The critical role of international collaborations to improve <u>conservation</u> and <u>utilization</u> of crop collections

Gayle Volk – USDA Fort Collins, Colorado, USA Clare Coyne – USDA Pullman, Washington state USA National Plant Germplasm System stations





The USDA National Plant Germplasm System (NPGS): USDA/ARS and university partners

597,400 accessions



Germplasm Distribution



- DISTRIBUTION OF NPGS GERMPLASM 2016-2020:
- 1.2 MILLION SAMPLES DISTRIBUTED TO FULFILL MORE THAN 55,000 GERMPLASM ORDERS. REQUESTORS SELECTED AN AVERAGE OF CA. 22 ACCESSIONS PER ORDER







Cool season food legumes

- Increasing the utilization of plant genetic resources
 - Empowering the global user community
- Current approaches
 - Phenotyping
 - Genotyping
 - High through-put phenotyping
 - Artificial intelligence: Machine Learning
- Future: Databases
 - Interoperability is key (BrAPI)

Utilization of PGR using GWAS and Genomic Selection

Graphic published by Esposito, S., Carputo, D., Cardi, T. and Tripodi, P., 2020. Applications and Trends of Machine Learning in Genomics and Phenomics for Next-Generation Breeding. *Plants*, *9*(1), p.34.





Genebanks are Centers of Innovations. Paraphrased from Dr. Peter Bretting, USDA

- <u>23,700</u> pea accessions packets shipped 2018-present
- First pea reference genome published (Kreplak et al., 2019)
- 163,902 SNP set available for pea core collection (Holdsworth et al., 2017)
 - Now mapped to the reference genome

Impact: High quality seed and SNP genotypes were provided to plant science researchers







GWAS – USDA Pea Core Collection







Preliminary data: year 1

Chromosome	SNP	Р	R ²
1	S1LG6_221982693	1.96E-08	0.064
4	S4LG4_93156666	1.16E-10	0.081
4	S4LG4_73177262	2.66E-08	0.063
6	S6LG2_235649991	2.23E-08	0.063



Increasing utilization: empowerment

- User-friendly SNP data sets easily downloadable from GRIN-Global or from the USDA National Agriculture Library or PulseDB.org
- Association mapping panel seed available from GRIN-Global

OPEN

Citation: Horticulture Research (2017) **4**, 17017; doi:10.1038/hortres.2017.17
www.nature.com/hortres

ARTICLE

A community resource for exploring and utilizing genetic diversity in the USDA pea single plant plus collection

William L. Holdsworth^{1,6}, Elodie Gazave¹, Peng Cheng², James R. Myers³, Michael A. Gore¹, Clarice J. Coyne^{2,4}, Rebecca J. McGee⁵ and Michael Mazourek¹



PEA.PROTEIN.2003.COYNE

← Previous page

Evaluation location: Washington, United States

Selected seeds of a single seed phenotype within each accession used previously in the mineral nutrient study of the pea core collection (Grusak, 2003, GRIN) were used for this analysis. Protein concentration (N H 6.5) of the ground seed was determined using a Leco nitrogen analyzer (Leco Corp., St Joseph, MI, USA). click for an Excel(.xls) spread sheet of Dr. Clare Coyne's protein data.

Show/hide col Showing 1 to 1	lumns Show 10 rows Excel CSV 10 of 479 entries	Download	Download EXCEL or CVS fi		
crop	method_name	🔶 taxon	♦ acp		
PEA	PEA.PROTEIN.2003.COYNE	Pisum abyssinicum	PI		
PEA	PEA.PROTEIN.2003.COYNE	Pisum abyssinicum	PI		
🛨 PEA	PEA.PROTEIN.2003.COYNE	Pisum abyssinicum	PI		

Empower: Phenotypic data download

Empower: Genotypic data







Search Accessions GRIN-Global (ars-grin.gov)

Maize Genetics | TASSEL

ICARDA Lentil Reference Collection Plus

- Genotyping GBS, 4195 SNPs
- Lentil Reference Genome v2.0 (Bett et al, 2021, in prep)
- ICARDA historical data

© The Author(s), 2021. Published by Cambridge University Press on behalf of NIAB ISSN 1479-2621 Plant Genetic Resources: Characterization and Utilization; 1–10 doi:10.1017/S147926212100006X

Genetic diversity and GWAS of agronomic traits using an ICARDA lentil (*Lens culinaris* Medik.) Reference Plus collection

