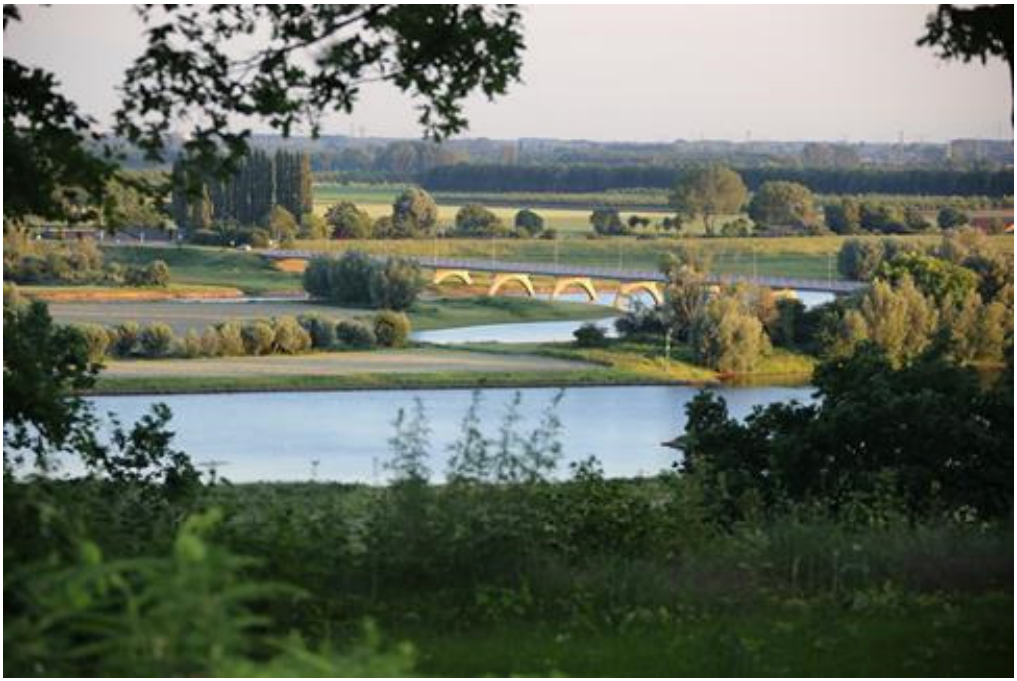
XVIth meeting of the Eucarpia Section Biometrics in Plant Breeding



Hotel “De Wageningsche Berg”, Wageningen NL 9 – 11 September 2015

Biometrics EUCARPIA 2015

WAGENINGEN UR
For quality of life

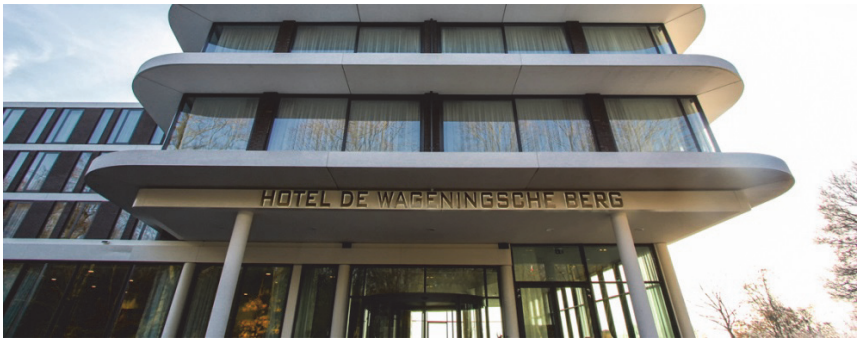
BOOK of ABSTRACTS

EUCARPIA

16th Meeting of EUCARPIA Section Biometrics for Plant Breeding

Wageningen

Wednesday 9th – Friday 11th September 2015



Local organizing committee

Marco Bink
Daniela Bustos Korts
Marcos Malosetti
Dinie Verbeek
Roeland Voorrips

International scientific committee

Fred van Eeuwijk	The Netherlands
Dietrich Borchardt	Germany
Alain Charcosset	France
Pawel Krajewski	Poland
Ian Mackay	United Kingdom
Chris Maliepaard	The Netherlands
Laurence Moreau	France
Hans-Peter Piepho	Germany
Carlotta Vaz Patto	Portugal

A word of welcome

It is an honour to welcome you to the 16th meeting of the Eucarpia section Biometrics in Plant Breeding. Like the other sections of Eucarpia, the section Biometrics in Plant Breeding aims at facilitating interactions and discussions between researchers in plant breeding. For the Biometrics in Plant Breeding section, the themes are related to the use of quantitative methods to improve breeding and selection. In plant breeding, genomic prediction and selection have been identified as highly promising techniques. I am therefore pleased that the scientific committee has been able to compose an excellent one day program dealing with genomic prediction and selection on Wednesday 9 September. Of course, also other topics remain of high interest and these will be presented and discussed on Thursday 10 and Friday 11 September.

The scientific program as a whole contains a keynote speaker, five invited speakers, 25 contributed speakers and more than 30 poster presenters. There will be ample opportunity to exchange ideas all along the conference, but in this context I would like to recommend especially the poster session, where drinks and food will be served as well. The social program provides a choice between three excursions on Thursday afternoon and a conference dinner on Thursday evening.

I wish to thank the local and scientific committees for the organization and the sponsors for extra financial support. I am also very happy that the invited speakers were willing to accept the invitations of the scientific committee and that so many researchers made the effort to submit contributed papers and posters.

On behalf of the scientific and local organizing committees I hope you will enjoy the meeting.

Fred van Eeuwijk,
Chairman of the Biometrics in Plant Breeding
Section of EUCARPIA

Fred van Eeuwijk

Premium Sponsors

VSNi
Bioscience Software & Consultancy

LemnaTec
We are Phenotyping

ENZA ZADEN

KWS
Seeding the future
since 1856

MONSANTO

Regular Sponsors:

KeyGene

Limagrain
from earth to life

syngenta

RIJK ZWAAN

SESVANDERHAVE
sugar beet seed

BAYER
Bayer CropScience

DUPONT
PIONEER.

BARENBRUG

Averis Seeds

Wednesday 9 September

07:30	**Registration**	
08:30	**WELCOME**	
08:30	Chair of Local Organizing Committee (M. Bink)	
08:40	President of Eucarpia section Biometrics for Plant Breeding (F. van Eeuwijk)	
08:50	Representative of Eucarpia Board (S. van der Heijden)	
09:00	**SESSION 1**	**Genomic and marker assisted breeding** *Chair: Marco Bink*
09:00	**John Hickey (invited)**	Sequence to Phenotype: Allocation of Resources
09:40	Chris Gaynor	Genomic Selection for Grain Yield in Kansas Wheat
10:00	Emi Tanka	On the joint use of pedigree and marker information for genomic selection in wheat breeding
10:20	**Coffee break**	
10:50	**SESSION 2**	**Genomic and marker assisted breeding** *Chair: Laurence Moreau*
10:50	**Luc Janss**	Genomic analysis in tetraploid potato using genotyping-by-sequencing
11:20	Willem Kruijer	Marker-Based estimation of heritability for plant traits
11:40	Christina Lehermeier	Assessment of genetic heterogeneity in structured plant breeding populations using multivariate whole-genome regression models
12:00	Daniela Bustos	Improving prediction accuracy by using models adapted to traits genetic architecture, and by considering the genetic distances to define the training set
12:20	**Group Picture**	
12:30	**Lunch at hotel**	
13:30	**SESSION 3**	**Genomic and marker assisted breeding** *Chair: Fred van Eeuwijk*
13:30	**Peter Visscher (keynote)**	Genomics and big data in human populations: combining genetics and epigenetics to predict phenotypes
14.20	**Roel Veerkamp**	Genomic selection in animal breeding: a success story
15:00	**Coffee break**	
15:30	**SESSION 4**	**Genomic and marker assisted breeding** *Chair: Ian Mackay*
15.30	Héloïse Giraud	QTL detection in a reciprocal recurrent design for silage in maize
15.50	Marcos Malosetti	Predictions for new genotypes and/or environments: models and designs
16:10	Jong Jiang	Modelling Epistasis in Genomic Selection
16.30	Hélène Muranty	Genomic selection in apple: a multi-year pilot study
16:50	**SESSION 5**	**Various themes** *Chair: Marcos Malosetti*
16:50	Donghui Ma (VSNi)	IBP - Breeding Management System (Sponsored Demo)
17:05	Poster Flashes	**1 slide & 1 minute each**
17:30	**Poster session**	*(with drinks & food)*

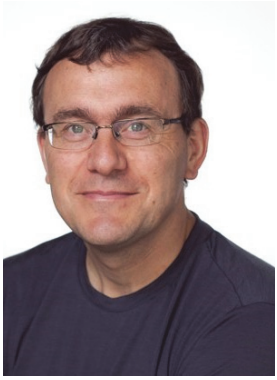
Thursday 10 September

08:00	Registration	
08:30	**SESSION 6**	**Implementaton of breeding strategies in public and private sector** *Chair: Hans-Peter Piepho*
08:30	**Neil Haussman (invited)**	Future Breeding Systems: view from DuPont Pioneer
09:10	**Andres Gordillo**	Genomic selection strategies and validation in hybrid maize and rye
09:40	Jaap Buntjer	Strategic use of genomic prediction models for parent selection in breeding programs.
10:00	**Coffee break**	
10:30	**SESSION 7**	**Analysis and use of high throughput data** *Chair: Chris Maliepaard*
10:30	**Dave Marshall**	The data challenges from the application of high throughput technologies in plant breeding and genetics
11:00	**Marco van Schriek**	Exploitation of digital phenotype markers for prediction of brassica napus field seed yield
11:30	Christine Hackett	Linkage map construction in blackcurrant using genotyping by sequencing data
11:50	Pawel Krajewski	On recommendations for metadata and data handling in plant phenotyping
12:10	Heike Sprenger	Discovery of Transcript and Metabolite Markers for the Breeding of Drought Tolerant Potato Cultivars using High Throughput Technologies
12:30	**Lunch at hotel**	
13:30	**Excursions**	(choices to be made via website or at arrival, return at hotel ~18:00 hours)
19:00	**Conference Dinner**	(hotel “De Wageningse Berg”)

Friday 11 September

08:30	**SESSION 8**	**Exploitation of genetic resources** *Chair: Carlotta Vaz Patto*
08:30	**Jens Riis-Jacobsen**	Accelerate genetic gain by taking advantage of additional data sources and
	(invited)	integrated data analysis – case studies from maize and wheat breeding at CIMMYT
09:10	Keith Gardner	A high density genetic linkage map for wheat using an eight-founder MAGIC population
09.30	Andreas Maurer	Modelling the genetic architecture of flowering time control in barley through nested association mapping
09:50	**Business Meeting**	**Eucarpia section Biometrics for Plant Breeding**
10:20	**Coffee break**	
10:50	**SESSION 9**	**Experimental design and analysis of phenotyping trials** *Chair: Pawel Krajewski*
10:50	**Alison Smith (invited)**	Experimental designs for expensive multi-phase traits
11:30	**Hans Peter Piepho**	The generation of efficient row-column designs for field trials
12:00	Fred van Eeuwijk	P-spline models for spatial variation in agricultural field trials
12:20	**Lunch at hotel**	
13:30	**SESSION 10**	**Various themes** *Chair: Dietrich Borchardt*
13:30	**Emma Huang (invited)**	Meta-alleles in multiparental populations
14:10	Peter Bourke	High-density linkage mapping and the double-reduction landscape in tetraploid potato
14:30	Chaozi Zheng	Multilocus haplotype reconstruction in outcrossing tetraploids
14:50	**Concluding remarks**	
15:00	**Farewell drinks**	

Keynote speaker



Peter Visscher

Peter Visscher's undergraduate studies were in the Netherlands. He moved to Edinburgh (UK) in 1987 for an MSc and subsequent PhD in genetics, working on the estimation of genetic parameters in large livestock pedigrees. A postdoctoral period in Melbourne (Australia) was followed by a return to Edinburgh, where he developed methods and software to map genetic loci underlying complex traits. In 1995 he moved to a faculty position at the Institute of Evolutionary Biology of the University of Edinburgh, developing gene mapping methods and software tools, with practical applications in livestock and human populations. Visscher joined the Queensland Institute of Medical Research in Brisbane (Australia) in 2005 and in 2011 moved to the University of Queensland where he is Professor and Chair of Quantitative Genetics and Director of the Centre for Neurogenetics and Statistical Genomics. Visscher is a Senior Principal Research Fellow of the Australian National Health and Medical Research Council and was elected a Fellow of the Australian Academy of Science in 2010. Visscher's research interests are focussed on a better understanding of genetic variation for complex traits, including quantitative traits and disease, and on systems genomics.

Invited speakers

Emma Huang

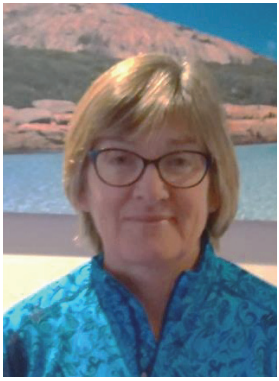
Dr. Emma Huang is a statistical geneticist whose research has focused on genetic population analysis in plants and humans. She received her doctorate in Biostatistics from the University of North Carolina in 2007 developing methods to map association between genes and disease in humans. From 2007 to 2015, she held a position at CSIRO, most recently as a senior research scientist working on experimental design, linkage map construction, and QTL mapping in biparental and multiparental populations. There she received numerous awards, including an Australian Research Council DECRA to develop an integrated analysis framework for

multiparental populations.

John Hickey is a quantitative geneticist who works at The Roslin Institute in Edinburgh (http://www.alphagenes.roslin.ed.ac.uk/). His area of research spans animal breeding, plant breeding, and human genetics. In particular he seeks to develop computational methods to generate and analyse huge data sets with whole genome sequence information, methods and breeding strategies that use genomic information to increase rates of genetic progress, and the concept of Genomic Selection 2.0. Software and algorithms developed by John Hickey underpin aspects of several of the largest breeding programs globally.

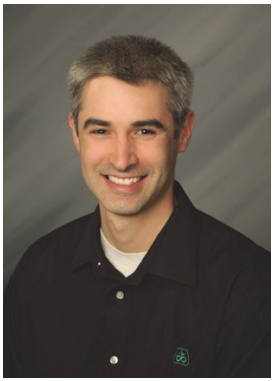
John Hickey

Alison Smith has worked as a biometrician for more than 25 years. During this time she has focussed on the design and analysis of data from plant breeding and crop improvement programs. In her PhD, completed in 2000, she developed a method for the analysis of yield data from multi-environment trials that is now used in all major plant breeding programs in Australia. Alison currently works for the University of Wollongong on the Grains Research & Development Corporation project “Statistics for the Australian Grains Industry”. Alison has published over 45 refereed journal articles and has been invited to present her research at a number of national and international statistical and scientific conferences. She has active links with industry, including most private and public plant breeding programs in Australia.



Alison Smith

Neil Hausmann is a Senior Research Manager at Pioneer Hi-Bred International in the Breeding Technologies group. Dr. Hausmann earned his Ph.D. degree in Integrative Biology from the University of California, Berkeley in 2004. Immediately thereafter, he joined Pioneer in Woodland, California leading a native drought characterization program in maize. In 2013, Dr. Hausmann moved to Johnston, Iowa to lead a team overseeing global Doubled Haploid technologies and field-based phenotyping efforts.



Neil Hausmann

Jens Riis-Jacobsen, Information Technology Unit, CIMMYT, Jens leads the informatics team at CIMMYT that work with the plant breeders and researchers on creating an integrated breeding informatics toolbox based on existing components from other institutions. Previously, Jens has worked on developing information systems for sweet potato, and potato breeding for CIP in Africa, and maize breeding at CIMMYT. He has also worked on other information systems including monitoring of rainforest reserve, land titling, municipal tax collection, finance, and branding iron registry. Jens has a Master in Technology Management from Aalborg University and an MSc in Forestry from University of Oxford.

Jens Riis-Jacobsen

Session 1

Genomic and marker assisted breeding

Sequence to Phenotype: Allocation of Resources

John M Hickey[1,], R Chris Gaynor[1], Ian J Mackay[2], Gregor Gorjanc[1]*

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Research Centre, Midlothian EH25 9RG, UK
[2]John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge, CB3 0LE, UK
[*] John.Hickey@roslin.ed.ac.uk

Background
Genomic selection is increasingly valued within the plant breeding community. To implement genomic selection large investments are needed in genomic data (markers and or sequence) and phenotypic data on which to train prediction equations. Choices about distributing these resources affect the return on investment

Results
A simulation was conducted which evaluated the long term benefit of three alternative breeding program designs: (i) a classical plant breeding program design; (ii) a minor modification to the classical design in which genomic prediction was used to increase the accuracy of preliminary yield trials; and (iii) a complete reorganization of the breeding program into a population improvement component driven by genomic selection and a product development component that was similar to i.

