Effort towards a world pea *(Pisum sativum* L.) germplasm core collection: The case for common markers and data compatibility

Smykal, P.[1] [1]Agritee Plant Research Ltd., Sumperk, Czech Republic
Coyne, C.J.[2] [2]USDA-ARS, Pullman, WA, USA
Ford, R.[3] [3]BioMarka, The University of Melbourne, Australia
Redden, R.[4] [4]Australian Temperate Field Crops Collection, Horsham, Victoria Australia
Flavell, A.J.[5] [5]University of Dundee at SCRI, Dundee, UK
Hybl, M.[1] [1]Agritec Plant Research Ltd., Sumperk, Czech Republic
Warkentin, T.[6], [6]University of Saskatchewan, Saskatoon, Canada
Burstin, J.[7] and Due, G.[7] [7]INRA URLEG, Dijon, France
Ambrose, M.[8] and Ellis, T.H.N.[8] [8]John Innes Centre, Norwich, UK

It is widely recognized that the genetic diversity of cultivated plants has narrowed as a result of thousands of years of domestication and associated bottlenecks. To avoid a permanent loss of diversity, conservation of plant genetic resources in the form of ex situ collections was pioneered by N.I. Vavilov (1). This activity has resulted in several million crop accessions being held in several hundred germplasm collections, including gene banks (2). In the case of pea (Pisum), the genus on which modern genetics was founded, several large germplasm collections are maintained worldwide (Table 1). More than ever, we are aware of the danger of diversity loss linked to cultivation of a limited number of high-yielding varieties from a small number of locations. Moreover, we are aware of the benefits resulting from the exploitation of older varieties, landraces and even wild crop relatives for breeding new varieties to cope with environmental and demographic changes (3). Consequently, within the last two decades the study of genetic diversity for both germplasm management and breeding has received much attention.

Table 1: ***Fx situ*** **nprmnlasm collections of** ***Pisum*** **with holdings in excess of 1000 accessions.**

Code	Country	Institute	pumber offF A ccessions	Web site
VIR	**Russia**	N.I. Vavilov Research Institute of Plant Industry, St. Petersburg	**6790**	http://www.vir.nw.ru/data
BAR	**Italy**	Istituto del Germoplasma, Bari	**4297**	http://www.ba.cnr.it/areagg34/germoplasma
SAD	**Bulgaria**	Institute of Plant Introduction and Genetic Resources, Sadovo	**2787**	http://www.genebank.hit.bg
NGB	**Sweden**	NordGen, Nordic Genetic Resource Centre, Alnarp	**2724**	http://www.ngb.se/sesto
CGN	**The Netherlands**	Centre for Genetic Resources, Wageningen	**1008**	http://www.cgn.wur.nl/pgr/
ATFC	**Australia**	Australian Temperate Field Crop Collection, Horsham	**6567**	http://www2.dpi.qld.gov.au/extra/asp/AusPGRIS
ICARDA	**Syria**	International Center for Agricultural Research in the Dry Areas, Aleppo	**6105**	http://www.icarda.cgiar.org
GAT	**Germany**	Leibniz Institute of Plant Genetics and Crop Plant Research, Gaterleben	**5336**	http://fox-serv.ipk-gatersleben.de/
ICAR	**China**	Institute of Crop Sciences, CAAS China	**3837**	http://icgr.caas.net.cn/cgris
USDA	**USA**	Plant Germplasm Introduction and Testing Research Station, Pullman	**3710**	http://www.ars-grin.gov
JIC	**UK**	John Innes Centre, Norwich	**3194**	http://www.jic.ac.uk/GERMPLAS/pisum
WTD	**Polland**	Plant Breeding and Acclimatization Institute Blonie, Radzikow	**2899**	http://www.ihar.edu.pl/gene_bank/
INRA	**France**	INRA, Station de Genetique et d'Amelioration des Plantes, Dijon	**1891**	http://www.dijon.inra.fr
UKR	**Ukraine**	Yurjev Institute of Plant Breeding, Kharkov	**1671**	http://www.bionet.nsc.ru
CZE	**Czech Republic**	AGRITEC, Research, Breeding and Services Ltd., Sumperk	**1273**	http://genbank.vurv.cz/genetic/resources
HUN	**Hungary**	Institute for Agrobotany, Tapioszele	**1188**	http://www.rcat.hu

