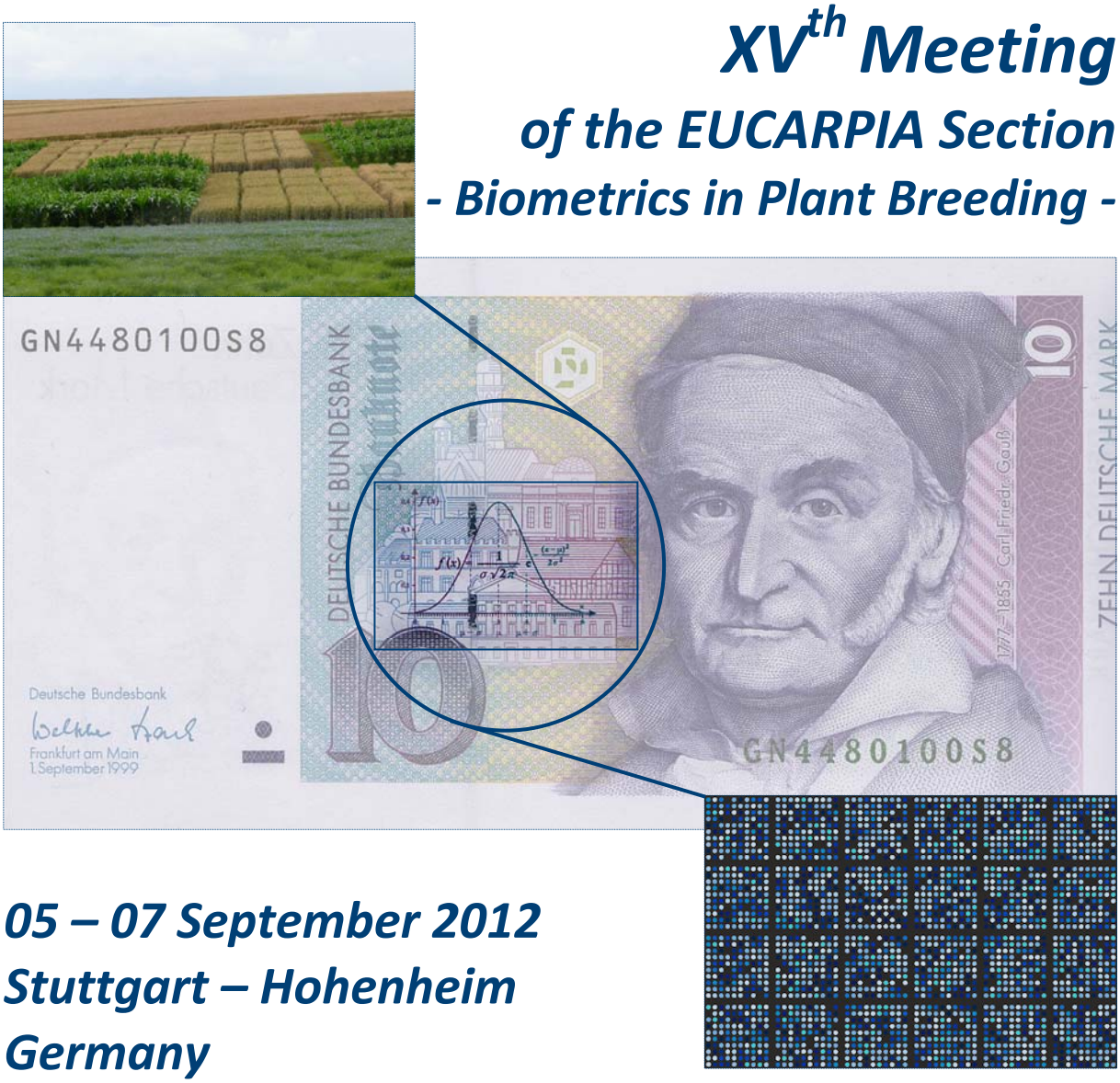
EUCARPIA

Programme, Information, Abstracts



XVth Meeting

of the EUCARPIA Section

- Biometrics in Plant Breeding -

05 – 07 September 2012

Stuttgart – Hohenheim

Germany

If you are wondering what is the message behind the title page, here is a brief explanation.
In the center is the old ten ‘Deutsche Mark’ bill with the drawing of Carl Friedrich Gauß, Mathematician and Astronomer. On the same bill, highlighted by the bordering rectangle, you see the ‘Gaussian’ normal distribution, representing “Biometrics”.

