by collaborators in each state. Vegetable pea breeding is conducted by a few private companies primarily based in the Pacific Northwest of USA, western Europe, China and India. The CGIAR system has generally ignored pea improvement over the years. ICRISAT breeds chickpea and pigeon pea, ICARDA breeds lentil, chickpea and fababean, while IITA breeds cowpea.

Future prospects for field pea production globally depend on several factors including: the ability of breeders to produce high yielding cultivars which are competitive in crop rotations with the dominant cereal and oilseed alternatives, the ability of agronomists to develop effective, sustainable crop production strategies, and the ability of the global pulse industry to market pea as a highly nutritious affordable food with diverse applications. Considering this situation and the continuing expansion of

global population and food needs, the time has come for CGIAR to re-invest in pea improvement, as it is the most affordable grain legume crop in the world (**Warkentin *et al.* 2015**). The key breeding objectives involve increasing yield potential by improving biotic and abiotic stress resistances, and enhancing quality for diverse food markets. Quality includes improved appearance of the seeds as well as improved nutritional value, cooking quality and flavour. The extension of model legumes for comparative functional genomics, together with 'omics' knowledge, is starting to provide candidate genes for QTL identification of genes involved in stress and quality traits. As genes are identified in model legumes and crop species comparison and transfer of candidate gene information from the model to the crop species is possible, favorable alleles for breeding and selection will be identified, and improved varieties will be developed by marker assisted selection (MAS) or genetic transformation. Future prospects for field pea production globally depend on several factors including: the ability of breeders to produce high yielding cultivars which are competitive in crop rotations with the dominant cereal and oilseed alternatives, the ability of agronomists to develop effective, sustainable crop production strategies, and the ability of the global pulse industry to market pea as a highly nutritious affordable food with diverse applications. Pea production will be challenged by climate change this century (Coyne *et al.* 2011). Rising temperatures pose the greatest threat to production of cool season pea as the traditionally temperate regions shift northward. For pea, failed seed-set from heat-stress causes the greatest damage to seed yields. This can be expected to exacerbate climate unpredictability and to result in unprecedented levels of heat and drought stress during the reproductive phase in agricultural areas of the temperate – sub-tropical zones worldwide, especially in the sub-Sahara and north central India (Coyne *et al.* 2011). Breeding aims to improve agronomically important traits by combining characters present in different parental lines of cultivars, species or their wide relatives. In conventional pea breeding programs, various crossing strategies are employed to incorporate desirable traits from one accession into another, more adapted background, including backcrossing, single seed descent and recurrent selection. This is a time-consuming and costly process, which may be speed up through the application of molecular markers used to determine the number, position and individual effects of loci associated with the trait of interest (reviewed in **Smýkal and Konečná 2014**). Markers also offer potential to advance pea breeding through accurate identity, pedigree, purity and hybrid determination, and analysis of genetic variation. The pace of appearance of new genomics technologies has tremendously increased during the last decade and unforeseen strategies for crop breeding have emerged. As in any other crop, there is long list of diseases and pests affecting pea. Among them fungal and viral pathogens are likely causing the most severe damage. The genetic basis of these diseases were approached by molecular tools and several genomic regions or even casusative genes were identified (reviewed in **Smýkal and Konečná 2014**). QTL mapping studies in pea are lagging behing other

economically important crops. These classical biparental mapping population approach was recently accompanied by association mapping, at best on whole genome level (Zhu *et al.* 2008) applied also in pea (Kwon *et al.* 2012; Cheng *et al.* 2015; Desgroux *et al.* 2016). With the availability of high density gene-based map (Tayeh *et al.* 2015a) and forecasted pea genome sequence these are anticipated to be applied for trait genomic prediction (Burstin *et al.* 2015; Tayeh *et al.* 2015b) as already seen in other economically more important crops.

Conclusions and future prospects

Pea belongs to the oldest domesticated plants, it has been extensively used in early hybridization studies and it was the model organism of choice for Mendel's discovery of the laws of inheritance, making pea part of the foundation of modern genetics. However its large genome rich in repetitive sequences has precluded its further use in genetics. Pea is important legume crop worldwide with rich genetic diversity preserved in germplasm collections. *Pisum* genus despite consisting of only two to three species, has interesting phylogenetic relationship to *Lathyrus* and *Vicia* genera. Although pea diversity has been molecularly and morphologically assessed and several core collections were established to facilitate its use, there is still gap in use of pea wild crop relatives. The example given with gene governing resistance to one virus exemplifies one of the possible germplasm exploration.

Moreover, progress in molecular tools and genomic resources offer the possibility of systematic use of diversity preserved in wild forms, species to be incorporated into modern varieties to provide genes, alleles which have not been used during domestication process in form of introgression lines. Our work on whole genome sequences provides an opportunity not only to clarify generic boundaries and intraspecific relationships, but coupled with environmental data extracted from collection sites offers great dataset to study species distribution in context of its adaptation to habitats. Wild pea provides an usefull model to study legume crop domestication process, especially two key traits: seed dormancy and pod dehiscence. With anticipated release of pea genome and adoption of precise high throughput phenotypic evaluation methods it will provide opportunity to link phenotype with genotype on a whole genome basis, a process initiated by Mendel about 150 years ago.

List of selected published papers used for the thesis

1. **Smýkal P,** Coyne C and 14 co-authors (2015) Legume crops phylogeny and genetic diversity for science and breeding. *Critical Reviews Plant Sciences* 34: 43-104.
2. **Smýkal P**, Vernoud V, Blair MW, Soukup A and Thompson RD (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Frontiers in Plant Sciences* 5: 351.
3. Mikič A, **Smýkal P** and 12 co-authors (2014) Beauty will save the world, but will the world save beauty? The case of the highly endangered *Vavilovia formosa* (Stev.) Fed. *Planta* 240: 1139-1146.
4. **Smýkal P** and 8 co-authors (2014) A comparative study of ancient DNA isolated from charred pea (*Pisum sativum* L.) seeds from an Early Iron Age settlement in southeast Serbia: inference for pea domestication. *Genet Resour Crop Evol.* 61: 1533–1544.
5. **Smýkal P** (2014) Pea (*Pisum sativum* L.) in biology prior and after Mendel's discovery. *Czech. J. Genet. Plant Breed.* 50: 52–64.
6. Konečná E, Šafářová D, Navrátil M, Hanáček P, Coyne C, Flavell A, Vishnyakova M, Ambrose M, Redden R and **Smýkal P** (2014) Geographical gradient of the eIF4E alleles conferring resistance to potyviruses in pea (*Pisum*) germplasm. *PLOS One* 9, 3: e90394.
7. Mikic A, **Smykal P,** Kenicer G *et al.* (2013) The bicentenary of the research on 'beautiful' vavilovia (*Vavilovia formosa*), a legume crop wild relative with taxonomic and agronomic potential. *Botanical Journal of the Linnean Society* 172: 524-531.
8. De Beukelaer H, **Smykal P**, Davenport G, Fack V (2012) Core Hunter II: fast core subset selection based on multiple genetic diversity measures using Mixed Replica search. *BMC Bioinformatics* 13: 312.
9. **Smýkal P**, Aubert G, Burstin J *et al.* (2012) Pea (*Pisum sativum* L.) in the Genomic Era. Review. *Agronomy* 2: 74- 115.
10. Jing R, Ambrose MA, Knox MR, **Smykal P** and 12 co-authors (2012) Genetic Diversity in European Pisum Germplasm Collections. *Theoretical and Applied Genetics* 125: 367-380.
11. Cieslarová J, Hýbl M, Griga M, Fialová E, **Smýkal P** (2012) Molecular Analysis of Temporal Genetic Structuring in Pea (*Pisum sativum* L.) Cultivars Bred in the Czech Republic and in Former Czechoslovakia Since the Mid-20th Century. *Czech J. Genet. Plant Breed.* 48: 61-73.
12. Cieslarová J, Hanáček P, Fialová E, Hýbl M, **Smýkal P** (2011) Estimation of pea (*Pisum sativum* L.) microsatellite mutation rate based on pedigree and single-seed descent analyses. *J Appl Genetics* 52: 391-401.

13. Upadhyaya HD, Dwivedi SL, Ambrose M, Ellis N, Berger J, **Smýkal P**, D. Debouck D, Duc G, Dumet D, Flavell A, Sharma SK, Mallikarjuna N, Gowda CL (2011) Legume genetic resources: management, diversity assessment, and utilization in crop improvement. *Euphytica* 180: 27-47.
14. **Smýkal P**, Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, Ford R, Coyne CJ, Maxted N, Ambrose MJ, Ellis THN (2011) Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. *Plant Genetic Resources* 9: 4-18.
15. Cieslarová J, Smýkal P, Dočkalová Z, Hanáček P, Procházka S, Hýbl M, Griga M (2011) Molecular evidence of genetic diversity changes in pea (*Pisum sativum* L.) germplasm after long-term maintenance. *Genet Resour Crop Evol.* 58: 439-451.
16. Akopian J, Sarukhanyan N, Gabrielyan I, Vanyan A, Mikić A, **Smýkal P,** Kenicer G, *et al.* (2010) Reports on establishing an ex situ site for 'beautiful' vavilovia (*Vavilovia formosa*) in Armenia. *Genet Resour Crop Evol.* 57: 1127–1134.
17. Jing R, Vershinin A, Grzebyta J, Shaw P, **Smýkal P**, Marshall D, Ambrose MJ, Ellis THN, Flavell AJ (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evolutionary Biology* 10: 44.
18. Mikič A, **Smýkal P**, Kenicer G, Sarukhanyan N, Akopian J, Gabrielyan I, Vanyan A, *et al.* (2010) Achievements in Research on Vavilovia (*Vavilovia formosa* (Stev.) Fed.), a Legume Crop Wild Relative. *Field Veg. Crop Res.* 47: 387-394.
19. **Smýkal P**, Šafářová D, Navrátil M, Dostalová R (2010) Marker assisted pea breeding: eIF4E allele specific markers to pea seed-borne mosaic virus (PSbMV) resistance. *Molecular Breeding* 26: 425-438.
20. Mikič A, **Smýkal P,** Kenicer G, Vishnyakova M, Sarukhanyan N, Akopian J, Gabrielyan I, Vanyan A, Čupina B, Mihailovič V, Ambrose M, Ellis N (2009) A revival of research on beautiful vavilovia (*Vavilovia formosa* syn. *Pisum formosanum*). *Pisum Genetics* 41: 11-14.
21. **Smýkal P**, Hýbl M, Corander J, Jarkovský J, Flavell A, Griga M (2008) Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. *Theoretical and Applied Genetics* 117: 413-424.
22. **Smýkal P**, Coyne CJ, Ford R, Redden R, Flavell AJ, Hýbl M, Warkentin T, Burstin J, Duc G, Ambrose M, Ellis THN (2008) Effort towards a world pea (*Pisum sativum* L.) germplasm core collection: The case for common markers and data compatibility. *Pisum Genetics* 40: 11-14.