Karthika Rajendran¹† ⁽ⁱ⁾, Clarice J. Coyne²* ⁽ⁱ⁾, Ping Zheng³, Gopesh Saha⁴‡, Dorrie Main³, Nurul Amin⁴§, Yu Ma³, Ted Kisha², Kirstin E. Bett⁵ ⁽ⁱ⁾, Rebecca J. McGee⁶ and Shiv Kumar¹



SAID (SAID

Fig. 2. Dendrogram based on UPGMA and the subpopulations (K = 3) calculated using the Bayesian clustering method of the software STRUCTURE based on SNP data for 176 lentil accessions (Pritchard et al., 2000).

ICARDA historical data

Figure S1. Frequency histogram of the phenotypic data from the historical record of evaluation of ICARDA lentil plant genetic resources.

Note on historic genebank data

+

0

- The data mining of historical genebank phenotypic data for GWAS is relatively new and mostly untested.
- Nguyen and Norton (2020) recently reviewed this approach for GWAS and genomic selection.
- Two examples of this wealth of data on barley and wheat was published by Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) (González *et al.*, 2018a; Philipp *et al.*, 2019).
- González et al., (2018b) published a strategy to utilize historical phenotypic data collected during seed regeneration to assemble large mapping populations of accessions to discover to genetic effects.
- The utilizing phenotypic historic data from *ex situ* genebanks was thought to "elevate them to bio-digital resources centers."

ICARDA Lentil Reference Collection Plus

GWAS results

		Chromoso		
Trait	Marker	me	P value	R ²
Days to	SLCCHR3_1469			
flower	44147	Chr 3	5.22E-05	0.11
Days to	SLCCHR2_8037			
maturity	2417	Chr 2	1.70E-05	0.17
Days to	SLCCHR5_2298			
maturity	89997	Chr 5	9.58E-05	0.17
Days to	SLCCHR6_5125			
maturity	7354	Chr 6	5.54E-05	0.16
Seeds per	SLCCHR2_2708			
pod	63768	Chr 2	9.44E-05	0.13
100 seed	SLCCHR2_1417			
weight	82657	Chr 2	3.73E-05	0.13

GWAS-To-Go Package

- Order 176 lentil accessions from ICARDA
- Download the SNP data file from PulseDB.org
- Collect phenotype(s) of interest
- Conduct GWAS in TASSEL
- Possible anywhere in world with internet

Root rot resistance in lentil

- GWAS
- ICARDA Lentil Reference Plus and USDA Lentil Core Collection
- 326 accessions
- GBS, SNPs

Article

Dissecting the Genetic Architecture of Aphanomyces Root Rot Resistance in Lentil by QTL Mapping and Genome-Wide Association Study

MDPI

Yu Ma¹, Afef Marzougui², Clarice J. Coyne³, Sindhuja Sankaran², Dorrie Main¹, Lyndon D. Porter⁴, Deus Mugabe⁵, Jamin A. Smitchger⁵, Chongyuan Zhang², Md. Nurul Amin⁶, Naser Rasheed⁷, Stephen P. Ficklin¹ and Rebecca J. McGee^{8,*}

Phenotyping

Root rot resistance in lentil

frontiers in Plant Science

ORIGINAL RESEARCH published: 16 April 2019 doi: 10.3389/fpls.2019.00383

Advanced Imaging for Quantitative Evaluation of Aphanomyces Root Rot Resistance in Lentil

Afef Marzougui¹, Yu Ma², Chongyuan Zhang¹, Rebecca J. McGee³, Clarice J. Coyne⁴, Dorrie Main² and Sindhuja Sankaran^{1*}

Root rot resistance in lentil

Seven QTL identified with both traditional scoring and multispectral imaging multiple trait values.

International Journal of Molecular Sciences

Article

Dissecting the Genetic Architecture of Aphanomyces Root Rot Resistance in Lentil by QTL Mapping and **Genome-Wide Association Study**

Yu Ma¹, Afef Marzougui², Clarice J. Coyne³, Sindhuja Sankaran², Dorrie Main¹, Lyndon D. Porter⁴, Deus Mugabe⁵, Jamin A. Smitchger⁵, Chongyuan Zhang², Md. Nurul Amin⁶, Naser Rasheed⁷, Stephen P. Ficklin¹ and Rebecca J. McGee^{8,*} Genomic prediction in plant genetic resources, an efficient strategy to identify useful germplasm.