Thus, in order to facilitate *Pisum sp.* germplasm management and increased efficiency of use, a core collection is being developed under the concept proposed by Frankel and Brown (4). Also, for breeding it is important to know the genetic background of cultivars, especially whether they have become too narrow in diversity to render crops more vulnerable to diseases or pests. Accessions genetically distinct from others are likely to contain the greatest number of novel alleles which can be exploited in breeding.

Germplasm accessions are classified based on known pedigree, passport data and morphological descriptors which are currently the only marker type accepted by the International Union for the

Protection of New Varieties of plants (UPOV). However, in recent years the genetic structure of major pea germplasm collections have been investigated by several molecular marker platforms, including microsatellite and retrotransposon-based markers. In particular, Simple Sequence Repeats (SSRs or microsatellites) have been popular for assessing *Pisum* genetic diversity because of their high polymorphism and information content, co-dominance and reproducibility (5, 6, 7, 8). A potential, but largely neglected, problem using SSRs to characterize highly diverse germplasm, is size homoplasy and the possibility of back-mutation exhibited by this marker type (9).

Alternately, marker systems based on retrotransposon insertion polymorphism have been extensively used for phylogeny and genetic relationship studies in pea. Retrotransposon-based insertional polymorphism (RBIP) based on presence and absence of specific insertions provides a highly specific, reproducible and easily scorable method suitable for deeper phylogeny and diverse germplasm studies (10, 11). High copy retroelements were successfully applied as markers in the multiplex IRAP-PCR (Inter-Retrotransposon Amplified Polymorphism) format (12) suitable for fast variety fingerprinting.

Using these markers several major world pea germplasm collections have been analyzed and core collections formed. Collections which have been characterized include, 1) 1200 pea accessions of Chinese origin contained within a larger set of over 2000 accessions were analyzed by 21 SSR loci (13), 2) 310 of 5394 USDA pea germplasm accessions were assessed with 37 RAPD and 15 SSR markers (14), and 3) INRA France used an extensive set of 121 protein and SSR markers to genotype 148 accessions (5, 7, 8). In addition, a pea collection (~100 accessions) held by the Crop Development Center in Saskatoon, Saskatchewan, Canada was studied using RAPD, ISSR and SSR (15) and the entire JIC pea germplasm (~ 3,500 accessions), comprised largely of expedition collections, was analyzed using 45 RBIP markers (Jing *et al.* in preparation). Finally, over 1,400 pea accessions held by the Czech National Pea Germplasm collection were genotyped using a combination of RBIP and SSRs (16). This latter study has shown that both SSRs and RBIPs have high information content and offer comparable diversity measurements. This is an important finding since SSRs, in spite of having multiple alleles, are more difficult to transfer between labs and RBIPs are more transferable. Although SSR and RBIP marker types are widely used, their potential is limited. Advances in model legume sequencing and increased knowledge of the legume genome there has been a shift to gene-based markers in pea (11). This trend is expected to continue with rapid advances in high throughput single-nucleotide polymorphism (SNP) detection assays based on next generation sequencing technologies (17). These provide powerful platforms with the potential for rapid genomic characterization of thousands of diverse pea genotypes provided adequate resources and support are available.

Improvements in marker methods have been accompanied by refinements in computational methods to convert original raw data into useful representations of diversity and genetic structure. The initially and still largely used distance-based methods (18) have been challenged by model-based Bayesian approaches. The incorporation of probability, measures of support, ability to accommodate complex models and various data types (19, 20, 21, 22) make Bayesian approaches more attractive and powerful.