Conclusions
The different breeding program designs gave different returns on investment. Complete reorganization of plant breeding programs into population improvement components driven by genomic selection and product development components was promising but its benefit was affected by costs.

Genomic Selection for Grain Yield in Kansas Wheat

Chris Gaynor[1,3], Jesse Poland[2], and Allan Fritz[3]

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland
[2]Wheat Genetics Resource Center, Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA
[3]Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA
[*] Chris Gaynor, chris.gaynor@roslin.ed.ac.uk

Genomic selection has the potential to improve breeding for complex, quantitative traits such as grain yield in wheat. Many of the published works showing this potential use simulations and/or conditions not representative of a wheat breeding program. This leaves some doubt as to how well genomic selection would perform in a real world situation. To examine this, genomic selection was performed using germplasm and yield data generated by the Kansas State University wheat breeding program during the course of its normal breeding operations. Using only data that would have been available prior to planting, multiple genomic selection models were constructed to predict grain yield for new entries in the 2013 and 2014 preliminary yield nurseries. Correlations between predicted yields and observed multi-location means for the best performing models were 0.21 and 0.22, for each year respectively. These values were comparable to observed phenotypic correlations in the breeding program for between-year multi-location means (0.06-0.30). Considerably higher correlations for genomic selection were obtained with models using a within-year cross-validation approach (0.38 and 0.57, respectively). These higher values are consistent with values reported in previous literature, but the lower values are more representative of how genomic selection would perform in a breeding program. These findings indicate genomic selection accuracies may be reduced by real world constraints in a breeding program, but genomic selection is still capable of predicting future performance. Further more, accuracies comparable to those obtained using phenotypic selection can be achieved.

On the joint use of pedigree and marker information for genomic selection in wheat breeding

Emi Tanaka[1], Alison Smith[1], Brian Cullis[1]*

[1]School of Mathematics and Applied Statistics, University of Wollongong
∗ presenting author and e-mail
[*]etanaka@uow.edu.au

Historically, prediction of additive genetic effects was based on the pedigree-based additive infinitesimal model (Fisher, 1918), however molecular markers paved ways to marker-assisted selection and now, with the availability of dense genome-wide molecular markers, to the use of genomic selection (Meuwissen et al., 2001).

In crop breeding, the use of line replication allows for genetic effects to be partitioned into additive and non-additive (or residual) effects using a linear mixed model approach. The additive effects are further partitioned into marker effects and marker lack of fit (or polygenic) effects with the joint use of marker and pedigree information (Garrick et al., 2009).

Models that jointly consider both marker and pedigree information have shown improvement to the prediction of quantitative traits to either marker-based models or pedigree-based models (Crossa et al., 2010, Burgueno et al., 2012). The improvement of the joint marker and pedigree-based model shown was marginal over marker-based models while significant over pedigree-based models.

In this talk, we present a comprehensive study of single-environment one-stage analysis of over 500 wheat breeding trials using a linear mixed model approach to show the effectiveness of the inclusion of the pedigree information.
We also address an important but often less mentioned aspect of genomic selection – model selection of the peripheral effects. We present a new diagnostic tool to assess the model fit.

Session 2

Genomic and marker assisted breeding

Genomic analysis in tetraploid potato using genotyping-by-sequencing

Luc Janss[1], Elsa Sverrisdóttir[2], Kåre Lehman Nielsen[2]

[1]Aarhus University, Department of Molecular Biology and Genetics
[2]Aalborg University
[*] luc.janss@mbg.au.dk

Standard genotyping systems are not suitable for tetraploids. The standard systems only distinguish one heterozygote, while tetraploids have three. We consider here the use of genotyping-by-sequencing (GBS) to genotype tetraploid potato. GBS is used in a quantitative way to estimate the allele-dose of a genotype. These allele-dose estimates have estimation error from using finite read-depth on a genotype, which equals in general (1-1/*P*)/*D*, where *P* is the ploidy level, and *D* the read-depth. Tetraploids have a lower homozygosity rate, which reduces the genetic variance by a factor 2 compared to diploids. This was indeed reflected in the GBS estimated genotypes in our potato genotypes. Genomic heritability estimates for some traits are presented using genomic relationship matrices based on GBS estimated genotypes corrected for error from low read-depth.

Marker-Based estimation of heritability for plant traits.

Willem Kruijer[1], M.P. Boer[1], M. Malosetti[1], F.A. van Eeuwijk[1]*
[1]Biometris, Wageningen UR
[*] willem.kruijer@wur.nl

Background
Heritability is an important parameter in quantitative genetics, determining among others the response to selection. While plant biologists and breeders commonly estimate repeatability or broad-sense heritability from phenotypic observations on replicates, there is an increasing body of methodology to estimate narrow-sense heritability from a dense set of markers, in particular for human traits. Here we investigate to which extent this methodology can be applied to wild plant populations, for which the number of genotyped individuals is typically much lower than in human genetics. Other differences are the strong genetic relatedness present in most plant populations, and the availability of genetically identical replicates.

Results
We investigate the statistical properties of marker-based heritability estimates for several populations of maize, rice and arabidopsis. When performing mixed model analysis on the genotypic means, sampling variances are considerably larger than what has been reported for many human traits, i.e. the extra information provided by the higher genetic relatedness does not compensate for the smaller sample size. This changes when mixed model analysis is performed on individual plant (or plot) data, in which case the sampling variance becomes several times smaller. This can also improve the accuracy of the corresponding G-BLUP. A disadvantage of 'plant-level' heritability estimators is that they become biased in the presence of independent non-additive genetic effects. We show that this bias also occurs for the estimator based on genotypic means, when the non-additive genetic effects are due to infinitesimal epistatic interactions. This occurs because the covariance structure induced by these effects resembles the structure of the relatedness matrix for the additive effects.

Conclusions
Marker-Based heritability estimation for plant traits can greatly benefit from mixed models at individual plant or plot level, but the currently available methodology needs further development in order to account for non-additive effects.
References

P.M. Visscher and M.E. Goddard (2015) A General Unified Framework to Assess the Sampling Variance of Heritability Estimates Using Pedigree or Marker-Based Relationships. *Genetics* (199), 223–232

D. Speed and D.J. Balding (2015) Relatedness in the post-genomic era: is it still useful? *Nature Reviews Genetics* 16 (1), 33-44

W. Kruijer et al. (2015) Marker-based estimation of heritability in immortal populations. *Genetics* 199, 379-398

W. Kruijer (2015) Misspecification in mixed-model based association analysis. (submitted)

Assessment of genetic heterogeneity in structured plant breeding populations using multivariate whole-genome regression models

Christina Lehermeier[1], Chris-Carolin Schön[1], Gustavo de los Campos[2]*

[1] Plant Breeding, Technische Universität München, 85354 Freising, Germany
[2] Department of Epidemiology & Biostatistics and Department of Statistics, Michigan State University, East Lansing, MI 48824, USA
[*]christina.lehermeier@tum.de

Plant breeding populations exhibit varying levels of structure and admixture; these features are likely to induce heterogeneity of marker effects across sub-populations. Traditionally, structure has been dealt with as a potential confounder, and various methods exist in genome-wide association studies and genome-based prediction to 'correct' for population stratification. However, these methods induce a mean-correction that does not account for heterogeneity of marker effects. The animal breeding literature offers a few recent studies that consider modeling genetic heterogeneity in multi-breed data using multivariate models. However, these methods have received little attention in plant breeding where population structure can have different forms. Here we address the problem of analyzing data from heterogeneous plant breeding populations using three approaches: (a) a model that ignores population structure, (b) a stratified (i.e. within group) analysis, and (c) a multivariate approach that uses multi-group data and accounts for data heterogeneity. The performance of the three approaches was assessed on different plant data sets with distinct population structure. The estimated genomic correlations between sub-populations varied from null to moderate depending on the genetic distance between sub-populations and traits. Our assessment of prediction accuracy features cases where ignoring population structure leads to a parsimonious more powerful model as well as others where the multivariate and stratified approach have higher predictive power. In general, the multivariate approach appeared slightly more robust than the other two approaches.

Improving prediction accuracy by using models adapted to traits genetic architecture, and by considering the genetic distances to define the training set

Daniela Bustos-Korts[1,2], Marcos Malosetti[1], Fred van Eeuwijk[1]*

[1] Biometris, Wageningen University and Research Centre, Wageningen, The Netherlands
[2] C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC), Wageningen, The Netherlands
** daniela.bustoskorts@wur.nl*

Accuracy in genomic prediction is influenced on the one hand by the genetic architecture of the trait, and on the other hand by the population that was used to train the model. Trait genetic architecture is relevant because it determines the relative importance that the model needs to give to particular regions along the genome, and whether additive and/or non-additive genetic effects need to be considered. Training set composition and size influence the ability of a prediction model to borrow information from observed to unobserved genotypes and/or from observed to unobserved environments. The higher the overlap of the training and the test population, the higher the expected accuracies of the predictions.
The two major objectives of this work were: 1) to compare different sampling methods to compose the training population set accounting or not for the population structure and 2) to compare prediction models that differ in the weighing of genomic regions. The sampling strategies that we evaluated were: a) random sampling, b) stratified sampling, and c) genetic distance sampling. The following prediction models were compared; GBLUP, QTL-based prediction, QTL+GBLUP and a Gaussian kernel to account for additive and non-additive genetic effects. Two traits, known to have different genetic architecture (heading date and yield) evaluated in a wheat diversity panel in Australia were used to evaluate these strategies and models.
Our results show that using genetic distances to define the training population set leads to higher prediction accuracy, compared to random or stratified sampling. As expected using QTL information gave an advantage for heading date (QTL+GBLUP>GBLUP) but not for grain yield (GBLUP>QTL+GBLUP). The use of a Gaussian kernel performed best or close to the best for both traits.

Session 3

Genomic and marker assisted breeding

Genomics and big data in human populations: combining genetics and epigenetics to predict phenotypes

Peter M. Visscher, University of Queensland

Driven by advances in genome technologies, the last 8 years have witnessed a revolution in our understanding of complex trait variation in human populations. Results from genome-wide association studies and whole-genome exome studies have shown that the mutational target in the genome for most traits appears to be very large, such that many genes are involved in explaining genetic variation. Genetic architecture, the joint distribution of the effect size and frequency of variants that segregate in the population, is becoming clearer and differs between traits. We will show new results from disparate complex traits including height, schizophrenia and gene methylation, to illustrate polygenicity and the power of experimental sample size. In addition, we will show emerging results that epigenetic information can be used to make predictions of complex traits and that gene methylation can be a predictor of past environmental exposures.

Genomic selection in animal breeding: a success story

Roel F. Veerkamp[1*]

[1] *Animal Breeding and Genomics Centre, Wageningen UR, P.O. Box 338, 6700 AH Wageningen, The Netherlands*
[*] *Roel.Veerkamp@wur.nl*

Since the availability of molecular techniques (e.g. micro-satellites, RFLP), there has been an interest in identifying variation in genes that regulate heritable traits. It has only been since the introduction of genomic selection, using a large number of SNP across the genome for genomic prediction (Meuwissen et al., 2001), that a big impact has been made in practical breeding programs. The aim here is to demonstrate, as an example, the introduction of genomic selection in cattle breeding, and the impact it has made.

In 2006, the Dutch breeding company CRV announced the first implementation of genomic selection. To date all major dairy cattle breeding companies have implemented genomic selection programs, where genomic prediction equations are trained in a reference population containing tens of thousands of bulls that have accurate phenotypes containing the records of many daughters.

The introduction of genomic selection has changed classical breeding programs top benefit from the genomic prediction. Classical breeding programs waited for phenotypes of some female offspring (cows) of test bulls before major selection decisions were taken on candidates. Waiting for the phenotypes of offspring was beneficial as it helped to increase the selection accuracy from 60% (for the parent average) to 85%. To date genomic prediction without having to wait for offspring records, gives an accuracy of about 75-80%. Even with relatively low density SNP panels (e.g. 50k SNP).

Compared with classical breeding programs, the current genomic based breeding programs achieve additional rates of gain of 60-70% (1.6-1.7x more genetic gain every year!). The extra gain is primarily observed because the generation interval has been reduced. In 2014, the age of sires (and dams) of selected AI bulls was around 1250 days (1100) (de Jong & Stoop, 2014, http://www.interbull.org/static/web/7_3_deJong.pdf), whereas in 2008 the age for parent selection was twice as long for sires 2500 days (1400).

The current success of genomic prediction is very much based on the use of long-range LD due to strong family relationships in the population. To increase the accuracy of genomic prediction even further and to predict across populations using sequence information, more sophisticated SNP training is required.

Session 4

Genomic and marker assisted breeding

QTL detection in a reciprocal multiparental design for silage in maize

H. Giraud[1], C. Bauland[2], A. Charcosset[2] , L. Moreau[2]*

[1]Université Paris-Sud, UMR Génétique Quantitative Evolution – Le Moulon, F-91190 Gif-sur-Yvette, France
[2]INRA, UMR Génétique Quantitative Evolution - Le Moulon, F-91190 Gif-sur-Yvette, France
[*]heloise.giraud@moulon.inra.fr

To valorize heterosis in maize, genetic diversity was structured in heterotic groups. Improvement schemes are based on a two stages approach: first a selection is carried out in each group based on the evaluation of hybrids obtained by crossing candidates with a reduced number of lines (testers) from the opposite group, and then improved lines of each group are crossed following an incomplete factorial design to identify the best hybrid combinations. As a consequence, selection is mainly based on the general combining ability (GCA) of the lines and poorly exploits the specific combining abilities (SCA) between the groups. In order to better understand the genetic architecture of GCA and SCA, we developed a reciprocal multiparental design between the two main heterotic groups used for maize silage production in Northern Europe. Eight inbred lines (four in each heterotic group) were crossed according to a half-diallel and approximately 900 inbred lines were developed in each group. The inbred lines were genotyped for approximately 20k SNPs and were crossed according to an incomplete factorial design in order to produce about 1000 hybrids. These hybrids were phenotyped in eight environments for traits related to silage: dry matter yield, dry matter content, female flowering, and plant height. We decomposed the genetic variance at the hybrid level and estimated the relative contributions of GCA and SCA. We performed QTL detection using different allele codings: (i) alleles observed at markers at the hybrid level (same effect assumed for the two groups), (ii) allele observed at markers but considering different allelic effects in each group, (iii) alleles transmitted from the eight parents. These different allele codings were found to be complementary, ranking in significance depending on the QTL region considered. They allowed us to simultaneously detect QTL for GCA in the two groups as well as QTL involved in the SCA which opens interesting prospects to revisit with markers the concept of reciprocal recurrent selection.

Predictions for new genotypes and/or environments: models and designs

Marcos Malosetti[1], Daniela Bustos-Korts[1,2], Martin P. Boer[1], Fred A. van Eeuwijk[1]

[1] *Biometris, Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands.*
[2] *C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC), The Netherlands.*
[*] *marcos.malosetti@wur.nl*

Background: A central objective in plant breeding is to predict the performance of genotypes across multiple environments to guide selection decisions. To this end genomic prediction approaches can be used, training on a limited set of environments with dense sets of markers and then predicting to new environments. Our objective was to compare the performance of different multi-environment genomic prediction models, not only in terms of accuracy of prediction but also in terms of the ability to predict in unobserved environments.
Results: Using the well-known Steptoe×Morex barley double haploid population as example, we showed the superiority of models that allow environment-specific effects (GE-BLUP) over main effect prediction models (G-BLUP). Our results showed that the distribution of genotypes in the training set over the tested environments can have an impact on the prediction accuracy. Finally, our results highlighted the importance of using explicit environmental information (covariables) in the prediction model to generalise predictions to unobserved environments. Conclusions: Models that account for G×E are needed to predict responses over multiple environments. While several options are available, we think that the inclusion of environmental covariables in the prediction model is a particularly attractive option. It naturally links with eco-physiological and crop growth models for G×E, and generalises predictions to unobserved future environments.

Modelling Epistasis in Genomic Selection

Yong Jiang[1], Jochen C. Reif[1]*

[1]Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, 06466 Stadt Seeland, Germany
[*] jiang@ipk-gatersleben.de

Background
Modelling epistasis in genomic selection is impeded by a high computational load. The extended genomic best linear unbiased prediction (EG-BLUP) with a marker-inferred epistatic relationship matrix and the reproducing kernel Hilbert space regression (RKHS) are two attractive approaches reducing the computational load.

Results
We proved that the EG-BLUP model is equivalent to an approach explicitly modelling epistatic effects. Based on this result, we gave an explicit proof on why the RKHS model with a Gaussian kernel captures epistatic effects among markers. We compared the EG-BLUP and RKHS model with the usual G-BLUP model using published data sets in wheat and maize inbred lines. We observed that the prediction accuracies of EG-BLUP and RKHS outperformed G-BLUP in wheat data sets but not in maize data sets.