 Graphic published by Esposito, S., Carputo, D., Cardi, T. and Tripodi, P., 2020.
 Applications and Trends of Machine Learning in Genomics and Phenomics for Next-Generation Breeding.
 Plants, 9(1), p.34.

'Turbocharging' Genebanks

Turbocharging Genebanks through the use of Genomic Selection: Sorghum case study

 Yu, X., Li, X., Guo, T., Zhu, C., Wu, Y., Mitchell, S.E., Roozeboom, K.L., Wang, D., Wang, M.L., Pederson, G.A. and Tesso, T.T., Schnable, Rex Bernardo and Jianming Yu 2016. Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nature Plants*, 2(10), pp.1-7.

Genomic prediction (selection)

- 91 Genomic Selection publications
- Machine Learning algorithms listed on both x and y axis
- No one GP algorithm performs best for all species and traits (next slide)
- Azodi et al. 2019. Benchmarking parametric and Machine Learning models for genomic prediction of complex traits. *G3: Genes, Genomes, Genetics, 9*(11), pp.3691-3702.

ML Models

Predicting plant height across crops: r ranges from a low of 0.08 to a high of 0.63. No one GP algorithm performed best for all species and traits

• Azodi et al. 2019. Benchmarking Parametric and Machine Learning Models for Genomic Prediction of Complex Traits. *G3: Genes, Genomes, Genetics, 9*(11), pp.3691-3702.

Harnessing genetic diversity through genomic prediction in the USDA Pea Germplasm Collection

- These data will facilitate the identification of accessions divergent from mainstream breeding germplasm that could serve as sources of novel, favorable alleles using genonic prediction modeling.
- Trait: pea seed protein concentration
- Machine learning
- Dr. Nonoy Bandillo, NDSU and Clare Coyne. 2020 CGC Evaluation grant.

Pea Core Data: GS to identify new alleles for pea seed protein concentration

- Agronomic features
 - Yield, time to flower, biomass etc (9)
 - Morphology (35)
- Disease reactions
 - Fungal (8)
 - Seed traits
 - Protein, oil, fiber, starch, mineral nutrients (11)
- Root traits (10)
- Abiotic (1)
- Insect (1)
- Nematodes (1)

Pea Phenotypic data in GRIN 193,101

Pea Genotypes VCF file in GRIN, SNPs 163,902

SNPS: Holdsworth, W.L., Gazave, E., Cheng, P., Myers, J.R., Gore, M.A., Coyne, C.J., McGee, R.J. and Mazourek, M., 2017. A community resource for exploring and utilizing genetic diversity in the USDA pea single plant plus collection. *Horticulture Research*, *4*(1), pp.1-8.

Genomic prediction in USDA pea PGR

- 1) How many individuals are needed to build a suitable "training population" for performing genomic selection?
- 2) What is the best statistical model for predicting genetic value?
- 3) How many markers are needed for genomic selection in a typical pea breeding program?
- 4) Can markers be computationally imputed to save even more on genotyping costs?
- 5) Do subsets of individuals exist that create a more accurate model than the entire population of individuals available? Or is it always advisable to use the maximum number of individuals available?
- 6) How does the relationship between selection candidate (i.e., individual being predicted) and training population affect prediction accuracy?

Md. Abdullah Al Bari, Hannah Worral, Stephen Szwiec, Yu Ma, Ping Zheng, Doreen Main, Clarice Coyne, Rebecca McGee and Nonoy Bandillo. 202x. Harnessing genetic diversity in the USDA pea (*Pisum sativum* L.) germplasm collection through genomic prediction (in preparation).

Landscape genomics or environmental association

 Lei, L., Poets, A.M., Liu, C., Wyant, S.R., Hoffman, P.J., Carter, C.K., Shaw, B.G., Li, X., Muehlbauer, G.J., Katagiri, F. and Morrell, P.L., 2019. Environmental association identifies candidates for tolerance to low temperature and drought. *G3: Genes, Genomes, Genetics, 9*(10), pp.3423-3438.

Figure 1 The genomic distribution of outlier SNPs identified in the F_{ST} comparisons of elevation (below 3,000m vs. above 3,000m), low latitude (below 30°N vs. 30-40°N), high latitude (30-40°N vs. above 40°N), growth habit (winter vs. spring), and association analysis of 21 bioclimatic variables, which are categorized into three classes (precipitation, temperature, and geographic variables) and ICs.