A large body of molecular data has been produced for germplasm collections and has been subsequently subjected both to genetic distance analysis and/or model-based Bayesian diversity analyses. 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, agronomic and phytopathological traits giving the data added value to the scientific and breeding communities.

Pisum ranks fourth among the world's most important grain legumes yet does not feature as a mandate crop within the CGIAR system. In recent times an international consortium (PeaGRIC) was formed to coordinate the international *Pisum* research community (23). Among the objectives of this consortium is the combining of available data sets into a virtual global collection and the development of a dispersed international reference collection. We feel the time is right for the establishment of a virtual, pea world core collection combining suitable molecular platforms with robust morphological parameters to address population structure and allow better cross-comparison of results. Existing examples of such worldwide core collections include the 372 wheat accessions chosen from about 4000 which were

validated to represent sufficient genetic diversity of the crop (24). In these lines we have already initiated RBIP markers analysis of Chinese-Mongolian origin pea accessions from core set (13), which will provide compatibility to JIC and Agritec germplasm data. As proposed, such a collection would provide a useful and powerful resource for the next generation markers such as single nucleotide polymorphims (SNPs) and for phenotypic analysis of agronomic traits. These would act as toolkits for association mapping, a strategy to gain insight on genes and genomic regions underlying desired traits (25). Compared to conventional linkage-mapping, based on time-consuming mapping population development; linkage disequilibrium (LD)-mapping, using the non-random associations of loci in haplotypes, is a powerful high-resolution mapping tool even for complex quantitative traits. In contrast to biparental crosses, the higher resolution and the possibility of historical trait data exploitation indicate that this approach has enormous potential in crop breeding and genetics. The prerequisites include a collection of accessions with a wide coverage of the available and existing genetic diversity, recording the phenotypic characteristics and finally genome-wide genotyping (25).

One very important issue is the deposition and availability of molecular, agronomical, and morphological trait data. So far, data held at the national level has not been broadly accessible. Although, the European EURISCO Web catalogue maintained by Biodiversity International and the USDA National Plant Germplasm System provide information of around two million accessions, this information is largely passport-based, thus limited. There is an attempt to develop and use database systems that will bring together passport, morphological and genotype data that will improve both germplasm management as well as enable data exploration across a wide range of data types. The example of such deposition can be seen at PANZEA (http://www.panzea.org) for maize or SoyBase (http://soybase.org) for soybean. In the case of the SoyBase, a Breeders Toolbox was already developed, providing a genetic map along with QTLs for related traits. Importantly, the GERMINATE (http://bioirf.scri.ac.uk/germinate_pea//app/index.pl) database (26) provides original RBIP-marker scores for the entire JIC pea collection with an interactive search interface.

Defining a pea core and a set of markers (SSR, RBIP and for the future - SNPs) can provide a basis for comparison of phenotypic and molecular analyses when only a part of this core and or identified markers are used, i.e. modern statistical procedures allow combination of data from partial inclusion from a core and partial inclusion of common markers. In this case many different studies can be combined and each new study incrementally adds to the virtual world database with participation of many different scientists worldwide. Phenotypic data from different sources can be converted to standard normal variates or converted into 1 - 9 scales for quantitative data to enable search engines to compare data from many sources. The idea is to provide an initial selection of genotypes / landraces for more detailed research / validation. Immediate steps should include an analysis of common reference accessions across the major diversity datasets and adoption of a common set of markers applied across the developed pea core sets, a common set of check cultivars for future genotyping, together with the exchange and deposition of both molecular and agronomic data. This will enable the creation of a virtual worldwide pea germplasm resource for common benefit.

Acknowledgements: The work on Czech pea germplasm was supported by the Ministry of Education of the Czech Republic (MSM 2678424601) project (2004-2010).