Conclusions
The prediction accuracy of genomic selection for selfing species can be improved by modelling additive by additive epistasis, which can be effectively implemented by EG-BLUP or RKHS. While for out-crossing species, more experimental data sets are required to examine the role of epistasis (probably involving dominance effects) in more detail.

Genomic selection in apple: a multi-year pilot study

Muranty H[1], Troggio M[2], Ben Sadok I[1], Al Rifaï M[1], Auwerkerken A[3], Stevanato P[4], van de Weg WE[5], Di Guardo M[2,5], Laurens F[1], Bink MCAM[6]*

[1]IRHS UMR1345, INRA, SFR 4207 QUASAV, F-49071 Beaucouze, France
[2]Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy
[3]Better3Fruit, Rillaar, Belgium
[4]University of Padova, Legnaro, Padova, Italy
[5]Wageningen UR Plant Breeding, Wageningen, Netherlands
[6]Wageningen UR Biometris, Wageningen, Netherlands
[*]Helene.Muranty@angers.inra.fr

Background The tremendous increase in throughput of genotyping techniques opened appealing perspectives for genomic selection in many crops, including fruit perennials. Genomic selection is expected to accelarate genetic progress in breeding programs, by increasing prediction accuracy, decreasing generation interval and increasing selection intensity. In the EU-FP7 project FruitBreedomics (www.fruitbreedomics.com), a pilot study in apple was performed with the objectives to (1) assess the accuracy of prediction and (2) study realized response to selection.
Results The training population comprised 977 individuals derived from 20 pedigree-related full-sib families and genotyped with the Infinium Illumina 20K SNP array. The same individuals were phenotyped for productivity, fruit aesthetic quality and internal fruit quality traits. Using the genotypic and phenotypic data in the training population, a genomic prediction model was built with the BayesC□□method in the publicly available software GS3. The application population was built using additional plant material from two commercial breeding programs. It comprised five full- sib families with a total of 1390 individuals that were phenotyped over two consecutive years at two complementary sites and genotyped on TaqMan OpenArray plates preloaded with 512 SNPs uniformly distributed over the apple genome that were a subset of the 20K SNP of the Illumina array. The genotypic data in the application population were imputed by using the software AlphaImpute. The accuracy of prediction across the 22 examined
traits reached a maximum value of 0.68 and had a median of 0.25. The accuracies were strongly affected by the phenotypic distribution and heritability of the traits. They were uncorrelated to the genomic relatedness of each family to the training population. Significant responses to genomic selection were obtained in 47 trait-family combinations out of 97, i.e., the

top 10% and bottom 10% differed by more than 0.6 within family standard error, and up to 2.4.

Conclusions The accuracy and selection response from genomic selection are very promising in apple and illustrate the great potential to accelerate breeding progress in perennial fruit tree crops that currently face long generation intervals and extensive phenotyping costs. A further evaluation of resource allocation is required to optimize the implementation of genomic selection.

Session 5

Various themes

Breeding Management System and its Analytics: Breeding View

Donghui Ma[1], Roger Payne[1], Darren Murray [1], Liz Jupe [1]*
[1] 2 Amberside House, Wood Lane, Hemel Hempstead, HP2 4TP UK.
[*]donghui@vsni.co.uk

The Breeding Management System (BMS), available at www.integratedbreeding.net, is an information management system developed by the Generation Challenge Programme's Integrated Breeding Project to help breeders manage the breeding process: from program planning to decision-making. The BMS suite of tools supports multiple breeding strategies including conventional breeding and many marker-assisted breeding programs.

For both conventional breeding and marker-assisted breeding, phenotypical data analysis and marker data analysis are the essential parts of the workflow. Within the BMS, these functions are performed by the Breeding View, a visual tool for running the analytical pipelines. The Breeding View provides a user-friendly interface to access breeding data analysis tools such as field trial analysis and QTL mapping. It contains a visual pipeline representation of the steps involved for an analysis ranging from data quality checks to producing final reports. Each visual pipeline includes a set of nodes that allow interaction to control the flow of the analysis. The results from an analysis are summarized within a combination of html reports, data export files (e.g. Microsoft Excel files) and image files. The Breeding View runs each analytical pipeline using an underlying statistical analysis engine. The main application used for the statistical analysis is GenStat, and also includes a facility to use R.

The Breeding View can be run as part of the Breeding Management System (BMS) within the Integrated Breeding Platform or as a standalone application. Due to its highly standardized and automated nature, using the Breeding View can significantly increase the efficiency and effectiveness of data analysis within the small time window available during the harvest campaign.

Session 6

Implementation of breeding strategies in public and private sector

Future Breeding Systems: view from DuPont Pioneer

Neil J. Hausmann, Andrew Baumgarten, Mark Cooper, Carlos D. Messina, Dean Podlich, Frank Technow, L. Radu Totir
DuPont-Pioneer, 7200 NW 62nd Ave, Johnston, IA USA 50131

Future breeding systems in DuPont Pioneer will be built upon development and deployment of predictive analytics capabilities explicitly integrating genomic, environmental, crop management and trait phenotypic knowledge. Improved understanding of the genetic architecture of traits, leveraging available sources of genetic diversity, will enhance the breeder's ability to utilise a range of model-based prediction methodologies to support the continually increasing scale of breeding programs. The accuracy of these predictions, especially when faced with complex traits displaying high levels of genotype by environment variability, will depend upon successful integration of quantitative genetics methodology, pedigree understanding, computer simulation and functional gene-to-phenotype models. The future breeding system require an adaptive field evaluation infrastructure that must meet needs of wide area testing while providing detailed phenotyping required for model parametrization. Examples will be drawn primarily from our experience gained through commercial maize breeding.

Genomic Selection Strategies and Validation in Hybrid Maize and Rye

A. Gordillo[1], M. Schmidt[1], P. Wilde[1], R. Bernardo[2], L.-M. Krchov[2,3], M. Schönleben[4], E. Bauer[4], C.-C. Schön[4]*

[1] KWS LOCHOW GMBH, Ferdinand-von-Lochow-Str. 5, 29303 Bergen, Germany
[2] University of Minnesota, Department of Agronomy and Plant Genetics, 411 Borlaug Hall, 1991, Upper Buford Circle, St. Paul, U.S.A
[3] KWS SAAT AG, Grimsehlstr. 31, 37555 Einbeck , Germany
[4] Technische Universität München, Plant Breeding, Liesel-Beckmann Str. 2, 85354 Freising, Germany
[*]andres.gordillo@kws.com

Genomic selection (GS) schemes can be classified according to the relationship between the training population (TP) and prediction population (PredP). GS schemes implemented in AgReliant Genetics' maize breeding program and KWS LOCHOW's rye breeding programs exemplify contrasting types of PredPs and different relationships between TP and PredP. In the hybrid maize GS scheme, large numbers of doubled haploid lines are developed within biparental populations. A portion of the lines is used as TP to predict untested lines from the same biparental population. Multi-environment validation experiments showed that prediction ability for grain yield is somewhat higher for lines in the TP and slightly lower for untested lines compared to phenotypic selection. Index selection using phenotypic and genomic grain yield values was consistently better than phenotypic or genomic selection alone (Krchov et al., 2015). Herein, GS allows predicting lines with insufficient seed set that otherwise would be delayed in their testcross evaluation by one year, which saves time and simplifies logistics. In the hybrid rye GS scheme, the TP includes independent populations tested in previous years. Validation experiments indicated that the prediction ability for grain yield was considerably higher for phenotypic selection than for GS when predicted lines were not included in the TP and GS was equivalent to phenotypic selection when predicted lines were included in the TP. This indicates a strong influence of the genetic background on estimates of marker-allele effects. Index selection using phenotypic and genomic grain yield values leads consistently to higher prediction abilities than phenotypic or genomic selection alone. Predictions based on TPs from previous selection cycles indicate a drop in prediction ability in the years 2013 and 2014. Validation experiments show that prediction abilities across selection cycles

depend on the proportion of parents and uncles of predicted lines in the TP. *Per-se* selection and variation in agronomic traits indirectly influencing yield may affect prediction ability. Using data from the evaluation of lines in multiple years and appropriate modelling of SNP by year effects as well as including parents of the PredP in the TP may be key to increase prediction abilities of GS across selection cycles.

References

Krchov, L.M., A. Gordillo, and R. Bernardo. 2015. Multienvironment Validation of the Effectiveness of Phenotypic and Genomewide Selection within Biparental Maize Populations.
Crop Sci. . doi:10.2135/cropsci2014.09.0608

Strategic use of genomic prediction models for parent selection in breeding programs

Niek Bouman[*1], *Timo Doeswijk*[1], *Jaap Buntjer*[1*]
[1]*Keygene N.V. P.O. Box 216 6700 AE Wageningen*
[*]*jbu@keygene.com*

In plant breeding, parent selection is an important application field for genomic prediction models. Parental lines may be selected based on genomic estimated breeding values (GEBVs) rather than on observed phenotypic trait values, which is in particular attractive for breeding practice in crops with long breeding cycles and for selection on traits that are difficult or expensive to measure.
Here we present and explain a novel method to select optimal subsets of parents from large candidate germplasm sets that are expected to yield the highest performance improvement in the next 2-5 breeding cycles. The principal difference is lain in not only using the GEBV prediction of the model as selection criterion, but also considering the genomic distribution of the allele substitution effects.
Simulation studies based on public variomic data sets of different crops demonstrate that application of this pre-selection method results in faster progress than selection on single parent breeding values.

Session 7

Analysis and use of high throughput data

The data challenges from the application of high throughput technologies in plant breeding and genetics

*David Marshall, *[1,2] Iain Milne[1], Paul Shaw[1], Gordon Stephen[1], Sebastian[1] Raubach, Sarah Hearne[1,2].*

[1]The James Hutton Insitute, Invergowrie, Dundee DD2 5DA, UK
[2]CIMMYT, Km. 45, Carretera, México-Veracruz El Batán, Texcoco CP 56237 Edo. de México. MEXICO
[*] David.Marshall@hutton.ac.uk

As new high through put technologies for sequencing, genotyping and phenotyping begin to impact on plant genetics, genetic diversity and breeding applications there is increasingly a need to develop and deploy the computational tools and infrastructure to deal with the resulting high volumes of data at every stage in the pipeline from generation to short and long term storage as well as support interactions with analysis and visualization tools. Some of these challenges are simply ones of resourcing i.e providing sufficient computing power either locally of through cloud solutions. However, in many existing software tools will not scale easily and in many plant breeding applications there may be a need to pass from data generation to analysis and decision- making in very short time frames. At local, national and international scales there are a number of developments such as such as the US funded iPlant project (http://www.iplantcollaborative.org/) which is now expanding into Europe to support analysis and the developing Plant Breeding API project (http://docs.brapi.apiary.io/) which is working towards the provision of a common API that can be used to integrated a variety of software tools and data sources in the domain of plant breeding and genetics. In this talk I shall discuss some of the major challenges that we face together with current or developing solutions where they exist.

Exploitation of digital phenotype markers for prediction of *brassica napus* field seed yield

Rod Snowdon[1], Marie Hohmann[1], Gert-Jan Speckmann[2], Koen Huvenaars[2], Marco van Schriek[2]
[1]Univ. of Giessen, [2]Keygene N.V. Wageningen, The Netherlands

Controlled phenotyping platforms or conventional pot trials are often used to compare plant performance under water stress and non-stressed conditions. Studies which compare stress reactions of plants under these controlled conditions with performance in the field are rare. Correlations between pot experiments and field trials are essential in order to identify and exploit morphological or physiological selection criteria for practical breeding approaches. To this end a selection of diverse winter oilseed rape cultivars known to show variable stress responses in the field were screened. All cultivars we grown in irrigated and non-irrigated field trials at multiple locations in Germany. The same cultivars were also grown under water-stressed and non-stressed conditions in two controlled experiments. Firstly a container experiment where the experiment was performed over a complete growing season so that seed yield could be measured. For the second experiment the cultivars were tested by digital phenotyping using the PhenoFab system. I will present correlations between early digital phenotypes and field yield-relevant parameters observed in this study.

Linkage map construction in blackcurrant using genotyping by sequencing data

Christine Hackett[1], Katharine Preedy[1]*
[1]Biomathematics and Statistics Scotland, Invergowrie, Dundee DD2 5DE, UK
[*]Christine.Hackett@bioss.ac.uk

Genotyping by sequencing is a recent technology that enables large numbers of single nucleotide polymorphisms (SNPs) to be detected and is suitable for the analysis of crops such as blackcurrant that do not have a sequenced genome. This presentation focuses on linkage map construction for a population of two parents and their F1 progeny. Our analysis starts with data on the frequency of the major and minor alleles in the parents and offspring and explores these graphically before using a method based on functional regression to call the genotypes for each individual. In the linkage map construction, we use an approach based on multi-dimensional scaling to obtain the order of SNPs along each linkage group rapidly, and to identify poorly-fitting markers. Further SNPs with unusual segregation patterns were detected and were mapped as quantitative traits using the allele count data. This identified that some of these unusual patterns were likely to be due to null alleles at the SNPs. The method gave a high-quality linkage map with up to 200 SNPs on each linkage group, and more approximate locations for many SNPs of lower quality.

On recommendations for metadata and data handling in plant phenotyping

Paweł Krajewski[1], Dijun Chen[2], Hanna Ćwiek[1], Aalt D.J. van Dijk[3], Fabio Fiorani[4], Paul Kersey[5], Christian Klukas[2], Matthias Lange[2], Augustyn Markiewicz[6], Jan Peter Nap[3], Jan van Oeveren[7], Cyril Pommier[8], Uwe Scholz[2], Marco van Schriek[7], Björn Usadel[4,9], Stephan Weise[2]*

[1]*Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland*
[2]*Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany*
[3]*Applied Bioinformatics, Bioscience, Plant Sciences Group, Wageningen University and Research Centre, Wageningen, The Netherlands*
[4]*Forschungszentrum Jülich, Jülich, Germany*
[5]*EMBL-EBI, Hinxton, United Kingdom*
[6]*Poznań University of Life Sciences, Poznań, Poland*
[7]*Keygene N.V., Wageningen, The Netherlands*
[8]*INRA-URGI, Versaille, France*
[9]*RWTH Aachen, Institute of Biology I, Aachen, Germany*
∗ *pkra@igr.poznan.pl*

Recent methodological developments in plant phenotyping, as well as the growing importance of its applications in plant science and breeding, are resulting in a fast accumulation of multi-dimensional data. There is great potential for expediting both discovery and application if these data are made publicly available for analysis. However, collection and storage of phenotypic observations is not yet sufficiently governed by standards that would assure interoperability among data providers and precisely link specific phenotypes and associated genomic sequence information. This lack of standards is mainly a result of a large variability of phenotyping protocols, the multitude of phenotypic traits that are measured, and the dependence of these traits on the environment. In the presentation we will discuss the current situation of standardization in the area of phenomics, point out the problems and shortages, and present the areas that would benefit from improvement in this field [1]. Besides, the foundations of the work that could revise the situation will be proposed, and practical solutions developed by the authors (based on the Minimum Information approach and the ISA-TAB format [2]) will be introduced.

[1] Krajewski P., Chen D., Ćwiek H. et al. (2015). Towards recommendations for metadata and data handling in plant phenotyping. J. Exp. Botany, in press.

[2] Sansone S-A., Rocca-Serra P., Field D. et al. (2012). Toward interoperable bioscience data. Nature Genetics 44:121-126.

The work was supported by the European Union under the FP7 thematic priority Infrastructures, Project No. 283496 transPLANT, Project No. 284443 EPPN, as well as in part by Project No. 244374 DROPS.

Discovery of Transcript and Metabolite Markers for the Breeding of Drought Tolerant Potato Cultivars using High Throughput Technologies

Heike Sprenger[1], Axel Fischer[1], Christian Schudoma[2], Alexander Erban[1], Katharina Rudack[3], Sylvia Seddig[3], Ellen Zuther[1], Dirk Walther[1], Joachim Kopka[1], Karin Köhl[1], Dirk K. Hincha[1]*

[1] Max-Planck-Institute of Molecular Plant Physiology, Potsdam, Germany
[2] The Sainsbury Laboratory, Norwich, UK
[3] Julius-Kühn-Institute, Groß Lüsewitz, Germany
** hsprenger@mpimp-golm.mpg.de*

Economic cultivation of starch potatoes (*Solanum tuberosum*) requires stable yields, which depend substantially on weather conditions. Because climate change scenarios and recent observations indicate increased likelihood of drought in early summer in Central Europe, potato cultivars need to be more drought-tolerant in the future. To accelerate breeding, we aim to develop molecular markers under both controlled and field conditions.
Initially, four potato cultivars of known drought tolerance were selected for drought stress experiments under greenhouse and field conditions. Plants were phenotyped for developmental stage, relative water content, biomass and yield parameters. Leaf samples of three independent greenhouse and field trials each, from two locations in Germany were taken in 2011 and 2012 to perform metabolite profiling by gas chromatography-mass spectrometry (GC-MS) and transcript profiling by Next Generation Sequencing (Illumina RNA-Seq).
We discovered differences in metabolite and transcript levels between control and drought conditions with high statistical confidence and identified transcripts and metabolites that might be responsible for higher drought tolerance in the selected potato cultivars. Therefore, our data analyses provide potential metabolite and transcript markers for drought tolerance and contribute to the molecular and physiological understanding of mechanisms underlying plant drought tolerance.
In parallel, we conducted drought stress experiments under greenhouse and rain-out shelter conditions on 30 additional cultivars of unknown tolerance. All cultivars were additionally grown on eight field sites across Germany under normal agricultural production conditions. The cultivars were scored for their drought tolerance based on starch yield stability. Leaf samples from these plants were used to test the validity of transcript and metabolite

marker candidates identified in the first phase of the project. Furthermore, the application of machine learning methods to transcript and metabolite profiling data enabled the prediction of drought tolerance from 20 selected molecular markers. With this approach we aim to provide validated molecular markers for marker assisted breeding to improve the performance of starch potatoes under drought conditions.

