Project proposal:

Towards comprehensive pea germplasm management for future use.

Applicant: ***Petr Smýkal ,Agritec Plant Research, s.r.o.***

1. Problem statement

Core collections of pea (*Pisum sativum* L.) are frequently established on national level, using various morphological and agronomical characters and molecular marker types. Since both markers and reference set of accessions are variable among different collections, direct comparison and compatibility of generated data are limited. Based on our previous work we would like to further test suitability of distance-based and Bayesian model-based methods for core collection establishment and data integration across different collections together with further utilisation for association mapping.

2. Justification and rationale

Both Czech and UK collections are well documented and characterized both on molecular and phenotypic levels. All partners are collaborating on evaluation and comparison of pea data sets. Current results and possibility of further testing of analytical methods of core collection establishment and data compatibility will permit better availability of germplasm accessions and their related information to all AEGIS users.

Selection criteria and methods determining in this project will aid the selection of accessions for the AEGIS project will be identified and made available to other curators of pea (and others autogamous grain legume crop species) collections.

3. Background

It is widely recognized that the genetic diversity of cultivated plants has narrowed as a result of domestication and connected bottlenecks. To avoid a permanent loss of diversity, conservation of plant genetic resources in form of *ex situ* collections were established. In order to facilitate their management, a core collection concept was developed (Frankel and Brown 1984). Knowledge of variety genetic basis, especially to see if it has not become too narrow to render crops more vulnerable to diseases or pests is important for breeding and additionally accessions most distinct from others are likely to contain the greatest number of novel alleles.

Traditionally, classifications are made by morphological descriptors, until now the only legitimate marker type accepted by the International union for the protection of new varieties of plants (UPOV), together with known pedigree and passport data. Moreover, in recent years genetic structures of major pea germplasm collections have been investigated by several molecular marker approaches. 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. 2008). A potentially existing but largely neglected problem in using SSRs for characterizing highly diverse germplasm, is homoplasy and the possibility of back-mutations exhibited by this marker type. Alternately, marker systems based on retrotransposon insertion polymorphism 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 (Smýkal et al. 2009). 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). This trend can be expected to further proliferate in line with rapid advances in high throughput single-nucleotide polymorphism (SNP) generation and detection assays. 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. Initially and still largely used 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 et al. 2007) make them more attractive and powerful. Altogether, a large amount of polymorphic data points have been produced for each collection, which were subsequently subjected both to genetic distance analysis and/or model-based Bayesian diversity analysis. However, after data processing, further use of such data is highly limited, especially in the absence of cross-comparison between collections. Furthermore, most of these accessions have been evaluated for morphological, agronomical and phytopathological traits, thus the data has enormous scientific and breeding potential.

Altogether, a large amount of polymorphic data points have been produced for each collection, which were subsequently subjected both to diversity analysis. In recent times an international consortium (PeaGRIC) has been formed to try to take on a coordinating function from the international Pisum community (Furman et al. 2006, Smýkal et al. 2009). Among the objectives of this consortium are stated combining of available data sets into a virtual global collection and the development of a dispersed international reference collection. It is time right for the establishment of virtual, pea world core collection combining suitable molecular platforms with robust morphological parameters to address population structure and allow better cross-comparison of results (Smýkal et al. 2009). As proposed, such a set would provide a useful and powerful resource for next marker generation such as single nucleotide polymorphims (SNPs) and importantly for phenotypic analysis of agronomic traits, both as toolkits for association mapping, as a strategy to gain insight on genes, genomic regions underlying desired traits (Zhu et al. 2008). One of the prerequisite is a collection of accessions with wide coverage of genetic diversity, recording the phenotypic characteristics and finally genome-wide genotyping (Zhu et al. 2008). One very important issue is the deposition and availability of both molecular and agronomical, morphological traits data. So far data held at national level are not broadly accessible. Importantly, GERMINATE(http://bioinf.scri.ac.uk/
germinate_pea//app/index.pl) database (Lee et al. 2005) provides original marker scores and allow thus searchable interphase.

4. Main objective and specific objectives The objective of the project is to assist in determining the selection of accessions for registration into AEGIS, to further test analytical methods of core collection establishment and to establish data compatibility of pea core collections for AEGIS .

Activity 1: Morphological data evaluation
Activity 2: Molecular data generation
Activity 3: Molecular data analysis
Activity 4: Molecular and morphological data integration
Activity 5: Data deposition into database
Activity 6: Data computing, core set

5. Materials and methods

Pisum germplasm collection kept in the applicant´s workplace includes recently 1307 accessions of field grain peas (*Pisum sativum L. ssp. sativum*)(79%) and 21% fodder peas (*Pisum sativum L. ssp. arvense*). The collection is desribed by the Descriptor List of Genus Pisum (Pavelková et al. 1986), passport part (19 descriptors, 179 traits) and (2) evaluation part (115 descriptors, 129 traits). The data are deposited in IS EVIGEZ database. The combination of the passport, pedigree, morphological descriptors and molecular markers, serve for the rationalisation of genetic resources in the frame of the AEGIS programme.

We will use the subset (164 accessions) of Czech national pea collection with focus on Czech/Czechoslovak accessions bred over last 70 years, which has been previously integrated into EU TEGERM and GLIP projects, together with selected world-wide reference accessions. In addition to already generated morphological and molecular data (Smýkal et al. 2008), these samples will be analyzed by additional RBIP, SSR and gene specific SNP loci (as *sbm1-2*/eIF4E gene etc.) to meet compatibility with other germplasm data. We will use established protocols (Corander et al. 2007, Smýkal et al. 2008 and submitted) both for molecular and morphological data generation and subsequent mathematical analysis. Moreover, relevant data are already available for entire collection as well as reasonably compatible data from JIC pea collection analyzed within the frame of previous TEGERM and GLIP EU projects.