1. Vavilov N.I. 1926. Bull. Appl. Bot. 26. Leningrad, USSR.
2. Laliberte, B. and Fowler, C. 2006. Acta Horticulturae 760:1-43.
3. Esquinas-Alacazar, J. 2005. Nature Rev. Genet. 6: 946-953.
4. Frankel, O.H. and Brown, A.H.D. 1984. *In* Genetics: New Frontiers, Vol. IV, Oxford and IBH Publ. Co., New Delhi, India
5. Burstin, J., Deniot, G., Potier, J., Weinachter, C., Aubert, G., Baranger, A. 2001. Plant Breed. 120:311¬ 317.
6. Ford, R., LeRoux, K., Itman, C., Brouwer, J.B., Tayler, P.W.J. 2002. Euphytica 124:397-405.
7. Baranger, A., Aubrey, G., Aran, G., Laine, A.L., Deniot, G., Potier, J., Burstin, J. 2004. Theor. Appl. Genet. 108:1309-1321.
8. Loridon, K., McPhee, K., Morin, J., Dubreuil, P., Pilet-Nayel, M.L., Aubert, G., Rameau, C., Baranger, A., Coyne, C., Lejeune-Henaut, I., Burstin, J. 2005. Theor. Appl. Genet. 111:1022-1031.
9. Mogg, R., Batley, J., Hanley, S., Edwards, D., O'Sullivan, H., Edwards, K.J. 2002. Theor. Appl. Genet. 105:532-543
10. Jing, R.C., Knox, M.R., Lee, J.M., Vershinin, A.V., Ambrose, M., Ellis, T.H.N., Flavell, A.J. 2005. Genetics 171:741-752.
11. Jing, R.C., Johnson, R., Seres, A., Kiss, G., Ambrose, M.J., Knox, M.R., Ellis, T.H.N., Flavell, A.J. 2007. Genetics 177:2263-2275.
12. Smykal, P., Horacek, J., Dostalova, R., Hybl, M. 2008. J. Appl. Genet. 49:155-166.
13. Zong, X., Redden, R.J., Liu, Q., Wang, S., Guan, J., Liu, J., Xu, Y., Liu, X., Gu, J., Yan, L., Ades, P., Ford, R. 2009. Theor. Appl. Genet. 118(2):193-204
14. Brown, A., Coyne, C. 2005. *In* ASA-CSSA-SSSA Annual Meetings Abstracts [CD-ROM], p. 8563.
15. Tar'an, B., Zhang, C., Wankertin, T., Tullu, A., Vandenberg, A. 2005. Genome 48:257-272.
16. Smykal, P., Hybl, M., Corander, J., Jarkovsky, J., Flavell, A.J., Griga, M. 2008. Theor. Appl. Genet. 117:413-424.
17. Rogers, Y.-H. and Venter, J.C. 2005. Nature 437:326-327.
18. Reif, J.C., Melchinger, A.E., Frish, M. 2005. Crop Sci. 45:1-7.
19. Beaumont, M.A., Rannala, B. 2004. Nature Rev. 5:251-261.
20. Corander, J., Gyllenberg, M., Koski, T. 2007. Bull Mathem Biol 69:797-815.
21. Coyne, C.J., Brown, A., Timmerman-Vaughan, G.M., McPhee, K.E., Grusak, M.A. 2005. Pisum Genetics 37:1-4.
22. Falush, D., Stephens, M., Pritchard, J.K. 2003. Genetics 164:1567-1587.
23. Furman, B.J. Ambrose, M. Coyne, C.J. and Redden, B. 2006. Pisum Genetics 38:32-34.
24. Balfourier, F., Roussel, V., Strelchenko, P., Exbrayat-Vinson, F., Sourdille, P., Boutet, G., Koenig, J., Ravel, C., Mitrofanova, O., Beckert, M., Charmet, G. 2007. Theor. Appl. Genet. 114:1265-1275.
25. Zhu, C., Gore, M., Buckler, E.S., Yu, J. 2008. Plant Gen. 1:5-20.
26. Lee, M.J., Davenport, G.F., Marshall, D., Ellis, T.H.N., Ambrose, M.J., Dicks, J., van Hintum, T.J.L., Flavell, A.J. 2005. Plant Physiol. 139:619-631.