Session 8

Exploitation of genetic resources

Accelerate genetic gain by taking advantage of additional data sources and integrated data analysis – case studies from maize and wheat breeding at CIMMYT

Jens Riis-Jacobsen[1], José Crossa[1], Juan Burgueño[1], Kai Sonder[1], Maria Tattaris[1], Victor Jun Ulat[1]*
[1]CIMMYT
[*]J.Riis@cgiar.org

Background: Exploitation of plant genetic resources is dependent on germplasm related data being transformed into useful information that supports decision making and enhances genetic gain. Traditional data sources in the breeding work are genealogy and phenotypic data, but in recent years additional sources such as genotypic data, environmental data, and different sensor data have become available at a low cost. Nevertheless so far, only the largest breeding companies have managed to take advantage of the new sources of data in integrated informatics and analytics platforms, and the majority of organizations involved in plant breeding are struggling with how to harness this potential. Taking the maize and wheat breeding programs at CIMMYT as the point of departure this paper analyses how a small to medium sized breeding institution can take advantage of new data sources, what benefits they may obtain, and what some of the challenges involved are.

Findings:

- Genealogy and phenotypic data remains the foundation data for crop genetic improvement, and with available tools it is possible to setup the core elements of a future integrated breeding information system
- Genotyping, climate, and remote sensing data can make valuable contributions in a plant breeding program as has been demonstrated with ad hoc studies, but tools that facilitate mainstreaming of this in plant breeding programs are not generally available
- While the informatics and biometric challenges in an integrated breeding platform are being addressed, plant breeding institutions will still be faced with challenges related to establishing a multidisciplinary team as well as change management capabilities that can implement the solutions

Conclusions: New data sources and new analytical capabilities like high throughput phenotyping can accelerate genetic gain, and plant breeding programs that incorporate these may benefit from added productivity. Informatics and biometric solutions are increasingly available, which will lower the barriers for using integrated approaches. However, the full potential will only be realized when breeding activities and investments are reorganized and smaller breeding programs may also struggle to access the broader set of competencies required.

A high density genetic linkage map for wheat using an eight-founder MAGIC population

Keith Gardner[1], Lukas Wittern[1,2], Ian Mackay[1]*

[1]NIAB, Huntingdon Road, Cambridge, CB3 0LE
[2] University of Cambridge, Downing Street, Cambridge CB2 3EA
∗ presenting author, email: keith.gardner@niab.com

The multiparent advanced generation intercross (MAGIC) is a mapping population created by several generations of intercrossing among multiple founder lines. Such populations contain greater diversity than bi-parental mapping populations, as well as increased opportunities for recombination and generation of novel haplotypes during the crossing scheme. The NIAB MAGIC population is based on eight U.K. winter wheat varieties chosen in consultation with U.K. wheat breeders to capture traits of importance among lines of use in contemporary breeding programmes. We have developed a high density genetic linkage map for the NIAB MAGIC population, based on the iSelect 90K SNP array. By comparison to a comprehensive UK association mapping panel, the MAGIC map captures more than 80% of the SNP diversity found in UK wheat germplasm.We describe the challenges involved in the construction of the map, including the need for a significant amount of manual re-scoring to improve the accuracy of SNP calls on the array. We demonstrate the value of this new map for improved trait mapping in wheat by comparison to published wheat genetic maps, and using examples of quantitative traits analyzed within the NIAB MAGIC population and in association mapping studies. Finally, we use the map to demonstrate how our MAGIC population provides novel insights into patterns of genomic diversity in UK wheat, in particular the identification of linkage blocks potentially under fertility or viability selection.

Modelling the genetic architecture of flowering time control in barley through nested association mapping

Andreas Maurer[1], Vera Draba[1,2], Yong Jiang[3], Florian Schnaithmann[1], Rajiv Sharma[3,4], Erika Schumann[1], Benjamin Kilian[3,5], Jochen C. Reif[3] & Klaus Pillen[1]*
[1] Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Betty-Heimann-Str. 3, 06120 Halle, Germany
[2] Interdisciplinary Center for Crop Plant Research (IZN), Betty-Heimann-Str. 3, 06120 Halle, Germany
[3] Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Stadt Seeland, OT Gatersleben, Germany
[4] Current address: University of Dundee at The James Hutton Institute, Invergowrie, Dundee DD2 5DA, United Kingdom
[5] Current address: Bayer CropScience NV, Technologiepark 38, 9052 Ghent, Belgium
∗ andreas.maurer@landw.uni-halle.de

Flowering time is a major agronomic trait determining yield potential and yield stability of crop plants. We developed the first nested association mapping (NAM) population, HEB-25, in an autogamous crop species, barley, and used it to study biodiversity and to dissect the genetic architecture of flowering time. Upon cultivation of 1,420 lines in multi-field trials and applying a genome-wide association study, eight major quantitative trait loci (QTL) were identified to control flowering time. These eight QTL accounted for 64% of the cross-validated proportion of explained genotypic variance (p_G) identifying them as main determinants of flowering time in barley (Maurer et al. 2015). These findings represented a first step towards the development of a causal model for flowering time regulation in barley. We could furthermore show that inclusion of haplotype information is one key to further improve the model fit. However, defining and modelling haplotypes is still a challenging issue. We discuss several techniques of efficiently capturing this information in our highly diverse germplasm. We conclude that HEB-25 and the exotic biodiversity present therein are valuable toolboxes to support dissecting the genetic architecture of important agronomic traits and to replenish the elite barley breeding pool with favorable, trait-improving exotic alleles.

Maurer, Andreas, et al. "Modelling the genetic architecture of flowering time control in barley through nested association mapping." *BMC genomics* 16.1 (2015): 290.

Session 9

Experimental design and analysis of phenotyping trials

Experimental designs for expensive multi-phase traits

Alison Smith[1], Daniel Tolhurst[2], Steve Jefferies[3], Brian Cullis[2,4]

[1]National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, NSW, Australia. Email: alismith@uow.edu.au*
[2]National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, NSW, Australia
[3]Australian Grain Technologies, Glen Osmond, SA, Australia
[4]CSIRO, Canberra, ACT, Australia

The importance of sound experimental designs for plant breeding trials cannot be underestimated. They are crucial to ensure valid inference and accurate prediction of genetic effects, whether they be effects for traditional or genomic selection or the identification of QTL. Many key traits involve multi-phase experiments, where grain samples are taken from a field experiment (Phase I) then processed further in one or more laboratory experiments (Phase II and higher). Typically the laboratory phases are costly relative to the field phase and this necessitates a limit on the total number of samples that can be tested. Historically this has been achieved by sacrificing field replication and testing a single composite sample for each variety, obtained by combining grain from all field replicates. Typically no replication or randomisation is employed in the laboratory phases. In this talk we describe the approach of Smith et. al. (2014) in which replication is achieved in all phases of the experiment. In terms of field replication, some varieties are tested as composite samples and some as individual replicate samples. Replication in the laboratory is achieved by splitting a relatively small number of field samples into sub-samples for separate processing. Model-based design techniques are used to obtain efficient designs for the laboratory phases, conditional upon the field design. Unlike the historical approach, this method allows the application of an efficient statistical analysis to the resultant data so that accurate predictions of genetic effects may be obtained.

The approach will be illustrated using an Australian wheat quality project that involved a series of field trials and subsequent measurement of a range of flour, dough and end-product traits.
A major challenge with this project was to develop experimental designs and protocols that were not only statistically valid, but also satisfied strict budgetary constraints and were pragmatic, in the sense of complying with

standard laboratory practice. We show how all of these issues were successfully addressed using the approach of Smith et al. (2014).

A.B. Smith, D.G. Butler, C.R. Cavanagh and B.R. Cullis. 2014. Multi-phase variety trials using both composite and individual replicate samples: a model-based design approach. The Journal of Agricultural Science, available on CJO2014. doi:10.1017/S0021859614000707.

The generation of efficient row-column designs for field trials

Piepho HP[1], Williams ER[2], Michel V[3]*

[1]Biostatistics Unit, Universität Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany
[2]Statistical Consulting Unit, Australian National University, Canberra, ACT 0200, Australia
[3]Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern, Dorfplatz 1, 18276 Gülzow, Germany
∗ piepho@uni-hohenheim.de

When generating experimental designs for field trials laid out on a rectangular grid of plots, it is useful to allow for blocking in both rows and columns. A potential problem with such designs is that occasionally treatment replications may be clustered in the field layout. This talk reviews strategy to avoid such clustering. When the design is resolvable, separation can be enhanced by latinizing or t-latinzing the design. When the design is non-resolvable, latinization is not possible, and randomized classical row-column designs may occasionally involve clustered placement of several replications of a treatment. In this talk we illustrate how a spatial variance-covariance structure can be used to achieve a more even distribution of treatments across the field and how such designs compare with classical row-column designs in terms of efficiency. We consider both equally and unequally replicated designs, including partially replicated designs.

References

Piepho, H.P., Michel, V., Williams, E.R. (2015): Beyond Latin squares: A brief tour to row-column designs. *Agronomy Journal* (under revision)
Piepho, H.P., Michel, V., Williams, E.R. (2015): Non-resolvable two-dimensional field layouts with an even distribution of treatment replications (in preparation).

P-spline models for spatial variation in agricultural field trials

Fred van Eeuwijk, María Xosé Rodríguez-Álvarez, Martin Boer, Paul Eilers*
Biometris, Wageningen UR
** fred.vaneeuwijk@wur.nl*

Traditionally, the modelling of the spatial effects in field trials has used correlated random noise. To make this work, simplifying assumptions, like a separable autoregressive structure, have to be made. In practice a lot of tuning is needed to get results.

We propose to model the spatial field with tensor products of B-splines. In the spirit of P-splines we use anisotropic difference penalties to tune smoothness. A special reformulation leads to six additive spatial components. They form the basis of a mixed model with six unknown variance parameters. On top of the field come fixed or random effects for genotypes, rows and columns and possibly sub-regions of the field.

Although the model contains up to ten variance components, estimation is stable and fast, using an EM-like algorithm. Hundreds of fields, with thousands of plots each, have been analyzed without human intervention. A package for R is available.

Our experience has shown that the effective dimensions connected to the model components play a crucial role in the estimation process. They are also very useful to summarize results. We illustrate our work with challenging experimental data.

Session 10

Various themes

Meta-alleles in multiparental populations

Bevan Emma Huang[1], Jorge Berny[2], Paula X. Kover[3], Richard Mott[4]*

[1]CSIRO, Digital Productivity Flagship, Brisbane QLD 4001 Australia
[2]University of California at Davis, Dept. of Plant Sciences, Davis CA 95616 USA
[3]University of Bath, Dept. of Biology and Biochemistry, Bath BA2 7AY UK
[4]University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN UK
[*]b.emma.huang@gmail.com

Multiparental populations have become increasingly popular in plant breeding due to their high genotypic and phenotypic diversity. In particular, MAGIC populations, which mix the genomes of multiple founders through several generations of recombination, offer relatively high resolution and power for investigation of many traits simultaneously.
Typically, models for QTL mapping in such populations follow two approaches, testing association either with the observed marker genotypes or the unobserved founder genotypes. If there is a single causal variant and it is genotyped, or is in strong linkage disequilibrium with a genotyped marker, then the first, simpler model is the most powerful possible test. The second, full model, allows each founder of the population to have a different effect, thereby allowing for multiple causal variants. However, it may be over-specified since it is unlikely that all founders have different effects.
Models intermediate in complexity that elucidate the number of distinct functional alleles should better represent the true genetic architecture of the trait, particularly in testing for interactions, where the number of effects in the full model can quickly outnumber the size of the population.
We consider here three approaches to collapsing founder alleles into 'meta-alleles'. The first, based on clustering haplotypes in sliding windows based on genomic similarity was proposed by Leroux et al. (2014). This data-driven approach was shown to have highest impact in a scenario with a huge number of medium/small-size families. We propose two alternate approaches with biological interpretations of the meta-alleles which may be more appropriate for MAGIC populations. One determines the set of distinct isoforms of each protein encoded by the founders of the population. These "protein alleles" are used to cluster the founders. The other clusters founders based on time to the most recent common ancestor.
We compare all three approaches to the simple and full models through application to a four-parent wheat and 19-parent Arabidopsis MAGIC

population. Further, we perform simulations based on the Arabidopsis population to quantify the gains achievable through use of these methods.

Leroux, D., A. Rahmani, S. Jasson, M. Ventelon, F. Louis *et al.,* 2014 Clusthaplo: a plug-in for MCQTL to enhance QTL detection using ancestral alleles in multi-cross design. Theor Appl Genet 127:921-933.

High-density linkage mapping and the double-reduction landscape in tetraploid potato

Peter M. Bourke[1], Roeland E. Voorrips[1], Chris Maliepaard[1]*
[1]Wageningen UR Plant Breeding, P.O. Box 386, Droevendaalsesteeg 1, 6708PB, Wageningen.
[*]peter.bourke@wur.nl

Linkage mapping is an important tool in the identification of causative loci for traits of interest in crop species. In polyploid species, the creation of linkage maps is somewhat limited by the lack of available software to create high-density maps. Autopolyploid species such as tetraploid potato (*Solanum tuberosum* L.) may also exhibit additional features such as double reduction, not normally encountered in diploid or allopolyploid species. In this study we genotyped an F1 population of potato of 235 individuals using the 18K SolSTW SNP array, resulting in 3273 mapped 1:1 segregating markers across 96 separate homologue maps using JoinMap 4.1. Double reduction segments in the dataset were identified, allowing a comparison of the rate of double reduction versus distance from the centromeres. Simulations allowed us to predict the frequency of multivalents required to generate similar levels of double reduction. We also tested the effect that multivalents and preferential pairing have on linkage mapping, showing that mapping based on the simplifying assumption of random bivalent pairing is relatively robust even when these assumptions are violated to some degree.

Multilocus haplotype reconstruction in outcrossing tetraploids

Chaozhi Zheng[1], Roeland E. Voorrips[2], Hans Jansen[1], Fred A. van Eeuwijk[1], Christine Hackett[3], Julie Ho[4], and Marco Bink[1]*

[1] Biometris, Wageningen University and Research Centre, Wageningen, The Netherlands
[2] Plant Breeding, Wageningen University and Research Centre, Wageningen, The Netherlands
[3] Biomathematics and Statistics Scotland, Dundee, Scotland
[4] Forage Genetics International Inc., Davis, CA USA

∗ chaozhi.zheng@wur.nl

Many plants such as potato and some animals such as salmon are polyploid. Unveiling the genetic architecture of complex traits is key for the genetic improvement of polyploids in agriculture. F1 progenies of a bi-parental cross are often used for quantitative trait loci (QTL) mapping in outcrossing polyploids, where haplotype reconstruction by identifying the parental origins of marker alleles is necessary. In this presentation, we build a statistical framework for multi-locus haplotype reconstructions in a full-sib tetraploid family from bi-allelic marker dosage data collected from single nucleotide polymorphism (SNP) arrays or next generation sequencing technology, given a genetic linkage map. Compared to diploids, in tetraploids additional complexity needs to be addressed including double reduction and possible preferential pairing of chromosomes. We evaluated our method by both simulation studies and real data from potato. We showed that the parental linkage phasing is robust to complex chromosome pairing behaviors during meiosis, to various marker segregation types, and to erroneous genetic maps, and that the subsequent probabilistic ancestral inference is accurate.