References

Abbo S, Lev-Yadun S, Gopher A (2010) Agricultural origins: centers and noncenters, a Near Eastern reappraisal. *Critical Reviews in Plant Science* 29: 317–328.

Abbo S, Lev-Yadun S, Heun M, Gopher A (2013) On the 'lost' crops of the neolithic Near East. *Journal of Experimental Botany* 64: 815–822.

Abbo S, Pinhasi van-Oss R, Gopher A, Saranga Y, Ofner I, Peleg Z (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in Plant Science* 19: 351-60.

Abbo S, Zezak I, Schwartz E, Lev-Yadun S, Gopher A (2008) Experimental harvesting of wild peas in Israel: implications for the origins of Near East farming. *Journal of Archaeological Science* 35: 922–929.

Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J (2008) Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology* 17: 5177-5188.

Alsos IG, Ehrich D, Thuiller W *et al.* (2012) Genetic consequences of climate change for northern plants. *Proceedings Royal Society of London B: Biological Sciences* 279: 2042-2051.

Alves-Carvalho S, Aubert G, Carrère S *et al.* (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* 4: 1-19.

Ambrose MJ (1995) From Near East centre of origin the prized pea migrates throughout world. *Diversity* 11: 118–119.

Ambrose MJ, Ellis THN (2008) Ballistic seed dispersal and associated seed shadow in wild *Pisum* germplasm. *Pisum Genetics* 40: 5–10.

Appelhagen I, Thiedig K, Nordholt N, Schmidt N, Huep G, Sagasser M, Weisshaar B (2014) Update on transparent testa mutants from Arabidopsis thaliana: characterisation of new alleles from an isogenic collection. *Planta* 240: 955-70.

Arumuganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9: 208–218.

Aryamanesh N, Zeng Y, Byrne O *et al.* (2014) Identification of genome regions controlling cotyledon, pod wall/seed coat and pod wall resistance to pea weevil through QTL mapping. *Theoretical and Applied Genetics* 127: 489-497.

Ascheron P and Graebner P (1910) Synopsis der mitteleuropaischen Flora Bd 6, Abt 2, IV Leipzig 1093 S

Atif RM, Patat-Ochatt EM, Svabova L, Ondrej V, Klenoticova H, Jacas L, Griga M, Ochatt S (2013) Gene transfer in legumes. Luttge U et al (eds) Progress in botany, vol 74. Springer, Berlin

Aubert G, Morin J, Jacquin F *et al.* (2009) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume. *Medicago truncatula. Theoretical and Applied Genetics* 112: 1024–1041.

Axelrod DI (1975) Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. *Annals of the Missouri Botanical Garden* 62: 280–334.

Baldev B (1988) Origin, distribution, taxonomy, and morphology. In: *Pulse crops.* pp. 3–51. B. Baldev S Ramanujam, and Jain, HK, Eds, New Delhi, India: Oxford and IBH Publishing Co.

Baranger AG, Aubert G, Arnau G et al. (2004) Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. *Theoretical and Applied Genetics* 108: 1309–1321.

Barba-Espín G, Diaz-Vivancos P, Job D, Belghazi M, Job C, Hernández JA (2011) Understanding the role of H2O2 during pea seed germination: a combined proteomics and hormone profiling approach. *Plant, Cell & Environment* 34: 1907-1919.

Barba-Espín G, Diaz-Vivancos P, Job D, Belghazi M, Job C, Hernández JA (2011) Understanding the role of H_2O_2 during pea seed germination: a combined

proteomics and hormone profiling approach. *Plant Cell and Environment* 34: 1907–1919.

Barulina H (1930) Lentils of the USSR and other countries. *Bulletin of Applied Genetics and Plant Breeding* 40: 1–319. [In Russian]

Barrow MV, Simpson CF, Miller EJ (1974) Lathyrism: a review. *The Quarterly Review of Biology* 49: 101–128.

Baskin CC (2003) Breaking physical dormancy in seeds– focussing on the lens. *New Phytologist* 158: 229–232.

Baskin JM, and Baskin CC(2000) Evolutionary considerations of claims for physical dormancy-break by microbial action and abrasion by soil particles. *Seed Science Research* 10: 409–413.

Bässler M (1966) Die Stellung des Subgenus *Orobus* (L.) Baker in der Gattung *Lathyrus* L. und seine systematische Gliederung. *Feddes Repertorium* 72: 69-97.

Bässler M (1973) Revision der eurasiatishen arten von *Lathyrus* L. Sect. *Orobus* (L.) Gren. et. Godr. *Feddes Repertorium* 84: 329-447.

Bastianelli D, Grosjean F, Peyronnet C, Duparque M, Regnier JM (1998) Feeding value of pea (Pisum sativum,L.). 1. Chemical composition of different categories of pea. *Animal Science* 67: 609-619.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews in Genetics* 5: 251-261.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews in Genetics* 5: 251-61.

Bedada G, Westerbergh A, Nevo E, Korol A, Schmid KJ (2014) DNA sequence variation of wild barley *Hordeum spontaneum* (L.) across environmental gradients in Israel. *Heredity* 112: 646–655.

Bellucci E, Bitocchi E, Rau D *et al.* (2014) Genomics of Origin, Domestication and Evolution of *Phaseolus vulgaris*. In: Tuberosa R, Graner A, Frison E eds. *Genomics of Plant Genetic Resources*. New York: Springer, 483-507.

Benedito VA, Torres-Jerez I, Murray JD *et al.* (2008) A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal* 55: 504-13.

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

Berger A (1928) Systematic botany of peas and their allies. In. Hedrick U (ed.) *The vegetables of New York*. Albany: 1: 10-18.

Bewley JD (1997) Seed Germination and Dormancy. *Plant Cell* 9: 1055–1066.

Bhullar NK, Zhang Z, Wicker T, Keller B (2010) Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene Pm3: a large scale allele mining project. *BMC Plant Biology* 10: 88.

Bieberstein M (1808) Flora taurico-caucasica exhibens stirpes phaenomagas, in Chersoneso Taurica et regionibus caucasicis sponte crescentes. Bd 2 Charkouiae, Typ Akad 477 S

Binladen J, Wiuf C, Gilbert TP, Bunce M, Barnett R, Larson G, Greenwood AD, Haile J, Ho SYW, Hansen AJ, Willerslev E (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* 172: 733–741.

Bisby FA, Buckingham J, Harborne JB (eds) 1994. *Phytochemical dictionary of the Leguminosae* Vol. 1: *plants and their constituents*. London: Chapman and Hall.

Bitocchi E, Bellucci E, Giardini *et al.* (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist* 197: 300-313.

Blixt S (1972) Mutation genetics in *Pisum*. *Agri Hort Genet* 30: 1-293.

Bogdanova V, Kosterin O, Yadrikhinskiy A (2014) Wild peas vary in their cross-compatibility with cultivated pea (*Pisum sativum subsp. sativum* L.) depending on alleles of a nuclear–cytoplasmic incompatibility locus. *Theoretical and Applied Genetics* 5: 1163-1172.

Bogdanova VS (2007) Inheritance of organelle DNA markers in a pea cross associated with nuclear-cytoplasmic incompatibility. *Theoretical and Applied Genetics* 114: 333-339.

Bogdanova VS, Galieva ER, Kosterin OE (2009) Genetic analysis of nuclear-cytoplasmic incompatibility in pea associated with cytoplasm of an accession of wild subspecies *Pisum sativum* subsp. elatius (Bieb.) Schmahl. *Theoretical and Applied Genetics*: 118: 801-809.

Bogdanova VS, Zaytseva OO, Mglinets AV, Shatskaya NV, Kosterin OE, Vasiliev GV (2015) Nucleic-cytoplasmic conflict in pea (*Pisum sativum* L.) is associated with nuclear and plastidic genes encoding Acetyl-CoA carboxylase subunits. *PLoS One* 10: e0119835

BogdanovaVS and Kosterin OE (2007) Hybridization barrier between *Pisum fulvum* Sibth.et Smith and *P. sativum* L. is partly due to nuclear-chloroplast incompatibility. *Pisum Genetics* 39: 8-9.

Bohra A, Pandey MK, Jha UC *et al.* (2014) Genomics-assisted breeding in four major pulse crops of developing countries: present status and prospects. *Theoretical and Applied Genetics* 127: 1263–1291.

Boissier E (1856) Diagnoses plantarum orientalum novarum. Lipsie, ser 2, v 3, pp 125.

Bordat A, Savois V, Nicolas M *et al.* (2011) Translational genomics in legumes allowed placing *in silico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *G3. Genes-Genomes-Genetics* 1: 93–103.

Borojevic K (2006)Terra and Silva in the Pannonian plain. Opovo agro-gathering in the late neolithic. Archaeopress, Oxford

Brauner S, Murphy L, Walling JG, Przyborowski J, Weeden NF (2002) STS markers for comparative mapping in legumes. *Journal of the American Society for Horticultural Science* 127: 616–622.

Breese EL (1989) Regeneration and multiplication of germplasm resources in seed genebanks: the scientific background. International Board for Plant Genetic Resources, Rome.

Brown AHD and Spillane C (1999) Implementing core collections-principles, procedures, progress, problems and promise. In: Johnson RCand Hodgkin T (eds.) *Core collections for today and tomorrow.* Rome, International Plant Genetic Resources Institute, Italy, pp. 1-9

Burstin J, Deniot G, Potier J, Weinachter C, Aubert G and Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. *Plant Breeding* 120: 311–317.

Burstin J, Salloignon P, Martinello M *et al.* (2015). Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics* 16: 105.

Byrne OMT (2005) Incorporation of pea weevil resistance for wild pea (*Pisum fulvum*) into cultivated pea (*Pisum sativum*). PhD Thesis, UWA

Byrne,OM, Hardie DC, Khan T, Yan, G (2008) Genetic analysis of pod and seed resistance to pea weevil in a *Pisum sativum x P. fulvum* interspecific cross. *Australian Journal of Agriculture Research* 59: 854-862.