The other two pictures represent two key technologies in ‘Plant Breeding’, i.e. field trials and genetic markers.

Table of contents

Conference venue

The XVth Meeting of the EUCARPIA Section Biometrics in Plant Breeding will be held at

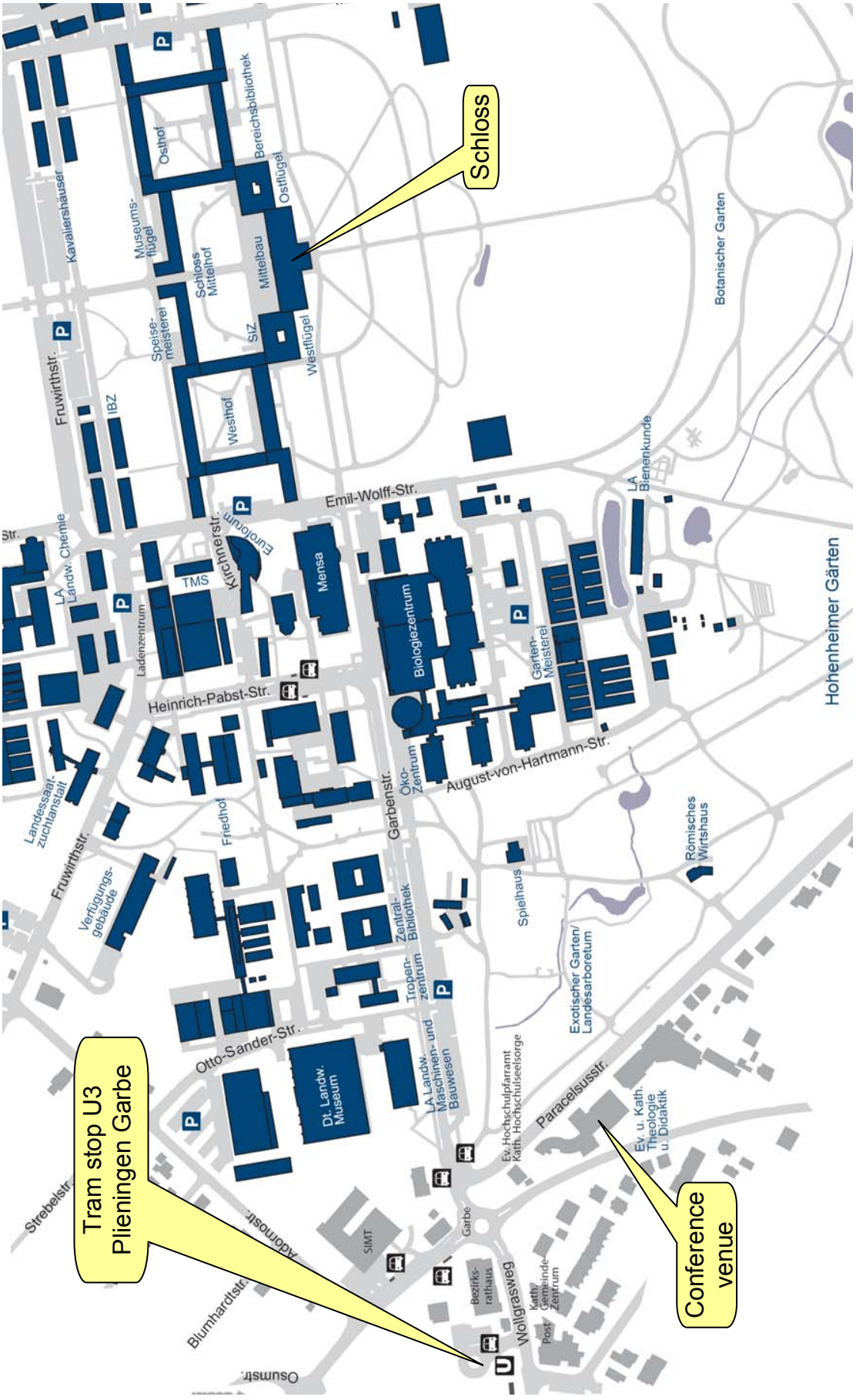
Akademie der Diözese Rottenburg-Stuttgart Tagungszentrum Hohenheim
Paracelsusstr. 91
70599 Stuttgart
Germany
Telefon: +49 (0) 711 451034-600
Telefax: +49 (0) 711 451034-898
Email: Hohenheim@akademie-rs.de