Poster session

Genome wide association study for maize flour aroma-related aldehydes

Mara Lisa Alves[1], Marcos Malosetti[2], Maria Belo[1], Andreia Bento da Silva[1], Elsa Mecha[1], Manuel Paulo[3], Cláudia Brites[3], Pedro Mendes Moreira[1,3], Maria do Rosário Bronze[1], Fred van Eeuwijk[2], Maria Carlota Vaz Patto[1]*

[1] Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157, Oeiras, Portugal. [2] Biometris-Applied Statistics, Wageningen University, Droevendaalsesteeg 1, 6708PB, Wageningen, The Netherlands. [3] Escola Superior Agrária de Coimbra - Instituto Politécnico de Coimbra | Centro de Estudos de Recursos Naturais Ambiente e Sociedade, Bencanta, 3045 601 Coimbra, Portugal.
∗mara@itqb.unl.pt

Broa is a Portuguese ethnic bread consisting of a mixture of maize flour (~70%) with wheat or rye. Traditionally this bread is baked using local maize open pollinated varieties, which adds value to the final product. Aroma strongly influences consumer preference and acceptance of baked goods and aldehydes have been identified as the main volatile compounds that contribute to aroma in cereals (Klensporf and Jelén, 2005). Indeed, six different aldehydes were identified as the main aroma-related volatiles in Portuguese maize open pollinated varieties (hexanal, heptenal, 2-heptanal (Z), nonanal, 2- nonenal (E) and decanal). The difficulty to track these compounds using traditional breeding methodologies conveyed the need to use other approaches such as marker assisted selection to breed for aroma-related volatiles. The present work aims to identify genomic regions controlling maize flour aroma to be explored on future maize breeding. A collection of 132 maize inbred lines was analyzed by gas chromatography coupled with mass spectrometry. This collection represents a uniquely composed germplasm panel as it contains inbred lines developed entirely from Portuguese open pollinated varieties. A genome-wide association study using phenotypic information from two years of field experiments, combined with the genotypic information from a 50K maize SNP array is currently undergoing. Marker-trait associations will be tested using a mixed linear model accounting for the genetic relatedness among inbred lines (kinship). A realized kinship matrix was already estimated using a subset of 2000 markers evenly-distributed over the genome. Results obtained from this analysis will be reported.

References

Klensporf D. and H. H. Jelén (2005). Analysis of volatile aldehydes in oat flakes by SPME-GC/MS. *Pol. J. Food Nutr. Sci.*, **14** (4): 389-395.

Multi-environment GWAS for drought tolerance in maize

Carina de Oliveira Anoni[1], Martin Boer[2], Marcos Malosetti[2], Maria Marta Pastina[3], Kaio Olímpio das Graças Dias[4], Sidney Netto Parentoni[3], Paulo Evaristo de Oliveira Guimarães[3], Edson Alves Bastos[5], Milton José Cardoso[5], Jurandir Vieira de Magalhães[3], Antonio Augusto Franco Garcia[1], Claudia Teixeira Guimarães[3], Lauro José Moreira Guimarães[3], Fred van Eeuwijk[2]*

[1]Luiz de Queiroz "College of Agriculture" - University of São Paulo, Brazil
[2]Wageningen University and Research Centre, The Netherlands
[3]Embrapa Maize and Sorghum, Brazil [4]Federal University of Lavras, Brazil
[4]Embrapa Mid-North, Brazil
[*]canoni@usp.br

Drought is a severe stress factor in maize production and causes significant reduction in grain yield. Genetic architecture of drought tolerance is complex and a better understanding of this trait is required. We used an association mapping panel of 190 inbred lines (52 classified as Dent, 17 as Flint, and 121 unclassified). The panel was genotyped with ~500K SNPs. Inbred lines were testcrossed with Dent and Flint testers and assessed in well-watered and water- stressed conditions in two years and two locations in Brazil. Grain yield, flowering time, plant and ear height were measured. First, mixed model analysis was performed per trial to obtain adjusted means per line. SNPs were used to estimate genetic relatedness among lines, and to classify lines as Dent or Flint. The classification was confirmed with the test cross performance of the lines when crossed with the Dent or Flint tester. The resulting classification and kinship information was used in a multi-environment association mapping mixed model. Heritability ranged from moderate to high for traits and environments, and genetic correlations between environments was low in general, but lower under water-stressed condition. Phenotypic and genotypic information allowed assigning most of the unclassified lines to the Dent or Flint heterotic group. Our mixed model analyses identified useful markers in Brazilian elite maize germplasm that were related to drought tolerance.

The prediction ability of genome wide selection (GWS) over inbreeding populations

Ricardo Augusto Diniz Cabral Ferreira[1], Isabela de Castro Sant'Anna[1], Vinicius Quintão Carneiro[1], Haroldo Silva Rodrigues [1], Joao Filipi Guimarães[1], Luciano Rogério Braatz de Andrade[1], Gabi Nunes Silva[2], Moysés Nascimento[2] , Marciane Silva Oliveira[1], Cosme Damião Cruz [(1,2)]*

[1] Federal University of Viçosa – Dep. of Biology
[2] Federal University of Viçosa – Dep. of Applied Statistics and Biometrics
**ricardo.cabral.agronomia@gmail.com*

Background: The advantage of GWS is the use of markers effects to select the individuals that were not phenotyped. The aim of this study was evaluated the capacity prediction of the GWS in different kinds of inbreeding population. It was simulated a genome of 10.1 Morgans with 10 linkage groups and one SNP per centimorgam. In each linkage group, two QTL's was simulated. The F_2 and four generation of inbreeding (S_1, S_2, S_3, S_4) was simulated with 500 individuals each. Twelve scenarios was created by the simulation of genetic and phenotype effects: 1) heritability (h^2) of 20% and average degree of dominance (*d/a*) of zero; 2) h^2 = 40% and *d/a*=zero; 3) h^2 = 60% and *d/a*=zero;4) h^2 = 80% and *d/a*=zero; 5) h^2 = 20% and *d/a*=0.5; 6) h^2 = 40% and *d/a*=0.5; 7) h^2 = 60% and *d/a*=0.5; 8) h^2 = 80% and *d/a*=0.5; 9) h^2 = 20% and *d/a*=1.0; 10) h^2 = 40% and *d/a*=1.0; 11) h^2 = 60% and *d/a*=1.0; 12) h^2 = 80% and *d/a*=1.0. To evaluated de capacity of prediction, the model (GBLUP) was trained in one generation and used to predict the subsequent self-generation. The squared of the correlation between the EGBV and the true genetic value was used to measure the capacity of prediction.
Results: The capacity of the use the markers effects varied considerably among the scenarios. Independent of the population used to estimate the markers effects the capacity of the prediction was worst in the scenario of *d/a* was 1.0. In the scenario of the h^2 of 20% and 40% for *d/a* equal zero, F_2 population showed the best population to be used to estimate the markers effects and as the *d/a* increase the inbreeding generations had best prediction. When h^2 were 60% and 80% independent of *d/a*, the inbreeding generation showed be better than F_2.
Conclusions: The results suggest that the use of the markers effects can be used to predict the EGBV in inbreeding population. The capacity of prediction depends on the heritability, average degree of dominance and in which population the effects of markers are estimated.

Germplasm Evaluation for Drought Resistance in Wheat

Habtamu Ayalew[1, 2], Hui Lui [1], Guijun Yan[1]*

[1] School of Plant Biology, Faculty of Science and the UWA Institute of Agriculture, The University of Western Australia, Crawley, WA 6009, Australia,
[2] Department of Horticulture, College of Agriculture and Natural Resources, Debre Markos University, P.O.Box 269 Debre Markos, Ethiopia
[*] habtamu.tamir@research.uwa.edu.au.

Background
Drought is one of the major abiotic factors that are affecting wheat production worldwide. It can happen at any stage of the crop development. Breeding strategies need to consider evaluating germplasm at various growth stages and for both above ground and below ground growth. However, screening for root traits has been one of the most difficult areas to practice over large number of genotypes. Hydroponic systems enable easy access to roots while high molecular weight polyethylene glycol (PEG) is commonly used to induce water stress. In this particular study, a total of 832 genotypes were evaluated for root length in a hydroponic system both under osmotic stress and non-stress growing conditions. Augmented complete block design with seven blocks, and six standard checks was used. Stress level of -0.82 MPa was applied using PEG6000 and plant nutrition in the form of half strength Hoagland's solution.
Result
Statistical analysis showed highly significant difference ($P<0.001$) both under stress and non-stress conditions. Osmotic stress has caused an average reduction of 54% on root length. The best control variety for drought resistance was significantly ($p<0.05$) outperformed by four new entries namely Colotana 296-52, Compare, Santa Elena, and Tammarin Rock, while the shortest adjusted root length was measured on genotypes Aus 16356, Elia, Camm, Portugal 3, and Sentinel. The six control varieties didn't show any significant difference among themselves under non-stress condition.
Conclusion
There was a change in the ranking of genotypes under the two water regimes which indicated the difficulty of selecting new drought resistant varieties for drought areas under optimum growing environments. Crossing among the most contrasting genotypes was carried out to analyze the inheritance of root length under water stress and to study the nature of the different gene actions. The resulting populations are being advance to RILs for QTL mapping.

Genomic selection in a hybrid rye breeding program using historical data: An approach considering genotype by year interaction

Angela-Maria Bernal-Vasquez[1], Andres Gordillo[2], Malthe Schmidt[3] Hans-Peter Piepho[4]*

[1,4]Biostatistics Unit, Institute of Crop Science, Universität Hohenheim. Fruwirthstrasse 23, 70599 Stuttgart, Germany
[2,3]KWS LOCHOW GMBH. Ferdinand-von-Lochow-Strasse 5, 29303 Bergen, Germany
[*] bernalva@uni-hohenheim.de

Breeding programs produce a large amount of data year by year, usually performing disconnected trials across years. The presence of genotype-by-year interaction is to be expected. Multi-year information could be used in genomic selection procedures, using a proper model that ensures that the prediction procedure does not divert part of the marker information into prediction of that interaction, rather than of the breeding values. Several authors have proposed the use of covariance functions or environmental covariables to model genotype-by-environment interactions and incorporate this effect into the genomic selection model.
In this study, we propose the use of the kinship matrix to be able to dissect the main genotype effect from the genotype-by-year effect under the genomic selection frame. We use an historical dataset from a rye hybrid breeding program and study different scenarios varying the training set configuration and using a forward validation approach.

Recombination, founder contribution, power and precision in the MAGIC winter wheat population

Tryntsje Boersma[1,2], Ian Mackay[2], Marco C. A. M. Bink[1], Fred A. van Eeuwijk[1]*

[1]Biometris, Wageningen University and Research Centre, PO Box 100, 6700AC Wageningen, The Netherlands
[2]NIAB, Huntington Road, Cambridge, CB3 0LE, UK
[*] tryntsje.boersma@wur.nl

The Multi-parent Advanced Generation Inter Cross (MAGIC) is an experimental population for genetic analysis in which multiple founders are intercrossed over several generations before recombinant inbred lines are generated. The increased diversity and recombination in MAGIC make it ideally suited to study genetic architecture of complex traits. At NIAB a winter wheat MAGIC population has been generated from eight commercial varieties as founders (Mackay *et al*., 2014). High quality dense SNP (n=18,625) data for 643 genotypes have been analyzed to test recombination frequencies and founder contributions in the progeny. This is done in two ways. First, with RABBIT, a newly released method using a hidden Markov model framework to reconstruct genome ancestry blocks from SNP array data (Zheng *et al.*, 2014) and secondly, with the R-package HAPPY (Mott *et al.*, 2000). In addition we assess the added value when using the founder origin probabilities from different methods when mapping QTL for traits yield, height and, heading date.

Transformational Innovations in Plant Breeding

William Bourdoncle and Franck Chopin

Monsanto Europe

Plant breeding is a foundation component of Monsanto's Research and Development (R&D) pipeline and is a core driver to deliver farmers better performing products more sustainably. We are at a unique inflection point in the evolution of plant breeding where data science, genomic breeding methods and predictive analytics are beginning to unlock previously untapped genetic potential. With these and other technological advances as the foundation, Monsanto is focused on continuous improvement of plant breeding. These advances also allow us to evolve our R&D approach to increase year-over-year genetic gain. Learn more about the Monsanto Plant Breeding program.

A preliminary model comparison of genomic predictions in diploid *Actinidia chinensis* (kiwifruit) seedlings

Can-Hong Cheng, Paul M. Datson, Elena Hilario, Cecilia Deng, Lorna Barron, Kelvina I. Manako Mark McNeilage*
The New Zealand Institute for Plant & Food Research Limited
Mount Albert Research Centre, Private Bag 92 169, Auckland, New Zealand
** Canhong.cheng@plantandfood.co.nz*

Kiwifruit are dioecious, long-lived, perennial, woody vines. The consequences for kiwifruit breeding include long breeding cycles, large planting areas, and having to estimate breeding values for fruit traits of potential male parents by female progeny performance, which is expensive and time-consuming. Also, the majority of economically important traits, such as yield, fruit quality and disease resistance, are complex because of their polygenic (quantitative) nature. Genomic selection in kiwifruit has arisen from the conjunction of development of restriction enzyme-based genotyping-by-sequencing (GBS) and new statistical methods to analyse the genotypic data. However, information on model robustness and guidelines for model selection are lacking in kiwifruit. In this study, 88 seedlings were randomly selected from a diploid *Actinidia chinensis* breeding population designed to breed kiwifruit with red-centred flesh. Of these, 78 were used as a training set and 10 were used as a validation set. GBS libraries were constructed using the restriction enzyme *Bam*HI and sequenced on Illumina HiSeq2000 platform in SE mode. A total of 8372 SNP markers were obtained on all 88 individuals. Three statistical models, RR-BLUP (ridge regression best linear unbiased predictor), GBLUP (genomic best linear unbiased predictor), and LASSO (least absolute shrinkage and selection operator), were used to estimate GEBVs (genomic estimated breeding values) for nine traits. The correlations between observed phenotypic data and predicted GEBVs of the validation set were calculated to measure the reliabilities of the predicted GEBVs. The correlations ranged from 0.48 to 0.91 for red intensity, fruit weight, yield, and fruit firmness at 60, 90 and 120 days of storage, whereas negative correlations were obtained for dry matter, soluble solids content and *Pseudomonas syringae* pv. *actinidiae* (Psa) resistance. Possible reasons for the negative correlations could have been the small number of individuals in the training data set and errors in the genotypes close to genes with major effects. Our results also showed that the reliabilities of the predicted GEBVs from different models for the same trait were highly correlated.

Data processing and visualization tool for plant breeding.

Meklit Chernet[1], Sebastiaan Vandewoude[1], Pieter Pypers[1] and Michele Malaise[1]*

[1] CropDesign N.V. - A BASF Plant Science Company
[*]meklit.chernet@cropdesign.com

Background
In plant breeding, data analysis and presentation of statistical results in an easily interpretable manner is crucial in performance evaluation and selection. The efficiency in doing so receives even higher attention in a high-throughput screening. Using available R and shiny tools, we have developed a semi-automated online system for data quality checks, statistical analysis and reporting.

Results
The interactive tool allows users to define the scope of the data to look at being univariate versus multivariate or at/across locations. This approach is valuable in data quality assessment such as outlier detection, understanding relationships between variables and assessing stability/variability. The tool includes data analysis features where several mixed effects models fitted to the data can be compared next to each other. The reporting system presents the complex statistical analysis results in simplified and easily interpretable outputs. We use interactive graphs that scientists can use to make selections in a multivariate, multi-location and multi-season environment. Selection criteria can be tuned to different purposes with next to none effort.

Conclusion
This system provides an intuitive and interactive way of working with data. The experience we have with this system highlights the higher efficiency obtained and the versatility of the possibilities we can offer to breeders and agricultural scientists to look at the result of data analysis. The tool allows scientists to obtain deeper understanding of interactions and answers questions even beyond the primary research question.

Evaluating the Potential of Genome-Wide Selection for Sugarcane Breeding: A Proof of Concept

De Bem, I.O.[1]; Mendes, C.A.[1]; Ribeiro, S.B.[1]; Alves, M.[1]; Nunes, C.M.C.[2]; Coelho, A.S.G.[1]*

[1] Escola de Agronomia, Universidade Federal de Goiás, Goiânia-GO, Brazil
[2] Centro Universitário de Goiás, Uni-Anhanguera, Goiânia-GO, Brazil
[*] ivonedebem@gmail.com

Background: The use of genetic markers in sugarcane breeding still represents a big challenge. Genetics of sugarcane is widely acknowledged as one of the most complex among crops due to the interspecific origin of its modern cultivars, its high and variable ploidy levels, the presence of aneuploidy, and its huge genome. The potential use of genomic-enabled predictions, derived from the development of Genome Wide Selection (GWS) methods, has been proven in different animal and plant breeding programs. In order to study the feasibility of GWS in sugarcane we evaluated models obtained by six different methods, including regression and Bayesian methods, under a strong linkage disequilibrium condition. A population, comprised of 157 individuals derived from selfing (88) and biparental crosses (69), was phenotyped for six traits, related to plant growth, biomass production and sugar content, using an augmented block design. A total of 7680 DArT markers were genotyped for each individual. GWS models were developed using the BGLR package and evaluated using the ten-fold cross validation approach. The accuracy was accessed through the Pearson correlation coefficient between breeding values and their genomic-enabled predictions.

Results: High values of accuracy were obtained for traits with high levels of genetic variance, ranging from 0.77 to 0.95. Near zero values of accuracy were obtained for traits with low genetic variance. Very similar results were obtained by the six different evaluated methods. A strong effect of the genetic structure between the two subpopulations (selfing and biparental crosses) was detected in the accuracies estimates.