Campbell CG (1997) Grass pea: *Lathyrus sativus* L. International Plant Genetic Resources Institute, Rome; Italy, 92p

Cieslarová J, Hýbl M, Griga M, Fialová E, Smýkal P (2012) Molecular analysis of temporal genetic structuring in pea (*Pisum sativum* L.) cultivars bred in the Czech Republic and in former Czechoslovakia since the mid-20th Century. *Czech Journal of Genetics and Plant Breeding* 48: 61-73.

Cieslarová J, Smýkal P, Dočkalová Z, Hanáček P, Procházka S, Hýbl M, Griga M (2011) Molecular evidence of genetic diversity changes in pea (*Pisum sativum* L.) germplasm after long-term maintenance. *Genetic Resources and Crop Evolution* 58: 439-451

Chaloupská M (2015) Genetická diverzita *Vavilovia formosa* a *Pisum sativum* subsp. *elatius* (Fabeae, Fabaceae). Master Thesis, UP Olomouc.

Clement SL, McPhee KE, Elberson LR, Evans MA (2009).Pea weevil, *Bruchus pisorum* L. (Coleoptera: Bruchidae), resistance in *Pisum sativum*×*Pisum fulvum* interspecific crosses. *Plant Breeding* 12: 478-486.

Conicella C, Errico A (1990) Karyotype variations in *Pisum sativum* ect Sect. *abyssinicum*. *Caryologia* 43: 87-97.

Constantin GD, Krath BN, MacFarlane SA, Nicolaisen M, Johansen IE, Lund OS (2004) Virus-induced gene silencing as a tool for functional genomics in a legume species. *The Plant Journal* 40: 622-631.

Cook DR (1999) *Medicago truncatula* a model in the making. *Current Opinion in Plant Biology* 2: 301–304.

Corander J, Gyllenberg M and Koski T (2007) Random partition models and exchangeability for Bayesian identification of population structure. *The Bulletin of Mathematical Biology* 69: 797-815.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* 205: 19-31.

Courtois B, Audebert A, Dardou A *et al.* (2013) Genome-Wide Association Mapping of Root Traits in a Japonica Rice Panel. *PLoS ONE* 8: e78037

Coyne CJ, Brown A, Timmerman-Vaughan GM, McPhee KE, Grusak MA (2005) Refined USDA-ARS pea core collection based on 26 quantitative traits. *Pisum Genetics* 37: 3-6.

Coyne CJ, McClendon MT, Walling JG *et al.* (2007) Construction and characterization of two bacterial artificial chromosome libraries of pea (*Pisum sativum* L.) for the isolation of economically important genes. *Genome* 50: 871-875.

Coyne CJ, McGee R, Redden RJ, Ambrose MJ, Furman BJ, and Miles CA (2011) Genetic Adjustment to Changing Climates: Pea. In. Crop Adaptation to Climate Change. Yadav, SS, Redden, RJ, Hatfield, JL, Lotze-Campen, Hall, AE Eds, Wiley Blackwell, Chichester UK. 2011, Chapter 8, 238-250.

Cruz VM, Kilian A, Dierig DA (2013) Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related species. *PLoS One* 8: e64062.

Cubero JI (1981) Origin, domestication and evolution. In: Lentils. pp. 15–38. Webb C. and Harwtin, GC (Eds.) Commonwealth Agricultural Bureau, Slough.

Cubero JL, Perez de la Varga M, Fratini R (2009) Origin, phylogeny and spread. In: The Lentil, Botany, Production and Uses. pp. 13–33. Erskine, W., Muehlbauer, F. J., Sarker, A., and Sharma, B., Eds., CABI, Wallingford.

Czefranova Z (1971) Conspectus systematic generis *Lathyrus* L. *Novititas Systematic Plant Vascular* (Leningrad) 8: 191-201.

Dalmais M, Schmidt J, Le Signor C *et al.* (2008) A. UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biology* 9: R43.

Davis PH (1970) *Pisum* L. In: *Flora of Turkey and East Aegean Islands*. Vol. 3 Davis PH (ed.). Edinburg: Edinburg University Press, pp. 370-373.

De Beukelaer H, Smykal P, Davenport G, Fack V (2012) Core Hunter II: fast core subset selection based on multiple genetic diversity measures using Mixed Replica search. *BMC Bioinformatics* 13: 312.

De Candolle A (1884) *Origin of cultivated plants*. Whitefish: Kessinger Publishing, 2006.

De Martino T, Errico A, Lassandro A, Conicella C (2000) Distorted segregation resulting from pea chromosome reconstructions with alien segments from *Pisum fulvum*. *Journal of Heredity* 91: 322-325.

Dempewolf H, Hodgins KA, Rummell SE, Ellstrand NC, Rieseberg LH (2012) Reproductive Isolation during Domestication. *The Plant Cell* 24: 2710-271.

Desgroux A, L'Anthoëne V, Roux-Duparque M *et al.* (2016) Genome-wide association mapping of partial resistance to *Aphanomyces euteiches* in pea. *BMC Genomics* 17: 124.

Deulvot C, Charrel H, Marty A *et al.* (2010) Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics* 11: 468.

Dolezel J, Greilhuber J (2010) Nuclear genome size. Are we getting closer? *Cytometry* 77, 635–642.

Dong Y, Yang X, Liu J, Wang BH, Bo-Ling Liu BL, Wang YZ (2014) Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature Communications* 5: 3352.

Dostálová R, Seidenglanz M, Griga M (2005) Simulation and assessment of possible environmental risks associated with release of genetically modified peas (*Pisum sativum* L.) into environment in central Europe. *Czech Journal of Genetics and Plant Breeding* 41: 51-63.

Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T and Ohashi M (1997) A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. *American Journal of Botany* 84: 541–554.

Duarte J, Rivière N, Baranger A *et al.* (2014) Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. *BMC Genomics* 15: 126.

Duque P, Chua NH (2003) IMB1, a bromodomain protein induced during seed imbibition, regulates ABA- and phyA-mediated responses of germination in *Arabidopsis*. *Plant Journal* 35: 787-99.

Elbaum R, Melamed-Bassudo C, Boareto E, Galili E, Lev-Yadun S, Levy AA, Weiner S (2006) Ancient olive DNA in pits: preservation, amplification and sequence analysis. *Journal of Archeological Sciences* 33: 77–88.

Eliášová A, Trávníček P, Mandák B, Münzbergová Z (2012) Autotetraploids of *Vicia cracca* show a higher allelic richness in natural populations and a higher seed set after artificial selfing than diploids. *Annals of Botany* 113: 159–170.

Ellis THN (2011) *Pisum*. In: C Kole, ed. *Wild Crop Relatives: Genomic and Breeding Resources*. Berlin-Heidelberg: Springer-Verlag, 237-248.

Ellis THN, Poyser SJ, Knox MR, Vershinin AV and Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. *Molecular and General Genetics* 260: 9-19.

Endo Y, Choi BH, Ohashi H and Delgado-Salinas A (2008) Phylogenetic relationships of New World *Vicia* (Leguminosae) inferred from nrDNA Internal Transcribed Spacer Sequences and floral characters. *Systematic Botany* 33: 356-363.

Erdős L, Cserhalmi D, Bátori Z, Kiss T, Morschhauser T, Benyhe B, Dénes A (2013) Shrub encroachment in a wooded-steppe mosaic: combining gis methods with landscape historical analysis. *Applied Ecology and Environmental Research* 11: 371-384.

Erdős L, Méri Á, Bátori Z, Gallé R, Körmöczi L. (2012) North south facing vegetation gradients in the Villány Mountains: a case study on the population and the community level. *Pakistan Journal of Botany* 44: 927-932.

Errico A, Conicella C, De Martino T, Ercolanola R, Monti LM (1996) Chromosome reconstructions in *Pisum sativum* through interspecific hybridization with *P. fulvum*. *Journal of Genetics and Breeding* 50: 309-313.

Errico A, Conicella C, Venora G (1991) Karyotype studies on *Pisum fulvum* and *Pisum sativum* using a chromosome image analysis system. *Genome* 34: 105–108.

Erskine W (1985) Selection for pod retention and pod dehiscence in lentils. *Euphytica* 34: 105–112.

Erskine W and Muehlbauer FJ (1991) Allozyme and morpfological variability, outcrossing rate and core collection formation in lentil germplasm. *Theoretical and Applied Genetics* 83: 119-125.

Eshed Y, Gera G, Zamir D (1996) A genome-wide search for wild-species alleles that increase horticultural yield of processing tomatoes. *Theoretical and Applied Genetics* 93: 877-886.

Esquinas-Alacazar J (2005) Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nature Reviews in Genetics* 6: 946-953

Etterson JR, Shaw RG (2001) Constraint to adaptive evolution in response to global warming. *Science* 294: 151-154.

Fairbairn A, Near J, Martinoli D (2005) Macrobotanical investigations of the North, South and KOPAL areas at Catalhoyuk. In Ian Hodder (Ed.), Inhabiting Catalhoyuk: Reports from the 1995-1999 seasons (pp. 137-201) Cambridge, England, U.K.: McDonald Institute for Archaeological Research.

Fedorov AA (1952) Supplement. In: Grossheim AA, ed. *Flora of the Caucasus 5*, Moscow—St. Petersburg: Academy of Science of the USSR, 453. (in Russian)

Ferguson ME, Maxted N, Slageren MV, and Robertson LD (2000) A re-assessment of the taxonomy of *Lens* Mill. (Leguminosae, Papilionoideae, Vicieae). *Botanical Journal of the Linnean Society* 133: 41–59.

Fernandes V, Triska P, Pereira JB, Alshamali F, Rito T, Machado A *et al.* (2015) Genetic Stratigraphy of Key Demographic Events in Arabia. *PLoS ONE* 10: e0118625.

Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García-del-Rey E, Arévalo JR, Whittaker RJ (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography* 38: 226–246.

Ferraro K, Alena L Jin AL, Nguyen TD, Reinecke DM, Ozga JA, Ro DK (2014) Characterization of proanthocyanidin metabolism in pea (*Pisum sativum*) seeds. *BMC Plant Biology* 14: 238.

Finch-Savage WE, and Leubner-Metzger G (2006) Seed dormancy and the control of germination. *New Phytologist* 171: 501–523.

Folkeson D (1990) Assignment of linkage segments to the satellite chromosomes 4 and 7 in *Pisum sativum*. Hereditas 112: 257–263.

Fondevilla S, Küster H, Krajinski F, Cubero JI, Rubiales (2011) Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology. *BMC Genomics* 12: 28.

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

Ford R, Le Roux K, Itman C, Brouwer J.B. and Taylor PWJ (2002) Genome-specific sequence tagged microsatellite site (STMS) markers for diversity analysis and genotyping in *Pisum* species. *Euphytica* 124: 397–405.