6. Expected outputs

The output of the project will rely on generation of methodologies for core collections establishment, as a prerequisite for association mapping strategies aimed at trait/marker identification using integration of phenotypic and genotyping data.

Moreover list of accessions together with molecular and morphological descriptors will be provided to inform the development and population of AEGIS database in conjunction with the ECPGR working group for Grain Legumes.

7. Benefits and impact

The expected benefits of the proposal are both for genetic resources conservation and use, as method of molecular and morphological data processing will be evaluated and established, as well as for breeders community as maximally diverse germplasm will be better described. Removal of duplicates and core set establishment will increase effectiveness and cost-efficiency of genetic resources maintenance.

8. Innovation

Project proposes innovative approaches for germplasm data compatibility, such as integration of morphological and molecular data types, using the current progress in mathematical analysis of genetic diversity.

9. Application of results

List of pea accessions of Czech origin for AEGIS database purposes will be available. Selection criteria that will determine the selection of accession into the AEGIS will be identified.

10. Workplan

Dr. Petr Smykal, Agritec Plant Research, s.r.o., Šumperk, Czech Rep. workplan

Main activities	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Activity 1												
Activity 2	■	■	■									
Activity 3			■	■	■	■						
Activity 4			■	■	■	■						
Activity 5												
Activity 6							■	■	■	■	■	■

Dr. Miroslav Hybl, Agritec Plant Research, s.r.o., Šumperk, Czech Rep. workplan

Main activities	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Activity 1	■	■	■	■	■							
Activity 2												
Activity 3												
Activity 4			■	■	■	■						
Activity 5												
Activity 6							■	■	■	■	■	■

Dr. Andrew Flavell, University of Dundee at SCRI, Inwergovrie, Dundee, UK workplan

Main activities	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Activity 1												
Activity 2												
Activity 3												
Activity 4												
Activity 5						■	■	■	■	■	■	■
Activity 6												

Dr. Mike Ambrose, John Innes Centre, Norwich, UK workplan

Main activities	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Activity 1												
Activity 2												
Activity 3												
Activity 4												
Activity 5						▨	▨	▨	▨	▨	▨	▨
Activity 6												

Prof. Jukka Corander, Abo University, Finland workplan

Main activities	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Activity 1												
Activity 2												
Activity 3												
Activity 4												
Activity 5												
Activity 6						▨	▨	▨	▨	▨	▨	▨

11. Budget

Agritec Plant Research, s.r.o., Šumperk, Czech Rep.

Staff time	1,600 Eur
Travel	2,000 Eur
Meetings	1,000 Eur
Supplies and/or services	4,000 Eur
Total	**8,600 Eur**
Funding from other sources	4,300 Eur
Amount requested from the Competive AEGIS Grant Scheme	**4,300 Eur**

University of Dundee at SCRI, Inwergovrie, Dundee, UK

Staff time	1,600 Eur
Travel	1,000 Eur
Meetings	700 Eur
Supplies and/or services	800 Eur
Total	**4,100 Eur**
Funding from other sources	2,050 Eur
Amount requested from the Competive AEGIS Grant Scheme	**2,050 Eur**

Crop Genetics, John Innes Centre, Norwich, UK

Staff time	1,600 Eur
Travel	1,000 Eur
Meetings	700 Eur
Supplies and/or services	800 Eur
Total	**4,100 Eur**
Funding from other sources	2,050 Eur
Amount requested from the Competive AEGIS Grant Scheme	**2,050 Eur**

Abo University, Finland

Staff time	1,600 Eur
Travel	1,000 Eur
Meetings	700 Eur
Supplies and/or services	800 Eur
Total	**4,100 Eur**
Funding from other sources	2,050 Eur
Amount requested from the Competive AEGIS Grant Scheme	**2,050 Eur**

12. Contributions foreseen by applicant

Agritec Plant Research, s.r.o., Šumperk, Czech Rep.
50 % of budget will be cover from Czech National Programme for Plant Genetic Resources supported by Czech Ministry of Agriculture.

University of Dundee at SCRI, Inwergovrie, Dundee,
50 % of budget will be obtained from core support from the University of Dundee comprising staff time (AJF, IT support staff, technical support staff) and basic consumables (plasticware, lab media and reagents).

Abo University, Finland
50 % of budget will be obtained from core support from the Abo University comprising staff and computational time .

13. Bibiography:

Baranger A. et al. (2004) Genetic diversity within Pisum sativum using protein and PCR-based markers. Theor Appl Genet 108: 1309-1321.

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

Corander J. et al. (2007) Random partition models and exchangeability for Bayesian identification of population structure. Bull Mathem Biol 69: 797-815.

Furman et al. (2006) Formation of PeaGRIC: An international consortium to co-ordinate and utilize the genetic diversity and agro ecological distribution of major collections of Pisum. Pisum Genetics 38:32-34.

Jing R.C. et al. (2005) Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in Pisum species. Genetics 171: 741-752.

Lee M.J. et al. (2005) GERMINATE. A Generic Database for Integrating Genotypic and Phenotypic Information for Plant Genetic Resource Collections. Plant Physiol. 139: 619–631.

Loridon K. et al. (2005) Microsatellite marker polymorphism and mapping in pea (Pisum sativum L.) Theor Appl Genet 111:1022-1031.

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

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

Smýkal P. et al. (2009) Effort towards a world pea (Pisum sativum L.) germplasm core collection: The case for common markers and data compatibility. Pisum Genetics 40: 11-15.

Zhu C., Gore M., Buckler E.S., Yu J. (2008) Status and prospects of association mapping in plants. The Pl. Genome 1: 5-20.