Conclusions: Results illustrate the potential levels of accuracy of GWS models for sugarcane breeding populations. Since maximum linkage disequilibrium was induced, the accuracies estimates should be considered optimal for similar experimental conditions. The similarity among results obtained by the different methods suggests a minor importance of the choice of method used to generate the GWS models. Results also showed that the inclusion of genetic structure effects is imperative for obtaining more realistic models.

Financial Support: FINEP/Ridesa/Petrobras/CAPES/CNPq

Estimation of additive and dominance effects in maize under well-watered and water-stressed conditions using high-density SNP markers and multi-environment trial analyses

Kaio Olímpio Das Graças Dias[1], Salvador Alejandro Gezan[2], Claudia Teixeira Guimarães[4], Sidney Netto Parentoni[4], Paulo Evaristo de Oliveira Guimarães[4], Edson Alves Bastos[4], Milton José Cardoso[4], Carina de Oliveira Anoni[3], João Candido de Souza[1], Jurandir Vieira de Magalhães[4], Maria Marta Pastina[4], Lauro José Moreira Guimarães[4]*

[1] *Universidade Federal de Lavras*
[2] *School of Forest Resources and Conservation, University of Florida*
[3] *Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo*
[4] *Embrapa Milho e Sorgo*
[*] *kaioolimpio@hotmail.com*

Background: Drought is one of the most common causes of severe crop production losses worldwide. In maize breeding programs, estimation of additive and dominance effects helps to understand the genetic architecture of target traits. In this study, additive (VA) and dominance (VD) variance components, broad and narrow sense heritabilities for grain yield (GY, ton/ha), ears per plot (EAR), anthesis-silking interval (ASI, days), female flowering time (FFT, days) and male flowering time (MFT, days) were estimated in a multi-environment trial (MET) analyses. Additive and dominance genomic relationship matrix for hybrids were obtained, based on approximately 50K SNP markers, and incorporated into the MET model. Traits were measured for 314 hybrids evaluated in eight sites, encompassing well-watered (WW) and water-stressed (WS) conditions, over two years and two locations in Brazil. The MET analysis was performed based on linear mixed model considering a random effect of hybrids within environments, and a unique genetic correlation coefficient between environments and specific genetic variance components for each environment. Results: Broad (H^2) and narrow (h^2) sense heritabilities varied considerably between traits and environments. For WS, H^2 and h^2 ranged from 0.25 (ASI) to 0.44 (MFT) and from 0.25 (ASI) to

0.38 (MFT), respectively. For WW these values, ranged from 0.20 (EAR) to 0.44 (ASI) and from 0.26 (EAR) to 0.48 (ASI and MFT), for H^2 and h^2, respectively. Estimates of VD/VA for WW and WS ranged from 0.39 (GY) to 0.12 (ASI) and from 0.51 (GY) to 0.19 (MFT), respectively.

Conclusions: Our results suggest that GY and EAR showed a high proportion of dominance effects in both conditions, while for ASI the dominance effects were more evident under WS condition. Differences were observed between WS and WW conditions for the VD/VA ratio for all traits. Low levels of genotype-by-environment interaction for FFT and MFT showed that it is possible to make successful selection across WW and WS conditions for both traits. These results contribute to a better understanding about the genetic architecture of important traits evaluated under WW and WS conditions in maize.

Numerical and graphical assessment of relationships between traits of the Iranian *Coriandrum sativum* L. core collection by considering genotype by irrigation interaction

Hamid Dehghani[1], Mostafa Khodadadi[1], Mokhtar Jalali-Javaran[1], Sajad Rashidi-Monfared[1], Jack (John) Christopher[2]*

[1] Plant Breeding and Biotechnology Department, Faculty of Agriculture, Tarbiat Modares University, Tehran, P.O. Box 14115-336, Iran

[2] Queensland Alliance for Agriculture and Food Innovation (QAAFI), Leslie Research Facility, The University of Queensland, PO Box 2282, Toowoomba, Qld 4350, Australia

[*] dehghanr@modares.ac.

Drought is a worldwide problem seriously influencing production of agricultural plants such as coriander, but development of tolerant genotypes is inhibited by a lack of effective selection criteria. The objectives of this study were to interpret the relationship between fruit yield and related traits in coriander to evaluate genotype by trait interaction by considering genotype × irrigation interaction. According to this aim, 16 Iranian endemic coriander genotypes were grown in a glasshouse under well watered (WW) and water stressed (WS) conditions. Structural equation modeling (SEM) was used, based on genotypic correlation coefficients and genotypic variance estimates obtained from combined analysis of WW and WS conditions and genotype-trait (GT) biplot analysis. Applying both types of analyze to the multiple trait data revealed that SEM and DC+YPr data based GT biplot results are similar. The GT biplot graphically displayed the interrelationships among traits and facilitated visual comparison of genotypes for selection. Predictors in the casual diagram could explain 70.8%, 80.9%, 100% and 41.7% of the total variation in fruit weight per plant (FWPP), fruit number per plant (FNPP), day to end of flowering (DTEOF) and umbel number per plant (UNPP), respectively. Also, the GT biplots explained 81.2% for WW data, 74% for WS data and 79.7% for DC+YPr data. It was found that selection for high SPAD chlorophyll content in the grain filling stage (SCCIGFS), UNPP, FNPP and branch number per plant (BNPP), and low shoot dry weight per plant (SDWPP) and DTEOF should be considered as priorities in breeding programs for coriander aiming for more productive and drought tolerant genotypes. Furthermore, G13 (TN-59-353) can be used as a drought tolerance donor in breeding programs.

Keywords: Coriander, Genotypic correlation, GT-Biplot, Structural equation modeling

Mapping of main and epistatic QTLs and QTL by environment interaction for rice grain yield and yield components under well watered and drought stress conditions

Hamid Dehghani[1], Mehdi Rahimi[1], Babak Rabiei[2], Ali R. Tarang[3]*
[1]- Department of Plant Breeding, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran.
[2]- Department of Agronomy and Plant Breeding, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran.
[3]- Agricultural Biotechnology Research Institute of North of Iran, Rasht, Iran.
[*]dehghanr@modares.ac.ir

Quantitative trait loci (QTLs) controlling yield and yield components were identified by using a F_5 population of 150 lines from a cross between 'Sepidroud' (*Indica*) and 'Gharib' (*Indica*) in normal and stress conditions. Based on the genetic linkage map containing 131 polymorphic SSR and 52 AFLP markers which covered 1063.14 cM of the rice genome with an average distance of 5.81 cM between markers, the main effect QTLs (M-QTLs) and epistatic QTLs (E-QTLs) for grain yield and its components were analyzed by QTL IciMapping software. Sixteen M-QTLs and three E-QTLs were identified for grain yield and its components in normal condition and 13 M-QTLs and 12 E-QTLs were found in stress condition for these traits including four M-QTLs and four E-QTLs for number of panicle per plant, three M-QTLs and three E-QTLs for grain yield, four M-QTLs and five E-QTLs for number of total spikelet per plant, six M-QTLs and two E-QTLs for number of filled grain per panicle, nine M-QTLs and one E-QTL for grain weight and three M-QTLs for spikelet fertility. The phenotypic variation explained by each M-QTL and E-QTL ranged from 5.86 to 22.87% and 6.64 to 18.86%, respectively. The QTL-by-environment interactions were detected by QTLNetwork program in the joint analyses of multi-environment phenotypic values. Thirteen M-QTLs and four E-QTLs were identified for the all studied traits. The phenotypic variation explained by each QTL and QTL-by-environment interaction ranged from 0.2 to 10.51% and 0.05 to 13.22%, respectively. As many as 29 M-QTLs were detected to be associated with yield and its components; some of them are being reported for the first time. Seven M-QTLs (*qNFGP4*, *qGW1*, *qGW3.1*, *qNTSP1* and *qNTSP11*) were identified with high phenotypic variance and could be considered to use in marker-assisted selection (MAS) programs for improving yield in rice. Also, the marker RM252, RM1, RM5800, RM202, RM42, E37M59-3 and E37M59-1 had a close linkage with these major QTLs and could be used in marker-assisted selection (MAS) programs.

Keywords*: Epistatic QTLs, Grain yield, Main QTLs, QTL-by-environment interaction, Rice.*

Abbreviations: NPP, number of panicle per plant; NTSP, number of total spikelet per panicle; NFGP, number of filled grain per panicle; SF, spikelet fertility; GW, 1000-grain weight; GY, grain yield; QTL, quantitative trait loci; M-QTL, Main QTLs; E-QTL, Epistatic QTLs.

Diallel cross analysis in melon (*Cucumis melo* L.) identification of best parents for yield and related traits

Zhila Jabbari[1]; Ehsan Feyzian[1]; Abdolmajid Rezai[2] and Mokhtar Jalali Javaran[1]*

[1] Department of Plant Breeding, Faculty of Agriculture, Tarbiat Modares University, P.O. Box 14115-136 Tehran, Iran.

[2] Department of Plant Breeding, Faculty of Agriculture, Isfahan University of Technology, P.O. Box Isfahan, Iran.

** jila7jabari@yahoo.ca*

A complete diallel cross study of 6 Iranian melon cultivar (Eyvanaki, Abasali, Tashkandi, Hose-sorkh, mashhadi and Nahavand) and one foreign cultivar (Ananasi) was carried out for number of fruits, average weight of fruits per plant, flesh thickness, yield and acceptable yield. The analysis of variance revealed that differences among the genotypes were significant for all characters. Additive gene effects were most important in governing fruit number per plant, average weight per fruit, flesh thickness and yield while dominance genetic effects mainly controlled acceptable yield. The reciprocal effect was only significant for average weight of fruits per plant. Nahavand was the best general combiner for average weight, flesh thickness and yield. Heterosis over the mid parent was found for all traits except number of fruits per plant while heterosis over the better parent was not found.

Keywords: *Cucumis melo L.*; Diallel; Heritability; Gene action; General combining ability; Specific combining ability.

The optimal design of experiments with a known genetic covariance structure: a case study of a nested association mapping population

A.M. Kelly[12] and D.G.Butler[12]*

[1] Department of Agriculture and Fisheries, Leslie Research Facility, Toowoomba, Queensland, Australia.
[2] Queensland Alliance for Agriculture and Food Industries, The University of Queensland, Brisbane, Australia.

[*] alison.kelly@daf.qld.gov.au

Evaluation of genetic material in field trials is an important part of the selection process in the early stages of plant breeding programs. It is well established that selection decisions are improved by incorporating information on genetic covariance into a linear mixed model analysis of these data. Information on genetic covariance in a breeding program traditionally arises from ancestral or pedigree information, defining an identity-by-descent relationship.

Phenotyping material for genetic analysis also relies on accurate determination of phenotypic performance in a field evaluation. In these studies, genetic covariance can be derived from molecular marker information, which maps the identity-by-state relationships. Of specific interest are nested association mapping populations, which are formed from a small set of diverse parents, or founders, used in a structured crossing program. For these populations, both marker and pedigree relationship data are usually available.

Experiments for the field evaluation of genotypes are typically designed assuming independent treatment (genotype) effects. This paper considers the design of experiments where there is a known genetic covariance structure among treatments. Here we consider the consequences for the optimal design of these experiments using a linear mixed model framework, and incorporating a known genetic relationship matrix. We show that the optimality criterion for design was minimised when incorporating genetic relatedness into the random allocation of genotypes to plots in a field trial.

Automatic IR laser scan technology for the identification of phenotypic markers for drought tolerance

Karin Köhl[1], Manuela Haas[1]*

[1] Max Planck Institute of Molecular Plant Physiology
[*] Koehl@mpimp-golm.mpg.de

Future crop yields will be increasingly reduced by drought stress. Global climate change models predict altered precipitation patterns that result in less reliable water supply during growth season. Yield security could be increased by breeding drought tolerant cultivars that produce stable yield at reduced water supply. Presently, breeding for drought tolerant crops is slowed down by the need to test many genotypes in field trials in the target environment. High-throughput screens might be achieved by marker assisted selection, if markers can be identify that predict drought tolerance. Genetic, transcript or metabolite analysis yield potential markers and can be performed in high-throughput workflows. Phenotypic markers may be more closely related to yield than transcript or metabolite markers and thus allow better predictions. However, phenotyping hitherto lacked high-throughput technics. In our project, we employed automatic infrared laser scans to gain time series of shoot surface growth in potato plants. To identify and validate phenotypic markers, the measurements were performed on a panel of 60 potato genotypes from a crossing population that segregated for drought tolerance. The panel was cultivated under optimal and reduced water supply in field trials and semi-controlled environments within the project VALDIS TROST. Within this project, the genotypes are characterized for drought tolerance on the basis of relative starch yield and for a range of metabolite and transcript markers (see contribution Sprenger et al.). Time-series measurements indicated that shoot surface growth is reduced under drought stress compared to optimal water supply. By time-series analysis of surface development, we aim to identify phenotypic markers with a significant genotype x environment interaction that allow predicting yield stability under drought stress from shoot phenotypes early in development.

Identifiability and Inference of Relatedness between Individuals

Fabien Laporte[1], Alain Charcosset[1], Tristan Mary-Huard[1,2]*
[1] UMR de Génétique Quantitative et Evolution, INRA/Univ. Paris-Sud/CNRS, 91190 Gif-sur-Yvette, France
[2] UMR 518 AgroParisTech/INRA, Paris 5e, France
[*] fabien.laporte@moulon.inra.fr

The relatedness between two individuals is a distribution of probabilities related to the number of alleles inherited from one or several common ancestors. It is an important concept in populations genetics, quantitative genetics and plant and animal breeding. For many years relatedness has been inferred from the pedigree infomation (Crow and Kimura 1970), but nowadays, thanks to genotyping techniques, relatedness between individuals of a panel can be inferred without pedigree. Genotyping based relatedness matrix has been used recently in genomic prediction.

In this work, we focus on the estimation strategy proposed by Milligan who models the relatedness distribution inference problem between two individuals as a mixture model. In this model, each observation corresponds to a marker, for which the observed variables are the 4 alleles (2 alleles for each individual), which define the Identity by State (IBS) mode of the marker. The hidden vari- ables are the ancestral origins of the alleles, which define Identity by Descent (IBD) mode of the marker. The objective is then to estimate the proportions of the different IBD modes over all available markers based on the IBS information. Maximum likelihood inference can be performed to estimate these proportions using the EM algorithm.

First we assess the identifiability of the model when inference is based on biallelic markers. Secondly, we propose several extensions of the work of Milligan. We extended the approach to phased haplotypes (i.e. the gametic origin of alleles is known) thanks to haplotypic reconstruction (in human genetics) or knowledge of parental inbred lines for plant hybrids. Moreover, it is now of common practice to consider multi-population panels. In such cases, allelic frequencies are specific to each population. This is explicitly taken into account within our methodology. Computational time and performances will be illustrated on simulated data.

Genomic selection for adaptation breeding: Whole-genome prediction of the response to environmental stress in bread wheat

Delphine Ly[1], Sylvie Huet[2], Arnaud Gauffreteau[3], Fabien Cormier[1,4], Renaud Rincent[1], Jean-Luc Jannink[5] and Gilles Charmet[1]*

[1]Institut National de la Recherche Agronomique - Blaise Pascal University, UMR1095, Genetics, Diversity and Ecophysiology of Cereals, 63039 Clermont-Ferrand, France
[2] Institut National de la Recherche Agronomique, MAIAGE – Applied Mathematics for Computer Sciences, Jouy-en-Josas, France
[3] Institut National de la Recherche Agronomique, UMR Agronomy, Thiverval Grignon, France
[4] Biogemma, Research center, Route d'Ennezat, Chappes, France
[5] USDA-ARS, R.W. Holley Center, 538 Tower Rd, and Cornell University, Section of Plant Breeding and Genetics, Ithaca, NY

**Delphine.ly@clermont.inra.fr*

In a climate change context, breeding for adaptation to environmental stresses has emerged as a major priority. Factorial Regression models the genotypic response to environmental stress covariates and is appropriate to predict adaptation. The contribution of genomics could enhance the predictive ability of the factorial regression model allowing predictions of untested genotypes' adaptation to untested environments. We propose a genomic extension of the factorial regression model, the FR-gBLUP (Factorial Regression-genomic Best Linear Unbiased Predictor) to model the genomic responses to an environmental stress. The predictive ability of our FR-gBLUP model and the standard additive gBLUP model for the response of grain number to nitrogen stress or to water deficit stress were compared in 4 cross-validation schemes, which simulated different combinations of tested or untested genotypes in tested or untested environments.
Comparing prediction accuracies of the FR-gBLUP model to the additive gBLUP, we showed gains in accuracy from 0 to 23% on real data. Highest gains were obtained when neither genotypes nor environments were in common between training and validation sets, in the case of the response to water deficit stress. We observed no decrease in accuracy for the response to nitrogen stress which explained only 4.2% of the genotype-by-environment interaction. Furthermore, simulation results showed indeed that greater

gains are achieved (from +11.64% to +53.20%) when the genomic response variance component is high (0.15 to 1.0 times the additive genetic variance).
To predict for new environments, the FR-gBLUP model is more advantageous than the additive BLUP, in terms of predictive ability. The FR-gBLUP is likely to be more effective in populations with high diversity of response to environmental stress, evaluated in diverse environments. Our study broadens the possibilities of adaptation breeding to genome-wide selection for adaptation.