Fraile A, Garcia-Arenal F (2010) The coevolution of plants and viruses: resistance and pathogenicity. *Advances in Virus Research* 76: 1–32.

Frankel OH and Brown AHD (1984) Current plant genetic resources - a critical appraisal. In Genetics: New Frontiers, Vol. IV, Oxford and IBH Publ. Co, New Delhi, India

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

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

Fuller D (2007) Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeological Insights from the Old World. *Annals of Botany* 100: 903-924.

Fuller DQ, Allaby RG, Stevens C (2010) Domestication as innovation: The entanglement of techniques, technology and chance in the domestication of cereal crops. *World Archaeology* 42: 13–28.

Fuller DQ, Willcox G, Allaby RG (2012b) Early agricultural pathways: Moving outside the 'core area' hypothesis in Southwest Asia. *Journal of Experimental Botany* 63: 617–633.

Fuller D *et al.* (2012a) Cultivation as slow evolutionary entanglement: comparative data on rate and sequence of domestication. *Vegetation History and Archaeobotany* 21: 131–145.

Funatsukia H, Suzukic M, Aya Hirosea A *et al.* (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences USA* 111: 17797–17802.

Gallardo K, Firnhaber C, Zuber H, Héricher D, Belghazi M, Henry C *et al.* (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds. *Molecular and Cellular Proteomics* 6: 2165–2179.

Gallego Lorrente M, Jones M.E.R, Eriksson A, Siska V *et al.* (2015) Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* 6262: 820-822.

Geltman D (1995) Guide Marschall von Bieberstein Herbarium. Komarov Botanical Institute, St. Petersburg, Russia. 1-23.

Gibbs AJ, Ohshima K, Phillips MJ, Gibbs MJ (2008) The Prehistory of Potyviruses: Their Initial Radiation Was during the Dawn of Agriculture. *PLoS ONE* 3: e2523.

Golubev AA (1990) Habitats, collection, cultivation and hybridization of *Vavilovia formosa* Fed. *Bulletin of Applied Botany, of Genetics and Plant Breedin*g 135: 67–75.

Gong Y, Xu S, Mao W, Hu Q, Zhang G, Ding J, Li Y (2010) Developing new SSR markers from ESTs of pea (*Pisum sativum* L.*). J Zhejiang Univ-Sci B.* 11: 702-707.

Govorov LI (1937) Goroch (Peas): In. *Kulturnaja Flora SSR*. Moscow: , p. 229-336 (in Russian)

Graeber K, Nakabayashi K, Miatton E, Leubner-Metzger G, Soppe WJJ (2012) Molecular mechanisms of seed dormancy. *Plant Cell and Environment* 35: 1769–1786.

Griga M, Smýkal P, Dostálová R, Seidenglanz M (2008) Field assessment of outcrossing rate in pea (*Pisum sativum* L.), a cleistogamous crop, as a background for release of genetically modified pea into environment. In: 10th international symposium on biosafety of genetically modified organisms, symposium handbook: biosafety research of GMOs: past achievements and future challenges. Wellington, New Zealand, p. 110.

Gronlund M, Olsen A, Johansen IE, Jakobsen I (2010) Protocol: using virus-induced gene silencing to study the arbuscular mycorrhizal symbiosis in *Pisum sativum. Plant Methods* 6: 28.

Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. *Trends in Plant Sciences* 15: 529–37.

Gunn C, and Kluve RJ (1976) Androecium and pistil characters for the tribe Vicieae (Fabaceae). *Taxon* 25: 563–575.

Gur A and Zamir D (2004) Unused genetic variation can lift yield barriers in plant breeding. *PLoS Biology* 2: e245.

Hall KJ, Parker JS, Ellis TH (1997) The relationship between genetic and cytogenetic maps of pea. I. Standard and translocation karyotypes. *Genome* 40: 744-54.

Hammer K (1984) Das Domestikationssyndrom. *Kulturpflanze* 11: 11–34.

Hammer K (2003) A paradigm shift in the discipline of plant genetic resources. *Genetic Resources and Crop Evolution* 50: 3–10.

Hampe A, Jump AS (2011) Climate Relicts: Past, Present, Future. *Annual Review of Ecology, Evolution, and Systematics*42: 313–33.

Hampton RO, Mink GI (1975) Pea seed-borne mosaic virus. CMI/AAB Descriptions of Plant Viruses. 146 n.

Han Y, Zhao X, Liu D, Li Y. *et al.* (2016) Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist* 209: 871–884.

Hancock JF (2012) Plant evolution and the origin of crop species. 3rd edn. CABI, Wallingford

Hanelt P, and Mettin D (1989) Biosystematics of the genus *Vicia* L. (Leguminosae). *Annual Review of Ecology, Evolution, and Systematics*20: 199–223.

Hardie DC, Baker GJ, Marshall DR (1995) Field screening of *Pisum* accessions to evaluate their susceptibility to the pea weevil (Coleoptera: Bruchidae). E*uphytica* 84: 155–161.

Harlan JR. (1975) Our vanishing genetic resources. *Science.* 188: 617-621.

Hedley CL, Lloyd JR, Ambrose MJ, Wang TL (1994) An analysis of seed development in *Pisum sativum*. XVII. The effect of the *rb* locus alone and in combination with r on the growth and development of the seed. *Annals of Botany* 74: 365-371.

Hecht V, Laurie RE, Vander Schoor JK *et al.* (2011) The pea GIGAS gene is a FLOWERING LOCUS T homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod. *Plant Cell* 23: 147-61.

Helbaek H (1959) Domestication of food plants in the Old World. *Science* 130: 365-72.

Helbaek H (1970) "The plant husbandry of Hacilar: A study of cultivation and domestication," in Excavations at Hacilar, vol. I. Edited by J. Mellaart, pp. i89-244. Edinburgh: Edinburgh University Press.

Hellens RP, Moreau C, Lin-Wang K *et al.* (2010) Identification of Mendel's white flower character. *PLoS ONE* 5: e13230.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Hjulsager CK, Olsen BS, Jensen DM, Cordea MI, Krath BN et al. (2006) Multiple determinants in the coding region of pea seed-borne mosaic virus P3 are involved in virulence against *sbm-2* resistance. *Virology* 355: 52–61.

Ho SY, Heupnik TH, Rambaut A, Shapiro B (2007) Bayesian estimation of sequence damage in ancient DNA. *Molecular Biology and Evolution* 24: 1416–1422.

Hoey BK, Crowe KR, Jones VM and Polans NO (1996) A phylogenetic analysis of *Pisum* based on morphological characters, allozyme and RAPD markers. *Theoretical and Applied Genetics* 92: 92-100.

Hofinger BJ, Jing HC, Hammond-Kosack KE, Kanyuka K (2009) High-resolution melting analysis of cDNA-derived PCR amplicons for rapid and cost-effective identification of novel alleles in barley. *Theoretical and Applied Genetics* 119: 851-865.

Hofinger BJ, Russell JR, Bass CHG, Baldwin T, dos Reis M *et al.* (2011) An exceptionally high nucleotide and haplotype diversity and a signature of positive selection for the *eIF4E* resistance gene in barley are revealed by allele mining and phylogenetic analyses of natural populations. *Molecular Ecology* 20: 3653–3668.

Holden D (2009) *Genetic studies of wide crosses between cultivated pea species, and the domestication of Pisum abyssinicum*. PhD Thesis, University of East Anglia. John Innes Centre & Unilever Research, Colworth.

Hopf M (1986) Archaeological evidence of the spread and use of some members of the Leguminosae family. In: The original and domestication of cultivated plants. pp. 35–60. Barigozzi, C. Ed, Oxford, Elsevier.

Hubner S, Gunther T, Flavell A, Fridman E, Graner A, Korol A *et al.* (2012) Islands and streams: clusters and gene flow in wild barley populations from the Levant. *Molecular Ecology* 21: 1115–1129.

Hubner S, Hoffken M, Oren E, Haseneyer G, Stein N, Graner A *et al.* (2009) Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Mol Ecology* 18: 1523–1536.

Chen H, Osuna D, Colville L, Lorenzo O, Graeber K, Küster H *et al.* (2013) Transcriptome-wide mapping of pea seed ageing reveals a pivotal role for genes related to oxidative stress and programmed cell death. *PLoS ONE* 8: e78471.

Cheng P, Holdsworth W, Ma Y *et al.* (2015) Association mapping of agronomic and quality traits in USDA pea single-plant collection. *Molecular Breeding* 35: 75.

Chetelat RT1, Meglic V, Cisneros P (2000) A genetic map of tomato based on BC(1) *Lycopersicon esculentum x Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. *Genetics* 154: 857-67.

Chimwamurombe PM, Khulbe RK (2011) Domestication. In: Pratap A, Kumar J, eds. *Biology and Breeding of Food Legumes*. Cambridge: CABI, 19–34.

Ibiza VP, Canizares J, Nuez F (2010) EcoTILLING in *Capsicum* species: searching for new virus resistance. *BMC Genomics* 11: 631.

Ido Y, Nakahara K.S, Uyeda I (2012) White clover mosaic virus-induced gene silencing in pea. *Journal of General Plant Pathology* 78: 127–132.

Isemura T, Kaga A, Tabata S, Somta P, Srinives P, Shimizu T *et al.* (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE* 7: e41304.

Isemura T, Kaga A, Tomooka N, Shimizu T, Vaughan DA (2010) The genetics of domestication of rice bean, *Vigna umbellata*. *Annals of Botany* 106: 927–944.

Jakob SS, Dennis Rödder D, Engler JO, Shaaf S, Özkan H, Blattner FR, Kilian B (2014) Evolutionary history of wild barley (*Hordeum vulgare subsp. spontaneum*) analyzed using multilocus sequence data and paleodistribution modeling. *Genome Biol Evol.* 6: 685–702.

Jang SJ, Sato M, Sato K, Jitsuyama Y, Fujino K, Mori H *et al.* (2015) A Single-Nucleotide Polymorphism in an Endo-1,4-β-Glucanase Gene Controls Seed Coat Permeability in Soybean. *PLoS ONE* 10: e0128527.

Jansen J, van Hintum T (2007) Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theoretical and Applied Genetics* 114: 421–428.

Jeong HJ, Kwon JK, Pandeya D, Hwang J, Hoang NH *et al.* (2012) A survey of natural and ethyl methane sulfonate-induced variations of *eIF4E* using high resolution melting analysis in Capsicum. *Molecular Breeding* 29: 349–360.