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Limited divergent adaptation despite a substantial environmental cline in wild pea

Timo Hellwig^{1,2} | Shahal Abbo¹ | Amir Sherman² | Clarice J Coyne³ | Yehoshua Saranga¹ | Simcha Lev-Yadun⁴ | Dorrie Main⁵¹⁰ | Ping Zheng⁵ | Ron Ophir²¹⁰

Pisum fulvum

- GBS genotyping
- Georeferenced
- Collection of 126 accessions deposited into GRIN
- Regeneration underway

GRIN-Global Curator Tool-everything you need to efficiently curate your crops at your finger tips with a few key strol s.

Resour

e GRIN-Global project's mission is to provide a scalable version of the Germplasm ource Information Network (GRIN) suitable for use by any interested genebank in the The GRIN-Global database platform has been and is being implemented at various ks around the world. For more information about GRIN-Global, review this ages or contact the GG International Help Desk.

www.grin-global.org

1.0.7, was released in December, 2011 in a joint effort by the Global

Hot-Sync Treeview with Data

installed the GRIN-Global documentati/ system and launched the Pakistani-C website by technical support provi/ Global Crop Diversity Trust and consultation of IT specialist V

Connected to: https://npgsweb.ars-grin.gov/GRINGlobal/GUI.asmx

(BrAPI?)

Germinate database

Received: 30 April 2020 Accepted: 16 June 2020

DOI: 10.1002/csc2.20248

Crop Science

SPECIAL ISSUE ARTICLE

Adapting Agriculture to Climate Change: A Walk on the Wild Side

From bits to bites: Advancement of the Germinate platform to support prebreeding informatics for crop wild relatives

Sebastian Raubach¹ | Benjamin Kilian² | Kate Dreher³ | Ahmed Amri¹⁹ | Filippo M. Bassi¹⁹ | Ousmane Boukar¹² | Douglas Cook⁸ | Alan Cruickshank¹⁵ | Christian Fatokun¹¹ | Noureddine El Haddad¹⁹ | Alan Humphries¹⁰ | David Jordan¹⁵ | Zakaria Kehel¹⁹ | Shiv Kumar¹⁹ | Sandy Jan Labarosa¹⁷ | Loi Huu Nguyen¹⁶ | Emma Mace¹⁵ | Susan McCouch¹⁴ | Ken McNally¹³ | David F. Marshall⁷ | Erick Owuor Mikwa⁹ | Iain Milne¹ | Loren H. Rieseberg⁵ | Roland Schafleitner¹⁸ | Shivali Sharma⁶ | Gordon Stephen¹ | Huynh Quang Tin¹⁶ | Abou Togola¹¹ | Emily Warschefsky⁵ | Peter Werner² | All our CWR Pre-Breeding Partners and Collaborators | Paul D. Shaw¹ |

BREEDING INSIGHT

Genotyping

- Information technology, genomics, and breeding process design that will partner with individual ARS pre-breeding and breeding groups and <u>NPGS</u> <u>genebanks</u>.
- The project will leverage investments in open source informatic tools that are already funded by USDA ARS, CGIAR, USAID, and the Bill and Melinda Gates Foundation.

Yes, plant genetic resource collections ARE the center of the universe

Graphic published by Esposito, S., Carputo, D., Cardi, T. and Tripodi, P., 2020. Applications and Trends of Machine Learning in Genomics and Phenomics for Next-Generation Breeding. *Plants*, *9*(1), p.34.

Acknowledgements

- USDA
 - Dr. Rebecca McGee
 - Britton Bourland
- Washington State University
 - Professor Dorrie Main
 - Professor Sindhuja Sankaran
 - Dr. Yu Ma
 - Afef Marzougui
 - Dr. Ping Zheng
- Cornell University
 - Professor Mark Mazourek
 - Dr. Bill Holdsworth
- ICARDA
 - Dr. Shiv Kumar
 - Dr. Karthika Rajendran

- University of Saskatchewan
 - Professor Kirstin Bett
- North Dakota State University
 - Professor Nonoy Bandillo
 - Dr. Md. Abdullah Al Bari
- The Hebrew University of Jerusalem
 - Professor Shahal Abbo
 - Professor Yehoshua Saranga
 - Dr. Timo Hellwig
- Volcani Center
 - Professor Ron Ophir
 - Professor Amir Sherman

Research sponsors and partners, present and past