A preliminary model comparison of genomic predictions in diploid *Actinidia chinensis* (kiwifruit) seedlings

Can-Hong Cheng, Paul M. Datson, Elena Hilario, Cecilia Deng, Lorna Barron, Kelvina I. Manako Mark McNeilage*
The New Zealand Institute for Plant & Food Research Limited
Mount Albert Research Centre, Private Bag 92 169, Auckland, New Zealand
** Canhong.cheng@plantandfood.co.nz*

Kiwifruit are dioecious, long-lived, perennial, woody vines. The consequences for kiwifruit breeding include long breeding cycles, large planting areas, and having to estimate breeding values for fruit traits of potential male parents by female progeny performance, which is expensive and time-consuming. Also, the majority of economically important traits, such as yield, fruit quality and disease resistance, are complex because of their polygenic (quantitative) nature. Genomic selection in kiwifruit has arisen from the conjunction of development of restriction enzyme-based genotyping-by-sequencing (GBS) and new statistical methods to analyse the genotypic data. However, information on model robustness and guidelines for model selection are lacking in kiwifruit.

In this study, 88 seedlings were randomly selected from a diploid *Actinidia chinensis* breeding population designed to breed kiwifruit with red-centred flesh. Of these, 78 were used as a training set and 10 were used as a validation set. GBS libraries were constructed using the restriction enzyme *Bam*HI and sequenced on Illumina HiSeq2000 platform in SE mode. A total of 8372 SNP markers were obtained on all 88 individuals. Three statistical models, RR-BLUP (ridge regression best linear unbiased predictor), GBLUP (genomic best linear unbiased predictor), and LASSO (least absolute shrinkage and selection operator), were used to estimate GEBVs (genomic estimated breeding values) for nine traits. The correlations between observed phenotypic data and predicted GEBVs of the validation set were calculated to measure the reliabilities of the predicted GEBVs. The correlations ranged from 0.48 to 0.91 for red intensity, fruit weight, yield, and fruit firmness at 60, 90 and 120 days of storage, whereas negative correlations were obtained for dry matter, soluble solids content and *Pseudomonas syringae* pv. *actinidiae* (Psa) resistance. Possible reasons for the negative correlations could have been the small number of individuals in the training data set and errors in the genotypes close to genes with major effects. Our results also showed that the reliabilities of the predicted GEBVs from different models for the same trait were highly correlated.

Small RNAs reflect environment from grandparental generations in apomictic dandelions

V. Preite[1], L. Morgado[2], J.F. de Carvalho[1], C. Oplaat[1], S. Anava[3], F. Johannes[2], O. Rechavi[3], K.J.F. Verhoeven[1]*

[1]Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands
[2]Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands
[3]Department of Neurobiology, Wise Faculty of Life Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel
[*] lionelmorgado@gmail.com

In plants, small RNAs (sRNAs) have important roles in transposable element silencing and in the post-transcriptional regulation of genes. Although it is known that sRNA abundance can be profoundly altered by environmental stressors, it remains unclear to which extent these changes can be inherited to subsequent generations that no longer experience the stressor.
To address this, plants of an apomictic dandelion lineage (*Taraxacum officinale hemicyclum*) were exposed to drought and salicylic acid stress and their progeny were grown unstressed for two generations. In the third generation, we deep-sequenced sRNA from both treatment groups as well as from controls (plants unstressed for three generations). We found consistent changes in the sRNA length composition in the treatment groups relative to the controls. The most pronounced changes occurred in sRNA of length 24nt and 21nt, which were significantly lower and higher in the treatment groups, respectively. Moreover, genes with the highest sRNA abundance changes, showed significant enrichment for GO terms concerning stress-related functions.
Our results demonstrate that ancestral environments can leave sRNA footprints at least three generations later, suggesting that sRNA play an important role in short-term adaptation to stressful environments.

Analysis of networks of differentially expressed genes for drought tolerance in Populus Nigra

Maria-João Paulo[1], Sabine Schnabe[1], Joost Keurentjes[2], Fred van Eeuwijk[1]*

[1] Biometris, Wageningen UR
[2] Laboratory of Genetics, Wageningen UR
** joao.paulo@wur.nl*

In an RNA-seq experiment in Populus Nigra, plants of three genotypes were grown under either drought or control condition. In the course of the experiment samples of seven distinct tissues were analysed for RNA-seq, and several static as well as dynamic phenotypes were measured.
We present the analysis of differentially expressed genes which was non trivial due to the complexity of the design. In addition we show preliminary result of the analysis of networks of differentially expressed genes, and discuss available methods to integrate different phenotypes and gene expression data.

A data analysis protocol for monitoring metabolomic changes in genetic experiments: a study of barley (*Hordeum vulgare*) leaves under drought stress

Aneta Sawikowska[*]*, Anna Piasecka, Piotr Kachlicki, Paweł Krajewski*

Institute of Plant Genetics PAS, Strzeszyńska 34, 60-479 Poznań, Poland
[*] *asaw@igr.poznan.pl*

Data processing methods applicable to large chromatographic data sets acquired in multifactorial experiments are presented. These methods lead from raw observations to the statistical descriptors allowing for interpretation of changes in identified compounds. We present an application of the algorithms to data obtained in a study of the reaction of barley RILs to drought. The protocol consists of several integrated and computationally optimized stages, some performed using publicly available software (R), some with our own scripts written for known algorithms, such as chromatogram alignment by correlation optimized warping [1, 2], and others with our own, new algorithms, such as for peak detection and deconvolution. The statistical approach is based on the analysis of variance performed with the help of the restricted maximum likelihood (REML) numerical procedures in Genstat package. To compare dependencies of metabolites under different conditions the correlation networks and differential correlation networks are constructed [3, 4]. The analysis includes also searching for quantitative trait loci using the method based on the mixed linear model [5]. The algorithms can be adapted to any chromatographic data.

References

[1] Nielsen N.P.V., Carstensen J.M., Smedsgaard J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. J. Chromatogr. A 805: 17-35.
[2] Skov T., van den Berg F., Tomasi G., Bro R. (2006). Automated alignment of chromatographic data. J. Chemom. 20: 484-497.
[3] Langfelder P., Horvath S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.
[4] Valcárcel B., Würtz P., Seich al Basatena N.-K., Tukiainen T., Kangas A.J., et al. (2011). A differential network approach to exploring differences between biological states: an application to prediabetes. PLoS ONE 6(9): e24702.

[5] Malosetti M., Ribaut J-M., van Eeuwijk F.A. (2013). The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. 4, 44.
This work was supported by the European Regional Development Fund through the Innovative Economy Program for Poland 2007-2013, project WND-POIG.01.03.01-00-101/08 POLAPGEN-BD „Biotechnological tools for breeding cereals with increased resistance to drought”.

Statistical Evaluation and Analysis of PACTS trials as a series of on-farm strip trials without replicates

Paul Schmidt[1], Hans-Peter Piepho*[1]

[1]*University of Hohenheim*
[*] *Schmidtpaul@hotmail.de*

In Germany, the performance evaluation of novel crop cultivars is done in the framework of *official trials* (OT, *Landessortenversuche*, a system of regional, replicated, small-plot, on-station trials) conducted by different German Federal States.
During a more recent development the concept of *head-to-head comparisons* emerged and was subject of controversial debates. Here, plots are laid out in large strips that can be hundreds of times larger compared to OT. The number of plots per field site is often reduced to <10. The number of tested varieties per location is smaller and treatments are unreplicated. Instead, resources are shifted towards raising the number of tested locations, which often leads to highly unbalanced data sets.
We had access to phenotypic data of rapeseed cultivars in Germany from 8 years of Pioneer's so-called Pioneer Accurate Crop Testing System (PACTS), which is a series of field trials that can be ascribed to the concept of head-to-head comparisons. Thus, this study had two major goals: (i) Examine and evaluate the discrepancies between the concepts of OT and head-to-head comparisons in statistical terms and (ii) after identifying a general framework to analyze the data with the help of mixed models, assess several model modifications and extensions terms of their potential benefit for the evaluation of variety performance.
As a conclusion it can be said that when single aspects are examined, head-to-head comparisons turn out to be beneficial in some and disadvantageous in other fields, but never unfeasible. Furthermore, contrasting genotypic and environmental variability found in this study to those found for official performance trials by Laidig et al. (2008) led to comparable results.
Secondly, after setting up a basis to allow for statistical analyses of the PACTS data and using the restricted maximum likelihood (REML) approach with mixed models, only the implication of information about respective cultivation areas as proposed by Graf et al. (2009) turned out to be truly promising, since both, its main as well as its interaction effect with varieties, were found to be significant. In all analyses, even for the basic model, differences between effects of tested varieties were found to be significant.

Graf, R., Michel, V., Roßberg, D. & Neukampf, R., 2009. Definition pflanzenartspezifischer Anbaugebiete für ein regionalisiertes Versuchswesen im Pflanzenbau. Journal für Kulturpflanzen, 61(7), pp. 247-253.
Laidig, F., Drobek, T. & Meyer, U., 2008. Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breeding 127, pp. 541-547.

Quantifying LD decay by quantile regression

Sabine K. Schnabel[1], Federico Torretta[2], Matthias Westhues[3]*

[1]Biometris, Wageningen University and Research Centre, The Netherlands
[2]Università di Palermo, Italy
[3]Universität Hohenheim, Germany
[*]sabine.schnabel@wur.nl

Through recent developments in genotyping more information on genetic markers become available. In order to perform powerful genome-wide association studies it is important to analyze linkage disequilibrium (LD) through pairwise comparisons of genetic markers. Large numbers of these markers pose new problems in terms of analysis and visualization. Here, we explore and quantify LD decay using monotone quantile regression. An example from a Maize population is used for illustration. The presented methods are currently applied also to other crops such as tomato and barley.

Genome-wide association studies have emerged as a great tool for the localization of QTLs in plant and animal breeding programs. However, a requirement to powerful GWAS is the investigation of the genetic relatedness (kinship matrix). For an appropriate kinship matrix, insight into LD between genetic markers is necessary. This matrix is best based on a set of independent markers. To find such a set of suitable markers we need to explore LD decay over the whole genome. This is commonly done globally per chromosome. We propose to investigate also the local
LD decay. LD is commonly measured in terms of the squared Pearson correlation coefficient R^2 between pairs of genetic markers. With increasing number of markers the number of comparisons makes modelling and especially visualization a challenge. Therefore we investigated different transformations of the correlation coefficients.
In our case study of how to explore and quantify local LD decay patterns in Maize we are using quantile regression with monotonicity constraints for a first summary of the LD decay. On top of that we are applying penalized splines to smooth the median local LD decay. These curves are easy to interpret.
While the presented steps are a good tool to quantify local LD decay, they have also been instrumental in identifying problems with the underlying genotypic data that have previously been overlooked. In this sense they can also serve as a diagnostic tool.

Exploring specific combining abilities in incomplete factorial cross designs with LABKEY software

Soula, Julie[1]; Vincourt, Patrick[2]; Bernard, Jean-Marc[3]; Bardet, Sébastien[1]; Royer, Frédéric[3]*

[1] Doriane SAS, 31 avenue Jean Médecin, 06000 Nice, France
[2] Open-Source-Biology, 12 avenue V. Segoffin, 31400 Toulouse, France
[3] Biosearch Data Management, 39 boulevard Dubouchage, 06000 Nice, France
**contact@doriane.com*

Plant breeders have to deal with several fields of quantitative interactions during their decision making process: for example genotype by environment (G*E) interactions, or, for hybrid crops, specific combining abilities (SCA) which represent the interactions between the male and female parents of each hybrid. Based on well-known concepts which, however, are not always implemented in a ready-to-use toolkit, LABKEY is offering solutions aiming at providing various tools to help the breeders to explore the space of these interactions. Among these tools, particular attention is paid to the AMMI (Additive Main Effect and Multiplicative Interaction) approach, which, in our case, amounts to performing Principal Components Analyses (PCA) on SCA: in a factorial cross design, the SCA components of a set of parental lines (e.g. male) are considered as genetic environments for the SCA components of the other set (e.g. female lines), and vice versa. The information brought by the first principal components is used to predict, under some hypotheses, the SCA value of non experimented combinations.
We propose here to illustrate how the LABKEY statistical and data analysis tools can give keys to the breeder for his future crosses. For that, we use a workflow made of the following steps:

- Extraction from the database managed by LABKEY of a pertinent subset of experimental data on hybrids with the purpose of General Combining Abilities (GCA) and SCA estimations
- Performing PCA on SCA
- Use of the first principal components found by the PCA to reconstruct estimated SCA for the non experimented hybrids.
- Valuation of the previous results in a crosses matrix highlighting promising and / or unpromising crosses and parents according to different criteria. At this step, the breeder can select the future crosses he wants to make: it will automatically generate the new materials in LABKEY.

Genetic Variation of Oleic Acid Content in IOPRI's Oil Palm Breeding Populations

Edy Suprianto[1], Heri Adriwan Siregar[1], Hernawan Yuli Rahmadi[1], Nanang Supena[1], Yurna Yenni[1], Abdul Razak Purba[1]*

[1]Indonesian Oil Palm Research Institute (IOPRI) Jl. Brigjen Katamso No. 51 Medan 20158
[*]edygrey@gmail.com

The breeding objectives of oil palm (*Elaeis guineensis Jacq*.) have been focused on the increasing of oil yield as well as the improvement of oil quality. Palm mesocarp oil, known as crude palm oil (CPO), typically contains about 35%-40% oleic acid, 40-50% palmitic acid and about 10% linoleic acid. A palm oil with a higher proportion of oleic acid is preferred and it could mean new market opportunities for edible palm oil. Study of genetic variation of oleic acid content has been carried out in Indonesian Oil Palm Research Institute (IOPRI) to identify genetic materials having high oleic acid contents among breeding populations. Research has been conducted on 323 dura x pisifera/tenera crosses from second cycle of reciprocal recurrent selection (RRS) scheme tested on four locations in North Sumatra and Riau, Indonesia. Identification of oleic acid content has also been conducted on existing varieties as well as oil palm germplasm collections including *E. oleifera* derived materials. Relatively large variation on oleic acid content have been found among the tested crosses, with the ranged between 12% to 57.84%. Oil palm varieties derived from La Mé and SP 540 materials showed high oleic acid content with average 44.3% and 42.3%, respectively. Oil with high oleic acid content has been identified on wild *E. oleifera* accessions from Suriname, about 64-67% of oleic acids. With relatively large genetic variation, it may be possible to create medium-high (55%-65%) oleic acid oil palm cultivars with conventional methods. Some QTL linked to oleic acid content found in other study may help to accelerate IOPRI's breeding program for development of high oleic acid oil palm cultivars.

Keywords: oil palm, germplasm, oleic acid, breeding, recurrent selection

Introduce two SAS code for finding shifted parameter and cut of point in cluster analysis in Shifted Multiplicative Model

Sajad Talaee [1*]
[1] *M.Sc.in Plant Breeding*
[*] *sajadtalaee@yahoo.com*

Genotype × environment interactions are known in all living existents from unicellular to higher organisms and are of high importance in various aspects of agriculture, genetics, evolution and statistical investigations. Although many statistical methods have been developed to study the Genotype × Environment interactions, but most methods are not able to detect significant differences for the separability of genotypic effects from environmental effects, separability of environmental from genotypic effects. Shifted multiplicative model (SHMM) is a strong tool for the separation genotypic effects from environmental effects and also separation environmental effect from genotypic effects. Lack of functional software program is a major problem for to use this model. In this paper: 1- introduce a SAS code that easily calculates the amount of F_{GH1}, F_{GH2} and F_1. This three Statistics can apply for cut of point in cluster analysis in SHMM model. 2- Also we prepare a sas code for calculate the shifted parameter (β) in SHMM model. This code can easily to use and don't need newton raphson algorithm. This two SAS code program can be applied by any breeder or student who knows a few basic principles of SAS software.

Validation and optimization of genomic prediction models

Patrick Thorwarth[1], Karl Schmid[1]*
[1]University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany
[*]Patrick.Thorwarth@uni-hohenheim.de

Background:
In genomic prediction different statistical models are evaluated using a cross-validation approach to find the model which is maximizing the prediction ability. In the meantime there is still uncertainty about the influence of different parameters such as the genetic architecture, the population size and the optimal number of markers on the prediction ability. We simulated 32 barley data sets differing in the composition of individuals (100 to 750), heritabilities (0.25 and 0.75) and linkage disequilibria (0.05 and 0.2) with a fixed number of markers (1000 per chromosome) to assess the influence of those parameters on the prediction ability. Additionally, we implemented a variable selection algorithm to optimize an Genomic Best Linear Unbiased Prediction model (GBLUP) to increase the prediction ability and compared it to state of the art models, such as BayesB, LASSO and Elastic Net.