Jin A, Ozga JA, Lopes-Lutz D, Schieber A, Reinecke MD (2012) Characterization of proanthocyanidins in pea (*Pisum sativum* L.), lentil (*Lens culinaris* L.), and faba bean (*Vicia faba* L.) seeds. *Food Research International* 46: 528–535.

Jing R, Ambrose MA, Knox MR *et al.* (2012) Genetic diversity in European *Pisum* germplasm collections. *Theoretical and Applied Genetics* 125: 367-80.

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis TH, Flavell AJ (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). *Genetics* 177: 2263-75

Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis THN, Flavell AJ (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evolutionary Biology* 10: 44.

Jing RC, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH, Flavell AJ (2005) Insertional polymorphism and antiquity of *PDR1* retrotransposon insertions in *Pisum* species. *Genetics* 171: 741-752.

Joannin S, Cornee J, Munch P *et al.* (2010) Early Pleistocene climate cycles in continental deposits of the Lesser Caucasus of Armenia inferred from palynology, magnetostratigraphy, and 40Ar/39Ar dating. *Earth and Planetary Sciences Letters* 291: 149–158.

Jones M, Brown T (2000) Agricultural origins: the evidence of modern and ancient DNA. *Holocene* 10: 769–776.

Juračková Ž (2012) Dormance semen planých forem hrachu (*Pisum sativum* subsp. *elatius*) v kontextu domestikace kulturního hrachu. Bachellor thesis. UPOL

Kalo P, Seres A, Taylor A, Jaka SA, Kevei J, Kereszt ZA, Ellis THN, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Molecular and General Genomics 272*, 235–246.

Kass E and Wink M (1996) Molecular evolution of the Leguminosae: Phylogeny of the Three subfamilies based on *rbcL*-sequences. *Biochemical Systematics and Ecology* 24: 365-78

Kassa MT, Penmetsa RV, Carrasquilla-Garcia N *et al.* (2012) Genetic Patterns of Domestication in Pigeonpea (*Cajanus cajan* (L.) Millsp.) and Wild Cajanus Relatives. *PLoS ONE* 7: e39563.

Kaur S, Pembleton LW, Cogan NO (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 13: 104.

Keiša A, Maxted N, Ford-Lloyd BV (2007) The assessment of biodiversity loss over time: wild legumes in Syria. *Gen. Res. Crop Evol.* 55: 603-612.

Kenicer GJ (2007) Systematics and biogeography of *Lathyrus* L. (*Leguminosae, papilionoideae*). PhD Thesis, Royal Botanical Garden Edinburgh.

Kenicer GJ, Kajita T, Pennington RT and Murata J (2005) Systematics and biogeography of *Lathyrus* (*Leguminosae*) based on internal transcribed spacer and cpDNA sequence data. *American Journal of Botany* 92: 1199–1209.

Kilian A, Wenzl P, Huttner E *et al.* (2012) Diversity Arrays Technology: a generic genome profiling

technology on open platforms. *Methods in Molecular Biology* 888: 67–89.

Kislev ME, Bar-Yosef O (1988) The Legumes: The Earliest Domesticated Plants in the Near East? *Current Anthropology* 29: 175–179.

Kloz J (1971) Serology of the Leguminosae. In: Harborne JB, Boulter D, Turner BL, eds. *Chemotaxonomy of the Leguminosae*. London: Academic Press, 309-365.

Knight TA. (1799) Experiments on the Fecundation of Vegetables. *Philosophical Transactions of the Royal Society* 89: 504-509.

Koizumi M, Kikuchi K, Isobe S, Ishida N, Naito, Kano H (2008) Role of seed coat in imbibing soybean seeds observed by micro-magnetic resonance imaging. *Annals of Botany* 102: 343-352.

Konečná E, Šafářová D, Navrátil M *et al.* (2014) Geographical gradient of the *eIF4E* alleles conferring resistance to potyviruses in pea *(Pisum*) germplasm. *PLOS One* 9: e90394.

Kongjaimum A, Kaga A, Tomooka N *et al.* (2012) The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-*gr. Sesquipedalis*. *Annals of Botany* 109: 1185–1200.

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

Kosterin OE and Bogdanova VS (2008) Relationship of wild and cultivated forms of *Pisum* L. as inferred from an analysis of three markers, of the plastid, mitochondrial and nuclear genomes. *Genetic Resources and Crop Evolution* 55: 735-755.

Kosterin O (2014) The lost ancestor of the broad bean (*Vicia faba* L.) and the origin of plant cultivation in the Near East. *Vavilov Journal of Genetics and Breeding* 18: 831-840.

Kosterin OE, Bogdanova VS (2014) Efficiency of hand pollination in different pea (*Pisum*) species and subspecies. *Indian Journal of Genetics* 74: 50-55.

Kosterin OE, Zaytseva OO, Bogdanova VS and Ambrose M (2010) New data on three molecular markers from different cellular genomes in Mediterranean accessions reveal new insights into phylogeography of *Pisum sativum* L. subsp *elatius* (Bieb.) Schmalh. *Genetic Resources and Crop Evolution* 57: 733-739.

Kovinich N,Saleem A,Arnason JT, Miki B (2011) Combined analysis of transcriptome and metabolite data reveals extensive differences between black and brown nearly-isogenic soybean(*Glycine max*) seed coats enabling the identification of pigment isogenes. *BMC Genomics* 12: 381.

Kroll H (1991) Southeast Europe. In: van Zeist, Wasylikowa K, Behre KE (eds) Progress in old world palaeoethnobotany. Balkema, Rotterdam, pp 161–177.

Kumar S, Gupta P, Barpete S, Sarker A, Amri A, Mathur PN, and Baum M (2013) Grass pea. pp. 269–305. In: Genetic and Genomic Resources of Grain Legume Improvement. Singh, M, and Upadhya, H, Eds, Elsevier, Netherlands.

Kupicha FK (1981) *Vicieae* (Adans.) DC. (1825) nom conserv prop. In. Polhill RM, Raven PH (ed.) *Advances in legume systematics* 1. Royal Botanical Gardens, Kew, pp. 377-381.

Kwon SJ, Brown AF, Hu J *et al.* (2012) Population genetic sub-structure within the USDA-ARS *Pisum* core collection and its potential as a platform for association mapping. *Genes and Genomics* 34: 305–320.

Ladizinsky G (1979) The origin of lentil and its wild genepool. *Euphytica* 28: 179–187.

Ladizinsky G (1986) A new *Lens* species from the Middle-East. *Not. R. Bot. Gard. Edinburg* 43: 489–492.

Ladizinsky G, and Sakar D (1982) Morphological and cytogenetical and characterization of *Vicia montbretii* Fisch and Mey (Synonym *Lens montbretii* (Fisch and Mey) Davis and Plitmann. *Botanical Journal of the Linnean Society* 85: 209–212.

Ladizinsky G, Braun D, Goshen D, and Muehlbauer FJ (1984) The biological species of the genus *Lens* L. (*Lens nigricans*). *Botanical Gazette* 145: 253–261.

Lamarck JB (1778) *Flore Francoise*. l'Imprimerie Royale, Paris. http://dx.doi.org/10.5962/ bhl.title.9461

Lamm R, Miravalle RJ (1959) A translocation tester set in *Pisum*. Hereditas 45: 417– 440.

Lamm R. (1977) Transpositions in Pisum. *Pisum Newsletter* 9: 28 –29.

Lamprecht H (1963) Zur Kenntnis von *Pisum arvense* L. *oect. abyssinicum* Braun, mit genetischen und zytologischen Ergebnissen. *Agri Hort Genetics* 21:35–55.

Lamprecht H (1966) Die Enstehung der Arten und hoheren Kaategorien. Wien: Springer Verlag.

Latham LJ, Jones RAC (2001) Alfalfa mosaic and pea seed-borne mosaic viruses in cool season crop, annual pasture, and forage legumes: susceptibility, sensitivity, and seed transmission. *Australian Journal of Agricultural Research* 52: 771–790.

Lavin M, Herendeen PS and Wojciechowski M (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* 54: 575–594.

Le Gall O, Aranda MA, Caranta C (2011) Plant resistance to viruses mediated by translation initiation factors. In: Caranta C, Aranda MA, Tepfer M & Lopez-Moya J, editors. Recent Advances in Plant Virology. Norfolk: Caister Academic Press. pp. 177–194.

Lehmann C (1954) Das morphologische system der Saaterbsen. *Der Zuchter* 24: 316-337.

Lenser T, Theißen G (2013) Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Sciences* 18:704–714.

Leubner-Metzger G (2005) beta-1,3-Glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening. *The Plant Journal* 41: 133-145.

Lev-Yadun S, Gopher A, Abbo S (2000) The cradle of agriculture. *Science* 288: 1602–1603.

Lewis G, Schrirer B, Mackinder B and Lock M (eds.) (2005) Legumes of the World. Royal Botanical Gardens, Kew

Li J, Redden RJ, Zong X, Berger J D, Bennett SJ (2013) Ecogeographic analysis of pea collection sites from China to determine potential sites with abiotic stresses. *Genetic Resources and Crop Evolution* 60: 1801-1815.

Lie TA (1984) Host genes in *Pisum sativum* L. conferring resistance to European *Rhizobium leguminosarum* strains. *Plant and Soil* 82: 415-425.

Linnaeus C (1753) *Species Plantarum*. Vol. 2. Stockholm: Salvius.

Liu C, Ji Hyung Jun JH, Dixon RA (2014) MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiology* 165: 1424–1439.

Lock M and Maxted N (2005) Tribe *Fabeae. In:* Lewis G, Schrire B, Mackinder B and Lock M (eds.) *Legumes of the World*. Richmond: Royal Botanic Gardens, Kew.

Loridon K, McPhee KE, Morin J *et al.* (2005) Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). *Theoretical and Applied Genetics*111: 1022-1031.

Lovisolo O, Hull R, Rosler O (2003) Coevolution of viruses with hosts and vectors and possible paleontology. *Advances in Virus Research* 62: 325–379.

Lucau-Danila A, Toitot C, Goulas E *et al.* (2012) Transcriptome analysis in pea allows to distinguish chilling and acclimation mechanisms. *Plant Physiology and Biochemistry* 58: 236-44.

Luo T, Luo S, Araújo WL, Schlicke H, Rothbart M, Yu J, Fan T, Fernie AR, Grimm B, Luo M (2013) Virus-induced gene silencing of pea CHLI and CHLD affects tetsaharapyrrole biosynthesis, chloroplast development and the primary metabolic network. *Plant Physiology and Biochemistry* 65: 17-26.

Macas J, Neumann P and Navrátilová A (2007) Repetitive DNA in the pea *(Pisum sativum* L. genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8: 427.