It is a comfortable and friendly conference centre conveniently located next to the beautiful Botanical Gardens and close to the main campus of the Universität Hohenheim.

University

The University is housed in and around the historical premises of Schloss Hohenheim. It is surrounded by a modern campus which, apart from the University institutions, includes the Botanic and Exotic Garden and three museums. Students attending the Universität Hohenheim study and read in the rooms in which Duke Carl Eugen once used to work. The departments of Natural Sciences, Agricultural Science and Business and Social Studies offer a broad range of interesting courses of study. With an innovative and international focus, the Universität Hohenheim works to continually improve the quality of human life in the areas of health and nutrition, sustainable agriculture, business, innovation and service management as well as communication.

Map of Universität Hohenheim with important locations



Hotels

Hotel Neotel

In the Hotel Neotel you will experience all the amenities of a modern 4-star-business-comfort hotel. The obliging and friendly staff will make you feel at home in Stuttgart. On the website you will find everything you need to know about the NEOTEL. To get to the venue you will need about 20 minutes via public transport. From the hotel you will reach the tram stop (U3, Vaihinger Straße) in about 5 minutes by foot.

Hotel Neotel, Vaihinger Str. 151, D-70567 Stuttgart/Möhringen,
Tel. 0711/7814-0; www.hotel-neotel.com; info@hotel-neotel.de

Hotel Gloria

In the Hotel Gloria you will enjoy a special service in a family-run private hotel. On the website you will find everything you need to know about the Hotel Gloria. To get to the venue you will need about 15 minutes via public transport. From the hotel you will reach the tram stop (U3, Sigmaringer Straße) in about 3 minutes by foot.