Results:
First results indicate an increase in the prediction ability based on our variable selection algorithm by 5-10% percent in comparison to state of the art methods.
In populations with a small amount of LD our algorithm keep s most of the QTLs, while under the high LD scenarios always some QTLs were removed.
Conclusion:
Our results so far indicate, that our variable selection algorithm can increase the prediction ability in all scenarios, but the performance in situations with a low LD and high heritability in combination with a large population size is most suitable.

Efficiency analysis of bioregulator application through seed on the initial growth of cotton plants (*Gossypium hirsutum* L.)

Vendruscolo, E. P.[1]; Lima, S. F.[2]; De Bem, I. O.[1]; Souza, E. J.[3]; Santos, O. F.[2]; Seleguini, A.[1].*

[1] Escola de Agronomia, Universidade Federal de Goiás, Goiânia-GO, Brazil
[2] Universidade Federal de Mato Grosso do Sul, Chapadão do Sul-MS, Brazil
[3] Universidade Estadual Paulista, Ilha Solteira-SP, Brazil
[*] agrovendruscolo@gmail.com

Background: New technologies as bioregulators are constantly applied to increment cotton yield, one of the most important fibrous cultures worldwide. However, the recommendation and use of these compounds are made without specific studies about their efficiency. The current study aimed to evaluate the effect of different bioregulator doses appliedas seed treatment,on the initial development of two cultivars of cotton. The study was conducted in a greenhouse located at the Federal University of Mato Grosso do Sul. Seeds of two cotton cultivars (FMT 701 and FMT 705) were exposed to six different doses of Stimulate® (0.0; 5.0; 10.0; 15.0; 20.0 and 25.0 ml 0.5 kg^{-1} seed), comprising the 12 treatments analyzed, conducted in a completely randomized design.
Results:Biometric characteristics of plant growth were evaluated weekly by measuring the height and the diameter, until 49 days after plant emergence.We used a simplest model to analyze the data (2x2 factorial) since the development ofa more complexmodelanalysis resulted inno significant effect (deviance analysis), in other words,the insertion of a split in time design does not cause differencesin the model explanation of the data. For the two variables the main effects were significant (cultivars and doses).However, as the interaction between cultivar and dose was significant, analysis of isolated effects can imply misinterpretation. The decomposition of the interaction effect pointed that a higher diameter average was obtained to cultivate FMT 701 using 10 ml 0.5 kg^{-1} seed of product, however, to cultivar FMT 705 the treatments 0, 10 or 15 ml 0.5 kg^{-1} seed of the product generate the highest averages, with no significant difference among them.
Conclusions:Considering the height variable, better results for cultivar FMT 701 were obtained by the application of 15 ml kg^{-1} seed, to cultivar FMT 705 the application of 0 or 10 ml kg^{-1} seed produced the best averages. The results showed that specific studies are required for recommending the application and the dosage of this bioregulator for cottoncultivars. Showing that the inclusion of inputs in the production must be carefully analyzed, allowing a better use of investments.

A case study on constructing a rearranged physical and genetic SNP marker map for tetraploid alfalfa

Roeland E. Voorrips[1], Johannes Jansen[2], Julie Ho[3] and Marco C.A.M. Bink[2]*
[1]Wageningen UR – Plant Breeding, Wageningen, the Netherlands
[2]Wageningen UR – Biometrics, Wageningen, the Netherlands
[3]Forage Genetics International, Davis, California, USA
[*] roeland.voorrips@wur.nl

While for tetraploid alfalfa (*Medicago sativa*) no physical sequence is available, a reference genome is available for the related diploid species *M. truncatula.* Based on these sequence data an Infinium array with 9277 SNPs was designed (Li et al 2014). We used this array to genotype two alfalfa Full-Sib progenies, each of about 170 individuals. Separate genetic maps were constructed for the 1:1 segregating markers (simplex x nulliplex, nulliplex x simplex and analogous configurations) and for the 1:4:1 segregating markers (duplex x nulliplex, nulliplex x duplex and analogous). Due to the narrow basis of the populations and the tetraploid nature of the crop many separate genetic linkage groups were obtained, often covering only a small fraction of a chromosome. We applied a regression approach to align these linkage groups with the physical marker map, with consistent results between the two populations. This alignment resulted in the identification of two rearrangements between the alfalfa and *M. truncatula* chromosomes and in the identification of several markers with inconsistent physical positions. A new, rearranged physical map was constructed, taking into account the chromosomal rearrangements and with consistent positions assigned to the aberrant markers. For this new map the recombination rate was estimated for each 0.5 MB segment, such that also an approximate genetic map was derived which we now use for QTL mapping in these populations.

Reference:

Li X, Han Y, Wei Y, Acharya A, Farmer AD, Ho J, Monteros MJ, Brummer EC (2014) PlosOne 9: e84329.

Haplotype allele assignment and imputation in pedigrees

Roeland E. Voorrips[1], Marco C.A.M. Bink[2], Johannes W. Kruisselbrink[2], Herma J.J. Koehorst - van Putten[1] and W. Eric van de Weg[1]*
[1] Wageningen UR – Plant Breeding, P.O. Box 386, 6700 AJ Wageningen, the Netherlands
[2] Wageningen UR – Biometris, P.O. Box 16, 6700 AA Wageningen, the Netherlands
[*] roeland.voorrips@wur.nl

With current genotyping platforms it is affordable to generate high-density SNP data for large pedigrees. However, the use of SNP genotypes for QTL mapping in pedigrees is limited, especially because multiple QTL alleles may be segregating whereas individual SNP markers have only two alleles. Using pedigree information, the FlexQTL software is able to phase the SNP genotypes to a large extent. It also provides information that allows the definition of "haploblocks": groups of tightly linked SNPs without any recombination events within the pedigree. Here we present complementary software (FQ-haplotyper) that uses the phased SNP genotypes and defined haploblocks to (1) impute missing SNP data, (2) delete and where possible correct erroneous SNP scores and (3) define and assign haploblock alleles. Full information on the imputations is provided, which may be used as a tool for curation of the data. The multi-allelic nature of the haploblock genotypes as assigned by FQ-haplotyper makes them much more informative than individual SNPs. Therefore they represent the information present in the SNP genotypes in a compact form which allows efficient QTL mapping. Our FQ-haplotyper software exports data in a form suitable for QTL mapping by FlexQTL and also produces data files for Pedimap software to visualize the flow of alleles through the pedigree. SNP data from an Illumina 20K array on a large apple pedigree will be used for demonstration.

Subgenomic diversity patterns caused by directional selection in bread wheat genepools

Kai Voss-Fels[1], Matthias Frisch[2], Lunwen Qian[1], Stefan Kontowski[3], Wolfgang Friedt[1], Sven Gottwald[1], Rod J. Snowdon[1]*

[1] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
[2] Institute for Agronomy and Plant Breeding II, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
[3] W. von Borries-Eckendorf GmbH & Co. KG, Hovedisser Str. 92, 33818 Leopoldshöhe, Germany
** Kai.P.Voss-Fels@agrar.uni-giessen.de*

Genetic diversity represents the fundamental key to breeding success, providing the basis for breeders to select varieties with constantly improving yield performance. On the other hand, strong selection during domestication and breeding have eliminated considerable genetic diversity in the breeding pools of major crops, causing erosion of genetic potential for adaptation to emerging challenges like climate change. High-throughput genomics technologies can address this dilemma by providing detailed knowledge to characterise and replenish genetic diversity in breeding programs. In hexaploid bread wheat (*Triticum aestivum* L.), the staple food for 35 percent of the world's population, bottlenecks during allopolyploidisation followed by strong artificial selection have considerably narrowed diversity, to the extent that in many regions yields appear to be unexpectedly stagnating. In this study we used a 90k single-nucleotide-polymorphism (SNP) wheat genotyping array to assay high-frequency, polymorphic SNP markers in 460 accessions representing different phenological diversity groups from Asian, Australian, European, and North American bread wheat breeding materials. Detailed analysis of subgroup diversity at the chromosome and subgenome scale revealed highly distinct patterns of conserved linkage disequilibrium between different gene pools. The data enable identification of genome regions in most need of rejuvenation with novel diversity, and provide a high-resolution molecular basis for genomic-assisted introgression of novel diversity into chromosome segments surrounding directionally selected meta-loci conferring important adaptation and quality traits.

Natural epigenetic quantitative trait loci in *Arabidopsis*

René Wardenaar[1] and Frank Johannes[1]*

[1] University of Groningen, Population Epigenetics and Epigenomics, Groningen Bioinformatics Centre, Faculty of Mathematics and Natural Sciences, Nijenborgh 7, 9747 AG Groningen, The Netherlands.
[*] r.wardenaar@rug.nl

Methylation of cytosines is an epigenetic mark involved in the silencing of transposable elements (TEs) and genes. Despite its functional conservation across many species, intra-specific surveys have revealed widespread variation in DNA methylation patterns within populations. A key challenge in population genetics is to show that epigenetic variants exist independently of cis- or trans-acting DNA sequence changes, are stably transmitted over many sexual generations, and are significantly associated with phenotypic variation. Addressing this challenge using natural populations continues to pose major technical difficulties.

We recently analyzed a population of (near) isogenic *Arabidopsis* lines that segregate experimentally induced DNA methylation changes at hundreds of regions across the genome[1,2]. We demonstrated that several of these differentially methylated regions (DMRs) act as bona fide epigenetic quantitative trait loci (QTL^{epi}) in this experimental population, accounting for 60 to 90% of the heritability for two complex traits, flowering time and primary root length[3].

Here we provide additional evidence that many of the candidate DMRs in the QTL regions are also variable among natural accessions of this species and do not significantly associate with cis and trans-acting DNA sequence variants. Phenotypic analysis revealed that several of these putative sequence-independent DMRs are significant predictors of flowering time among accessions.

These DMRs may therefore contribute to heritable phenotypic variation in natural populations of *Arabidopsis* and provide an epigenetic basis for Darwinian evolution.

References

[1] Johannes et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* **5**:e1000530.

[2] Colomé-Tatché et al. (2012) Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc Natl Acad Sci USA* **109**:16240–16245.

[3] Cortijo et al. (2014) Mapping the epigenetic basis of complex traits. *Science* **343**:1145-1148.

Genome-scale diversity in oilseed rape breeding pools

Christian Werner[*] and Rod Snowdon

Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

**christian.werner@agrar.uni-giessen.de*

High-density single-nucleotide polymorphism (SNP) genotyping provides a powerful platform to investigate diversity between and within gene pools on a genome-wide scale. In rapeseed, the analysis of molecular diversity patterns can provide further insight into intra- and interspecific diversification as well as genetic narrowing (bottleneck effect), both phenomena which predominantly arose due to intense artificial selection since the spontaneous hybridization of *Brassica rapa* and *Brassica oleracea*.
In this study we are investigating genome-wide diversity patterns between genetically distinct winter, semi-winter and spring oilseed rape breeding populations by genotyping a collection of 1055 homozygous lines with the Brassica 60k SNP Illumina consortium genotyping array. Analysis of linkage disequilibrium (LD) and genome diversity at the population and chromosome level lay the foundation for the identification of highly conserved chromosomal regions and the introgression of novel variation by the interchange of genetic material between *Brassica napus* pools, thereby promoting the generation of new breeding diversity. Subsequent genome-wide association studies (GWAS) will reveal the impact of these genomic sections on agronomic traits, facilitate targeted recombination and enable identification of the genes underlying these traits.
Genome-wide evaluation of molecular diversity indicates strongly eroded diversity across several chromosome regions. Extensive patterns of LD can be observed within all *B. napus* subgroups examined in this study, especially on the C-genome. These regions represent key targets for improvement of agronomic traits within the respective gene pools by introgression of novel allelic diversity, simultaneously promoting the establishment of heterotic pools for rapeseed hybrid breeding.

Comparison of processing methods for circular chromosome conformation capture (4C-seq) data

Dimitrios Zisis[1], Iris Hövel[2], Rurika Oka[2], Maike Stam[2], Paweł Krajewski[1]*

[1] Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34 Poznań, Poland
[2] Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
** dzis@igr.poznan.pl*

Circular chromosome conformation capture (4C) is a high resolution methodology, which combined with high throughput sequencing can study DNA contacts made across the genome by a given genomic site of interest (referred to as a "viewpoint" or "bait"). Recently some methods for the analysis of 4C-seq data have been developed such as 4Cseqpipe [1], fourSig [2] and a method based on DESeq algorithm for gene expression analysis [3]. They provided basic algorithms for the preprocessing of next-generation sequencing reads, the creation of in-silico library of restriction fragments, read alignment, and contact strength estimation. Application of those methods for *Arabidopsis thaliana* data revealed some limitations. Here we aim at providing a new method for the analysis of 4C-seq data starting from pre-processing of the NGS reads, fragments preparation, mapping, and ending with normalization and statistical comparison of samples, in the same package. We apply the algorithm eXpress [4] for correction of estimated coverage with respect to non-uniquely mapped reads. We compare our algorithm with the existing ones in terms of the efficiency and interpretation of each step. For the comparison we use the data obtained in the experiment with *Arabidopsis thaliana* and FLC locus as the viewpoint.

References

[1] Van De Werken H.J., Landan G., Holwerda SJ. et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods 9(10), 969–972.
[2] Williams R.L. Jr., Starmer J., Mugford J.W. et al. (2014). fourSig: a method for determining chromosomal interactions in 4C-Seq data. Nucleic Acids Res. 42(8), e68.
[3] Ghavi-Helm Y., Klein F.A., Pakozdi T. et al. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512(7512), 96–100.
[4] Roberts A., Pachter L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. Nature Methods 10, 71.

This work was supported by the EpiTRAITS project 316965 which is funded by the European Commission within its 7th Framework Programme. Further information about the project can be found at http://www.epitraits.eu/.

RNA-Seq based custom SNP array for Marker Assisted Selection in Potato Breeding

janss K[1], de Vetten N[2], de Haas J[3], Maliepaard CA[4], Voorrips RE[4], Noback M[5], Wedema R[5], Nap JP[5], Jansen RC[1]*

[1] Groningen Bioinformatics Centre, University of Groningen, The Netherlands
[2] Averis Seeds BV, Valthermond, The Netherlands
[3] HZPC B.V. R&D, Metslawier, The Netherlands
[4] Wageningen UR Plant Breeding, Wageningen University and Research Centre, Wageningen, The Netherlands
[5] Institute for Life Science & Technology, Hanze University of Applied Sciences, Groningen, The Netherlands
∗ k.zych@rug.nl

Background

The cultivated potato is one of the most important food crops. Unfortunately, potato breeding is slow and troublesome. This could be solved by implementing Marker Assisted Selection (MAS). MAS postulates using genetic markers in place of classical phenotypes for selection purposes. This allows investigating traits of interest on seedling stage, saving time, workload and surface needed to breed new cultivars.

MAS is not yet fully realized in potato breeding. The main obstacle is the lack of genomics resources caused by complex genetics of tetraploid potato. In our study we address those shortcomings.

Results

We generated a population of tetraploid potato, consisting of a large bi-parental cross (1000 clones) and a panel of cultivars and breeding clones (200). Together with a diploid cross (200) it forms a unique resource for genomics studies.

In order to discover SNPs with the highest chance to segregate in our population, we performed RNA-Seq on samples from parents of the cross. The best 60 000 of those SNPs were used for custom SNP array. This is one of very few SNP arrays optimized for tetraploid potato.

We genotyped our population using this custom array. Genotype calling was performed with the fitTetra software [1]. Since genotype calling in tetraploids is an extremely challenging task, we updated the software to assure the best performance.

We verified quality of genotypes with SNP calls from parental RNA-Seq data and genotypes produced with commercially available potato SNP array.

Conclusions

We constructed a custom SNP array and used it to genotype largest-to-date experimental population of tetraploid potato. Together with the follow up association studies it will bring MAS one step closer to reality in potato breeding.

Our results clearly show that tailoring an array for population under study results in much higher numbers and quality of genotyped SNPs. Description of the methods and tools used in this study as well as new version of the fitTetra software will soon become openly available. This will facilitate future genomics studies in potato and other polyploids.

[1] R. E. Voorrips, G. Gort, B. Vosman, “Genotype calling in tetraploid species from bi-allelic marker data using mixture models.”, BMC Bioinformatics 12:172, 2011

Premium Sponsors

VSNi Bioscience Software & Consultancy

LemnaTec We are Phenotyping

ENZA ZADEN

KWS Seeding the future since 1856

MONSANTO

Regular Sponsors

KeyGene

Limagrain from earth to life

syngenta

RIJK ZWAAN

SESVANDERHAVE sugar beet seed

DUPONT PIONEER

BARENBRUG

Averis Seeds

BAYER Bayer CropScience