Magee AM, Aspinall S, Rice DW *et al.* (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research* 20: 1700-1710.

Majeed M, Safdar W, Ali B, Mohammad A, Ahmad I, Mumtaz AS (2012) Genetic assessment of the genus *Pisum* L. based on sequence specific amplification polymorphism data. *Journal of Medicinal Plants Research* 6: 959–967.

Makasheva RK. (1979) Gorokh (pea) In: Korovina ON (ed.) Kulturnaya flora SSR. Leningrad: pp. 1-324 (in Russian)

Mallikarjuna N, Saxena KB (2005) A new cytoplasmic nuclear male-sterility system derived from cultivated pigeonpea cytoplasm. *Euphytica* 142: 143-148.

Mallikarjuna N, Sharma HC, Upadhyaya HD (2007) Exploitation of wild relatives of pigeonpea and chickpea for resistance to *Helicoverpa armigera*. *An Open Access Journal published by ICRISAT* 3: 1-4.

Marbach I, and Mayer AM (1974) Permeability of seed coats to water as related to drying conditions and metabolism of phenolics. *Plant Physiology* 54: 817–820.

Marles SMA, and Gruber MY (2004) Histochemical characterization of unextractable seed coat pigments and quantification of extractable lignin in the Brassicaceae. *Journal of the Science of Food and Agriculture* 84: 251–262.

Martín-Sanz A, Caminero C, Jing R, Flavell AJ, Pérez de la Vega M. (2011) Genetic diversity among Spanish pea (*Pisum sativum* L.) landraces, pea cultivars and the world *Pisum* sp. core collection assessed by retrotransposon-based insertion polymophisms (RBIPs). *Spanish Journal of Agricultural Research* 9: 166-78.

Marx GA (1977) Classification, genetics and breeding. In: Sutcliffe JF, Pate JS (ed) *Physiology of the Garden Pea.* In: Sutcliffe JF and Pater JS (eds.) New York: Academic Press, pp. 21-43.

Mass B, Usongo M (2007) Changes in seed characteristics during the domestication of the lablab bean (*Lablab purpureus* (L.) Sweet: Papilionoideae). *Australian Journal of Agricultural Research* 58: 9-19.

Matamoros MA, Loscos J, Dietz K-J, Aparicio-Tejo PM, Becana M (2009) Function of antioxidant enzymes and metabolites during maturation of pea fruits. *Journal of Experimental Botany* 61: 87–97.

Maxted N, Ambrose M (2001) Peas (*Pisum* L.). In: Maxted N, Bennett SJ, eds. *Plant Genetic Resources of Legumes in the Mediterranean*. Dordrecht: Kluwer Academic Publishers, 181-190.

McCallum C.M, Comai L, Greene EA and Henikoff S (2000) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiology* 123: 439–442.

McCouch S (2004) Diversifying Selection in Plant Breeding. *PLoS Biology* 2: e347.

McPhee K, Kraft JM, Tullu A, Muehlabuer F (1999) Resistance to Fusarium Wilt Race 2 in the Pisum Core Collection. ournal of the American Society for Horticultural Science. *American Society for Horticultural Science* 124: 28-31.

Melamed Y, Plitman U, Kislev ME (2008) *Vicia peregrina*: an edible early Neolithic legume. *Vegetation History and Archaeobotany* 17: S29–S34.

Mendel JG (1866) Versuche über Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereins in Brünn* 4: 3–47. (see also http://www.mendelweb.org/)

Meyer CJ, Steudle E, and Peterson CA (2007) Patterns and kinetics of water uptake by soybean seeds. *Journal of Experimental Botany* 58: 717–732.

Meziadi C, Blanchet S, Richard MS *et al.* (2016) Bean pod mottle virus: a new powerful tool for functional genomics studies in *Pisum sativum*. *Plant Biotechnology Journal* 14: 1777–1787.

Mikič A (2012) Origin of the words denoting some of the most ancient Old World pulse crops and their diversity in modern European languages. *PLoS One* 7: e44512.

Mikič A, Smýkal P, Kenicer G *et al.* (2013) The bicentenary of the research on 'beautiful' *Vavilovia* (*Vavilovia formosa*), a legume crop wild relative with taxonomic and agronomic potential. *Botanical Journal of the Linnean Society* 172: 524–531.

Mikič A. (2015) The First Attested Extraction of Ancient DNA in Legumes (Fabaceae). *Frontiers in Plant Sciences* doi.org/10.3389/fpls.2015.01006

Miller P (1768) *The Gardener's Dictionary, Containing the methods of cultivating and improving the kitchen, fruit and flower garden*. etc. (8th Edn), London

Moïse JA, Han S, Gudynaite-Savitch L, Johnson DA, Miki BLA (2005) Seed coats: structure, development, composition, and biotechnology. *In Vitro Cellular & Developmental Biology Plant* 41: 620–644.

Moreau C, Ambrose MJ, Turner L, Hill L, Ellis TH, Hofer JM (2012) The B gene of pea encodes a defective flavonoid 3,5-hydroxylase, and confers pink flower color. *Plant Physiology* 159: 759–768.

Muratova VS (1931) Common beans (*Vicia faba* L.). *Bull. Appl. Bot. Genet. Pl. Breed. Suppl.* 50: 1–298.

Murfet IC, and Reid JB (1993) Developmental mutants. In Peas: Genetics, Molecular Biology and Biotechnology, R. Casey and D.R. Davies, eds (Wallingford, UK: CAB International), pp. 165–216.

Musil M (1966) Uber das Vorkommen des Virus des Blattrollens der Erbse in der Slowakei (Vorlaufige Mitteilung). *Biologia 21*: 133–138.

Liu N, Guwen Zhang G, Xu S, Mao W, Qizan Hu Q, Gong Y (2015) Comparative Transcriptomic Analyses of Vegetable and Grain Pea (*Pisum sativum* L.) Seed Development. *Frontiers in Plant Sciences* doi.org/10.3389/fpls.2015.01039

Nakhutsrishvili G (2013) High-mountain vegetation. In: Nakhutsrishvili G (ed) The Vegetation of Georgia (South Caucasus) Springer Berlin Heidelberg, 119-208.

Nanni L, Bitocchi E, Bellucci E *et al.* (2011) Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 123: 1341-1357.

Nasiri J, Haghnazari A and Saba J (2009) Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on SSR markers. *African Journal of Biotechnology* 15: 3405-3417.

Neumann P, Navrátilová A, Schroeder-Reiter E *et al.* (2012). Stretching the rules: monocentric chromosomes with multiple centromere domains. *PLoS Genetics* 8: e1002777.

Neumann P, Pavlíková Z, Koblížková A *et al.* (2015). Centromeres off the hook: massive changes in centromere size and structure following duplication of CenH3 gene in Fabeae species. *Molecular Biology and Evolution* 32: 1862–1879.

Neumann P, Pozárková D, Vrána J, Dolezel J, Macas J (2002) Chromosome sorting and PCR-based physical mapping in pea (*Pisum sativum* L.). *Chromosome Research 10*: 63-71.

Neumann P, Schubert V, Fuková I, Manning JE, Houben A, Macas J (2016) Epigenetic histone marks of extended meta-polycentric centromeres of *Lathyrus* and *Pisum* chromosomes. *Frontiers in Plant Sciences* 7: 234.

Nieto C, Piron F, Dalmais M *et al.* (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. *BMC Plant Biology* 7: 34.

Nikolaeva MG (1969) Physiology of Deep Dormancy in Seeds. Izdatel'stvo Nauka. Leningrad (Translated from Russian by Z. Shapiro, NSF, Washington, D.C.)

Nonogaki H (2014) Seed dormancy and germination—emerging mechanisms and new hypotheses. *Frontiers in Plant Sciences* 5: 233.

Ohlemüller R, Anderson BJ, Araújo MB, Butchart SH, Kudrna O, Ridgely RS, Thomas CD (2008) The coincidence of climatic and species rarity: high risk to small-range species from climate change. *Biology Letters* 4: 568-572.

Ochatt SJ, Benabdelmouna A, Marget P, Aubert G, Moussy F, Pontecaille C and Jacas L (2004) Overcoming hybridization barriers between pea and some of its wild relatives. *Euphytica* 137: 353–359.

Ochatt SJ, Abirached-Darmency M, Marget P, Aubert G (2007) The Lathyrus paradox: 'poor men's diet' or a remarkable genetic resource for protein legume breeding? In: Ochatt S, Jain SM, editors. Breeding of neglected and under-utilised crops, spices and herbs. Enfield, NH: Science Publishers, pp. 41–60.

Olsen K.M, Wendel JF (2013) A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. *Annu. Rev. Plant Biol.* 64: 47–70.

Özkan H, Kafkas S, Sertac Ozer M, Brandolini A (2005) Genetic relationships among South-East Turkey wild barley populations and sampling strategies of *Hordeum spontaneum*. *Theoretical and Applied Genetics* 112: 12–20.

Palmer JD, Jorgensen RA and Thompson WF (1985) Chloroplast DNA Variation and Evolution in *Pisum*: Patterns of Change and Phylogenetic Analysis. *Genetics* 109: 195-213.

Palmer S, Moore JD, Clapham AJ, Rose P, Allby RG (2009) Archelogical evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS One* 4: e6301.

Pang YZ, Peel GJ, Wright E, Wang Z, Dixon RA (2007) Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. *Plant Physiology* 145: 601–615.

Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics* 106: 239–250.

Parween S, Nawaz K, Roy R *et al.* (2015) An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Scientific Reports* 5: 12806.

Patil G, Valliyodan B, Deshmukh R *et al.* (2015) Soybean (*Glycine max*) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics* 16: 520.

Pearce SR, Knox M, Ellis TH, Flavell AJ and Kumar A (2000) Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Molecular and General Genetics* 263: 898-907.

Pistorius R (1997) Scientists, plants and politics. A history of plant genetic resources movement. IPGRI, Rome, Italy.

Poets AM, Fang Z, Clegg MT, Morrell PL (2015) Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biology* 16: 173.

Polans NO and Saar DE (2002) ITS sequence variation in wild species and cultivars of pea. *Pisum Genetics* 34: 9-13.

Polans NO, Moreno RR (2009) Microsatellite and ITS sequence variation in wild species and cultivars of pea. *Pisum Genetics* 41: 3-6.

Pourkheirandish M, Hensel G, Kilian B *et al.* (2015) Evolution of the grain dispersal system in barley. *Cell* 162: 527-539.

Pradhan S, Adams RLP (1995) Distinct CG and CNG methyltransferases in *Pisum sativum*. *Plant Journal 7*: 471-481.

Pradhan S, Bandhiwal N, Shah N, Kant C, Gaur R, Bhatia S (2014) Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Frontiers in Plant Sciences* 2014, 5: 698.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Provvidenti R (1990) Inheritance of Resistance to Pea Mosaic Virus in *Pisum sativum*. *Journal of Heredity* 81: 143-145.

Puonti-Kaerlas J, Eriksson T, Engstroem P (1999) Production of transgenic pea (*Pisum sativum* L.) plants by Agrobacterium-mediated gene transfer. *Theoretical and Applied Genetics* 80: 246–252.

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457:843–848.

Radchuk R, Emery NJR, Weier D *et al.* (2010) Sucrose non-fermenting kinase 1 (SnRK1) coordinates metabolic and hormonal signals during pea cotyledon growth and differentiation. *The Plant Journal* 61: 324-338.

Ramirez-Villegas J, Khoury C, Jarvis A, Debouck DG, Guarino L (2010) A gap analysis methodology for collecting crop gene pools: A case study with *Phaseolus* beans. *PLoS ONE* 5: e13497.

Raven PH, Polhill RM (1981) Biogeography of the Leguminosae. In: Polhill, RM, Raven PH, eds. *Advances in Legume Systematics* Kew, RBG, 27-34.

Reeves PA, Panella LW, Richards CHM (2012) Retention of agronomically important variation in germplasm core collections: implications for allele mining. *Theoretical and Applied Genetics* 124: 1155–1171.

Reif JC *et al.* (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Science* 45: 1-7.

Riebeseel E, Häusler RE, Radchuk R *et al.* (2010) The 2-oxoglutarate/malate translocator mediates amino acid and storage protein biosynthesis in pea embryos. *Plant* Journal 61: 350-363.

Riehl S, Zeidi M, and Conard, NJ (2013) Emergence of agriculture in the foothills of the Zagros Mountains of Iran. *Science* 341: 65–67.

Rigola DJ, van Oeveren J, Janssen A *et al.* (2009) High throughput detection of induced mutations and natural variation using KeyPoint technology. *PLoS ONE* 4: e4761.

Rodriguez M, Rau D, Bitocchi E *et al.* (2015) Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytologist* 209: 1781-1794.

Rosenberg NA (2002) Genetic structure of human populations. *Science* 298: 2381-2002.

Rubiales D, Moreno MT, Sillero JC (2005) Search for resistance to crenate broomrape (*Orobanche crenata*) in pea germplasm. *Genetic Resources and Crop Evolution* 52: 853-861.

Rubio M, Nicolaı M, Caranta C, Palloix A (2009) Allele mining in the pepper gene pool provided new complementation effects between *pvr2-eIF4E* and *pvr6-eIF(iso)4E* alleles for resistance to pepper veinal mottle virus. *Journal of General Virology* 90: 2808–2814.

Ruffel S, Gallois J-L, Moury B, Robaglia C, Palloix A *et al.* (2006) Simultaneous mutations in translation initiation factors *eIF4E* and *eIF(iso)4E* are required to prevent Pepper veinal mottle virus infection of pepper. *Journal of General Virology* 87: 2089–2098.

Saar DE and Polans NO (2000) ITS sequence variation in selected taxa of *Pisum*. *Pisum Genetics* 109: 195-213.

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* 7: 511-23.

Saha GC, Vandemark GJ (2012) Evaluation of expression stability of candidate references genes among green and yellow pea cultivars (*Pisum sativum* L.) subjected to abiotic and biotic stress. *American Journal of Plant Sciences* 3: 235–242.

Sarıkamış G, Yanmaz R, Ermiş S, Bakir M, Yüksel C (2010) Genetic characterization of pea (*Pisum sativum*) germplasm from Turkey using morphological and SSR markers. *Genetics and Molecular Research* 9: 591–600.

Sato S, NakamuraY, Kaneko T *et al.* (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Research* 15: 227–239.

Saxena KB, Kumar RV, Srivastava N, Bao S (2005) A cytoplasmic-nuclear male-sterility system derived from a cross between *Cajanus cajanifolius* and *Cajanus cajan*. *Euphytica* 145: 289-294.

Severin AJ, Woody JL, Bolon YT *et al.* (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology* 10: 160.

Schaefer H, Hechenleitner P, Santos-Guerra A *et al.* (2012) Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. *BMC Evolutionary Biology* 12: 250.

Schindler F (1890) Welches Werthverhaltniss besteht zwischen den Landrassen landwirthschaftlicher Kulturpflanzen und den sogenannten Zu¨chtungsrassen ? Internationaler landund forstwirtschaftlicher Congress zu Wien 1890. Section I: Landwirthschaft. Subsection: Pflanzenbau. Frage 5. Heft 13: 19–24.

Schlee M, Göker M, Grimm GW, Hemleben A (2011) Genetic patterns in the *Lathyrus pannonicus* complex (Fabaceae) reflect ecological differentiation rather than biogeography and traditional subspecific division. *Botanical Journal of the Linnean Society* 165: 402-421.

Schlumbaum A, Tensen M, Jaenicke-Despre´s V (2008) Ancient plant DNA in archaeobotany. *Vegetation History and Archaeobotany* 17: 233–244.

Schmalhausen I (1895) Flora Srednei y Yuzhnoj Rossii, Kryma i Severnogo Kavkaza, vol 1. pp. 468 (in Russian)

Schmutz J, Cannon SB, Schlueter J *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178-183.

Scholthof KB, Adkins S, Czosnek H *et al.* (2010) Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology* 12: 938-954.

Schrire BD, Lavin M, Lewis GP (2005) Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biologiske Skrifter* 55: 375-422.

Sindhu A, Ramsey L, Sanderson LA *et al.* (2014) Gene-based SNP discovery and genetic mapping in pea. *Theoretical and Applied Genetics* 10: 2225-2241.

Smartt J (1990) *Grain Legumes: Evolution and Genetic Resources*. Cambridge: Cambridge University Press.

Smýkal P (2006) Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. *Journal of Applied Genetics* 47: 221–230.

Smýkal P (2014) Pea (*Pisum sativum* L.) in biology prior and after Mendel's discovery. *Czech Journal of Genetics and Plant Breeding* 50: 52–64.

Smýkal P, Coyne C, Ambrose MJ *et al.* (2015) Legume crops phylogeny and genetic diversity for science and breeding. *Critical Reviews in Plant Sciences* 34: 43-104.

Smýkal P, Coyne C, Redden R, Maxted N (2013) Peas. In: Singh M, Upadhyaya HD, Bisht IS, eds. *Genetic and Genomic Resources of Grain Legume Improvement.* Amsterdam: Elsevier.

Smýkal P, Coyne CJ, Ford R *et al.* (2008a) Effort towards a world pea (*Pisum sativum* L.) germplasm core collection: The case for common markers and data compatibility. *Pisum Genetics* 40: 11-14.

Smýkal P, Horáček J, Dostálová R, Hýbl M (2008b) Variety discrimination in pea (*Pisum sativum* L.) by molecular, biochemical and morphological markers. *Journal of Applied Genetics* 49: 155-166.

Smýkal P, Hýbl M, Corander J, Jarkovský J, Flavell A, Griga M (2008c) Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. *Theoretical and Applied Genetics* 117: 413-424.

Smýkal P, Kalendar R, Ford R, Macas J, Griga M (2009) Evolutionary conserved lineage of *Angela*-like retrotransposons as a genome-wide microsatellite repeat dispersal agent. *Heredity* 103: 157–167.

Smýkal P, Kenicer G, Flavell AJ *et al.* (2011) Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. *Plant Genetic Resources: Characterization and Utilization* 9: 4–18.

Smýkal P, Konečná E (2014) Advances in Pea Genomics. In: *Legumes in the Omic Era*. Chapter 15. Gupta S, Nadarajan N, and Gupta D, Eds, Springer, Dordrecht, Netherlands.

Smýkal P, Kosterin O (2010) Towards introgression library carrying wild pea (*Pisum fulvum*) segments in cultivated pea (*Pisum sativum*) genome background. In: *Book of Abstracts of Vth International Congress on Legume Genetics and Genomics*. Asilomar, USA, 123.

Smýkal P, Vernoud V, Blair MW, Soukup A, Thompson RD (2014) The role of the testa during development and in establishment of dormancy of the legume seed. *Frontiers in Plant Science* 5: 351.

Somers DA, Samac DA, Olhoft PM (2003) Recent advances in legume transformation. *Plant Physiology* 131: 892-899.

Sprent JI, Ardley JK, James EK (2013) From North to South: a latitudinal look at legume nodulation processes. *South African Journal of Botany*, 89: 31-41.

Spurny M (1963) Cell wall structure of epidermal cells of the pea seed coat (*Pisum sativum* L.) studied by microcinematography. *Mikroskopie* 18: 272–279.

Spurny M (1964) Changes in the permeability of the seed coat inconnection with the development of suberin adcrustations of the macrosclereids from the

seed coat of the pea(*Pisum sativum* L.). *Flora* 154: 547–567.

Steele KP and Wojciechowski MF (2003) Phylogenetic analyses of tribes *Trifolieae* and *Vicieae*, based on sequences of the plastid gene *matK* (*Papilionoideae: Leguminosae*) In: Klitgaard BB and Bruneau (eds.) *Advances in Legume Systematics*. Royal Botanical Garden, Kew, pp. 355-370.

Sudheesh S, Sawbridge TI, Cogan NOI, Kennedy P, Forster JW, Kaur S (2015) De novo assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics* 16.

Sun L, Miao Z, Cai C *et al.* (2015) *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nature Genetics* : doi:10.1038/ng.3339

Švabová L, Smýkal P, Griga M, Ondrej V (2005) *Agrobacterium*-mediated transformation of *Pisum sativum in vitro* and *in vivo*. *Biologia Plantarum* 49: 361-370.

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063-1066.

Tanno K I and Willcox G (2006) How Fast Was Wild Wheat Domesticated? *Science* 311: 1886.

Tar'an B, Zhang C, Warkentin T, Tullu A and Vandenberg A (2005) Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters. *Genome* 48: 257–272.

Tayeh N, Aluome C, Falque M *et al.* (2015a) Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high density, high resolution consensus genetic map. *The Plant Journal* 84: 1257-1273.

Tayeh N, Klein A, Le Paslier M-C *et al.* (2015b) Genomic prediction in pea: effect of marker density and training population size and composition on prediction accuracy. *Frontiers in Plant Science* 6: 941.

Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M and Davenport GF (2009) Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinformatics* 10: 243.

Thompson R, Burstin J, Gallardo K (2009) Post-genomics studies of developmental processes in legume seeds. *Plant Physiology* 151: 1023–1029.

Townsend CC (1968) Contributions to the Flora of Iraq: V: Notes on the Leguminales. *Kew Bulletin*, 435-458.

Triques K, Sturbois B, Gallais S *et al.* (2007) Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. *The Plant Journal* 51: 1116–1125.

Troszyńska A, Ciska E (2002) Phenolic compounds of seed coats of white and coloured varieties of pea (*Pisum sativum* L.) and their total antioxidant activity. *Czech Journal of Food Sciences* 20: 15–22.

Upadhyaya HD, Dwivedi SL, Ambrose M *et al.* (2011) Legume genetic resources: management, diversity assessment, and utilization in crop improvement. *Euphytica* 180: 27-47.

Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, Gowda CL, Hoisington D, Singh S (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biology* 16: 106.

Van Hintum TJL (1999) The general methodology for creating a core collection. In: Johnson RC, T. Hodgkin T (eds.) *Core collections for today and tomorrow*. Rome: International Plant Genetic Resources Institute, pp. 10-17.

van Oss R, Abbo S, Eshed R *et al.* (2015) Genetic Relationship in Cicer Sp. Expose Evidence for Geneflow between the Cultigen and Its Wild Progenitor. *PLoS ONE* 10: e0139789.

Varshney RK, Song C, Saxena RK *et al.* (2013a) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology* 31: 240–246.

Vavilov NI (1926) Studies on the origin of cultivated plants. (Russian) *Bulletin of Applied Botany and Plant Breeding* 14: 1–245.

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Translated from the Russian by K. Starchester. *Chronica Botanica* 13: 1–364.

Vavilov NI (1992) Origin and Geography of Cultivated Plants. Cambridge University Press, Cambridge.

Vaz Patto MC, Rubiales D (2014) *Lathyrus* diversity: available resources with relevance to crop improvement – *L. sativus* and *L. cicera* as case studies. *Annals of Botany* 113: 895–908.

Verdier J, Dessaint F, Schneider C, Abirached-Darmency M (2013a) A combined histology and transcriptome analysis unravels novel questions on *Medicago truncatula* seed coat. *Journal of Experimental Botany* 64: 459–470.

Verdier J, Torres-Jerez I, Wang M *et al.* (2013b) Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *The Plant Journal* 74: 351–362.

Verdier J, Zhao J, Torres-Jerez I *et al.* (2012). MtPAR MYB transcription factor acts as a nonswitch for proanthocyanidin biosynthesis in *Medicago truncatula*. *Proceedings of the National Academy of Sciences USA* 109: 1766–1771.

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ and Ellis NTH (2003) Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. *Molecular Biology and Evolution* 20: 2067-2075.

Vilmorin de P (1913) "Fixite des Races de Froments." IVe Conference Internationale de Genetique – Paris 1911. *Comptes-rendus et Rapport Edite s par Philippe de Vilmorin*. Masson, Paris, France, pp. 312–316.

von Proskowetz E (1890) Welches Werthverhaltniss besteht zwischen den Landrassen landwirthschaftlicher Kulturpflanzen und den sogenannten Zu¨chtungsrassen ? Internationaler land- und forstwirtschaftlicher Congress zu Wien 1890. Section I: Landwirthschaft. Subsection: Pflanzenbau. Frage 5. Heft 13:3–18.

Waines JG (1975) The Biosystematics and Domestication of Peas (*Pisum* L.). *Bulletin of the Torrey Botanical Club* 102: 385–395.

Wang A, Krishnaswamy S (2012) Eukaryotic translation initiation factor 4E mediated recessive resistance to plant viruses and its utility in crop improvement. *Molecular Plant Pathology* 13: 795–803.

Wang ET, Martínez-Romero J, Martínez-Romero E (1999) Genetic diversity of rhizobia from *Leucaena leucocephala* nodules in Mexican soils. *Molecular Ecology* 8: 711–724.

Wang H.L, Grusak MA (2005) Structure and development of *Medicago truncatula* pod wall and seed coat. *Annals of Botany* 95: 737–747.

Wang J, Chu S, Zhang H, Zhu Y, Cheng H, Yua D (2016) Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Scientific Reports* 6: 20728.

Wang Z, Luo, Y, Li X *et al.* (2008) Genetic control of floral zygomorphy in pea (*Pisum sativum* L.) *Proceedings of the National Academy of Sciences USA* 105: 10414–10419.

Warkentin T, Smykal P, Coyne CJ *et al.* (2015) Pea. In: De Ron AM, ed. *Grain Legumes. Handook of Plant Breeding, Vo. 10*. New York: Springer, 37-85.

Warkentin TD, Jordan MC, Hobbs LA (1992) Effect of promoter-leader sequences on transient reporter gene expression in particle bombardment pea (*Pisum sativum* L.) tissues. *Plant Sciences* 87: 171–177.

Warschefsky E, Penmetsa RV, Cook DR, von Wettberg EJ (2014) Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American Journal of Botany* 101: 1791-1800.

Weeden NF (2007) Genetic changes accompanying the domestication of *Pisum sativum*: Is there a common genetic basis to the 'domestication syndrome' for legumes? *Annals of Botany* 100: 1017-1025.

Weeden NF and Wolko B (2001) Allozyme analysis of *Pisum sativum* spp. *abyssinicum* and the development of genotypic definition for this subspecies. *Pisum Genetics* 33: 21–25.

Weeden, NF, Wolko B (2001) Allozyme analysis of *Pisum sativum* spp. *abyssinicum* and the development of genotypic definition for this subspecies. *Pisum Genetics* 33: 21-25.

Weigelt K, Kuster H, Radchuk R *et al.* (2008) Increasing amino acid supply in pea embryos reveals specific interactions of N and C metabolism, and highlights the importance of mitochondrial metabolism. *The Plant Journal* 55: 909-926.

Weigelt K, Küster H, Rutten T *et al.* (2009) ADP-glucose pyrophosphorylase-deficient pea embryos reveal specific transcriptional and metabolic changes of carbon-nitrogen metabolism and stress responses. *Plant Physiology* 149: 395-411.

Weiss E, Kislev ME, Hartmann A (2006) Autonomous Cultivation Before Domestication. *Science* 312: 1608-1610.

Weiss E, Wetterstrom W, Nadel D, Bar-Yosef O (2004) The broad spectrum revisited: Evidence from plant remains. *Proceedings of the National Academy of Sciences USA* 101: 9551-9555.

Weitbrecht K, Kerstin Muller K, Leubner-Metzger G (2011) First off the mark: early seed germination. *Journal of Experimental Botany* 62: 3289-309.

Weller JL, Hecht V, Liew LC. *et al,* (2009) Update on the genetic control of flowering in garden pea. *Journal of Experimental Botany* 60: 2493–2499.

Weller JL, Liew LC, Hecht VFG *et al.* (2012) A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proceedings of the National Academy of Sciences USA* 109: 21158–21163.

Werker E, Marbach I, Mayer AM (1979) Relation between the anatomy of the testa, water permeability and the presence of phenolics in the genus *Pisum*. *Annals of Botany* 43: 765–771.

Westphal E (1974) Pulses in Ethiopia, their taxonomy and agri-cultural significance. *Versl Landbouwkundl Onderzoek*, Wageningen, The Netherlands

Willcox G (2005) The distribution, natural habitats and availability of wild cereals in relation to their domestication in the Near East: multiple events, multiple centres. *Vegetation History and Archaeobotany* 14: 534-541.

Wilson VE and AG (1972) Natural crossing in Lens esculenta Moench. *J. Soc. Hortic. Sci.* 97: 142–143.

Wojciechowski MF, Lavin M and Sanderson MJ (2004) A phylogeny of Legumes (*Leguminosae*) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846 – 1862.

Yarnell SH (1962) Cytogenetics of the vegetable crops III. Legumes. A. Garden peas, *Pisum sativum* L. *Botanical Review* 28: 465–573.

Young JPW and Matthews P (1982) A distinct class of peas (*Pisum sativum* L.) from Afghanistan that show strain specificity for symbiotic *Rhizobium*. *Heredity* 48: 203—210.

Zahran HH (1999) *Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews* 63: 968-989.

Zamir D (2001) Improving plant breeding with exotic genetic libraries. *Nature Review Genetics* 2: 983-989.

Zaytseva O, Gunbin KV, Mglinets AV, Kosterin OE (2015) Divergence and population traits in evolution of the genus *Pisum* L. as reconstructed using genes of two histone H1 subtypes showing different phylogenetic resolution. *Gene* 556: 235-244.

Zhang RF, Zhang FX, Zhang MW, Wei ZC, Yang CY, Zhang Y, Tang XJ, Deng YY, Chi JW. (2011) Phenolic composition and antioxidant activity in seed coats of 60 chinese black soybean (*Glycine max* L. Merr.) varieties. *Journal of Agricultural and Food Chemistry* 59: 5935-5944.

Zhang S, Shi Y, Cheng N, Du H, Fan W, Wang C (2015) De Novo characterization of fall dormant and nondormant alfalfa (*Medicago sativa* L.) leaf transcriptome and identification of candidate genes related to fall dormancy. *PLoS ONE* 10: e0122170.

Zhou Z, Yu Jiang Y, Wang Z, Gou Z *et al.* (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* 33: 408-416.

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *The Plant Genome* 1: 5–20.

Zhukov VA, Zhernakov AI, Kulaeva OA, Ershov NI, Borisov AY, Tikhonovich IA (2015) De Novo Assembly of the Pea (*Pisum sativum* L.) Nodule Transcriptome. *International Journal of Genomics* 695947.

Zlatković B, Mikić A, Smýkal P (2010) Distribution and new records of *Pisum sativum* subsp. *elatius* in Serbia. *Pisum Genetics* 42: 15-18.

Zohary D and Hopf M (2000) *Domestication of plants in the Old World.* Oxford: Oxford University Press.

Zong X, Guan JP, Wang SM, Liu Q, Redden R, Ford R (2008) Genetic diversity and core collection of alien *Pisum sativum* L. germplasm. *Acta Agronomy Sinica* 34: 1518–1528.

Zong X, Redden RJ, Liu Q *et al.* (2009a) Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local collection with microsatellite markers. *Theoretical and Applied Genetics* 118: 193–204.

Zong XX, Ford R, Redden RR, Guan JP and Wang SM (2009b) Identification and analysis of genetic diversity structure within *Pisum* genus based on microsatellite markers. *Agricultural Sciences in China* 9: 591-600.