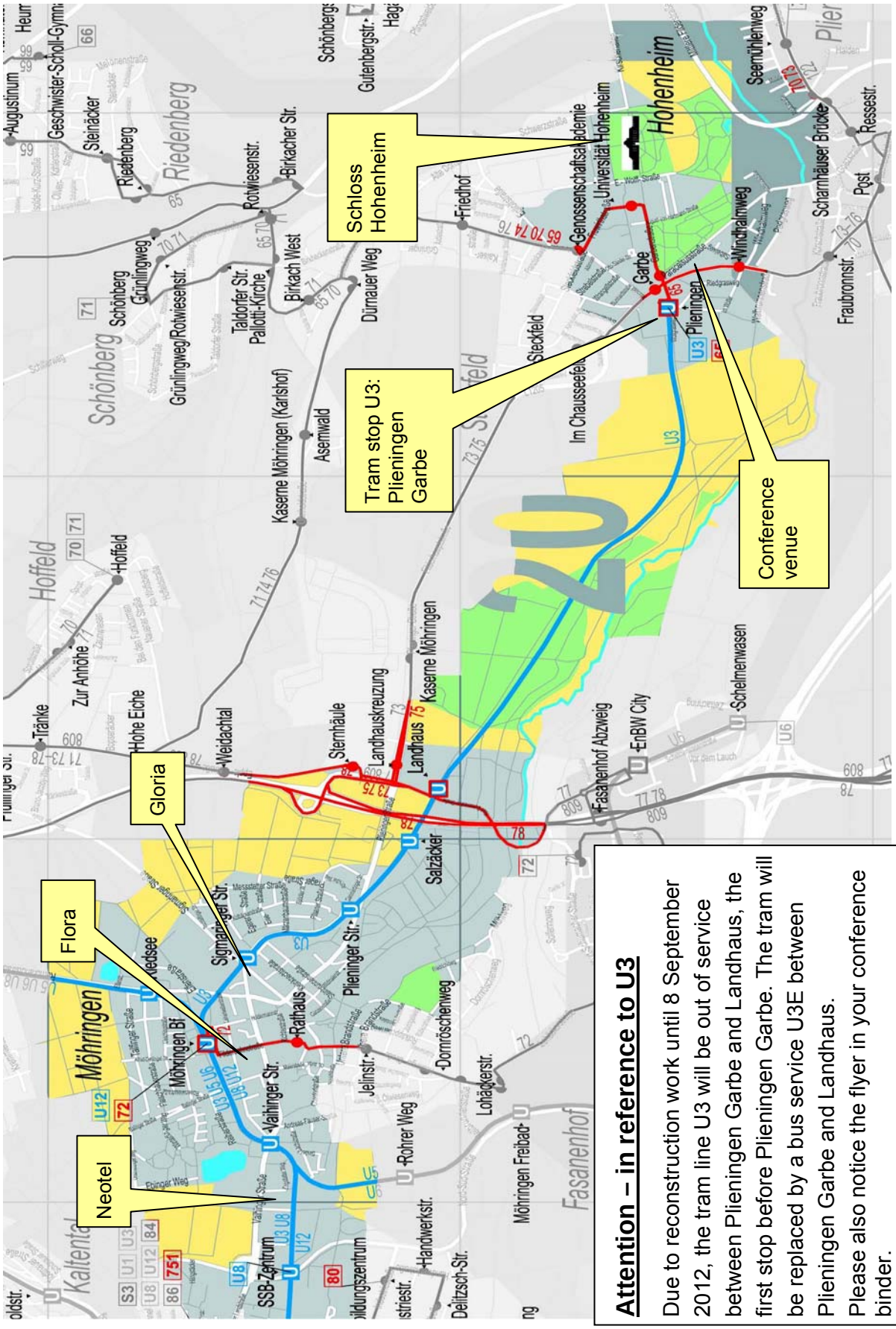
Tagungs- und Musicalhotel Gloria, Sigmaringer Str. 59, D-70567 Stuttgart/Möhringen
Tel. 0711/7185-0 ; www.hotelgloria.de; info@hotelgloria.de

Hotel Flora

The Hotel Flora is ideal for business and leisure with comfortable and spacious guest bedrooms some of which have a balcony or terrace, its breakfast area and a welcoming team. Discreetly located and surrounded by a generous garden, the hotel has a cosy and family atmosphere. On the website you will find everything you need to know about the hotel. To get to the venue you will need about 20 minutes via public transport. From the hotel you will reach the tram stop (U3, Möhringen Bahnhof, platform 4) in about 6 minutes by foot.

Hotel Flora Stuttgart-Möhringen, Filderbahnstr. 43, 70567 Stuttgart/Möhringen
Tel. 0711/71608-0; www.flora-hotel-stuttgart.de, info@florahotel.de

Map showing tram line between Universität Hohenheim, conference venue and hotels



Attention – in reference to U3

Due to reconstruction work until 8 September 2012, the tram line U3 will be out of service between Plieningen Garbe and Landhaus, the first stop before Plieningen Garbe. The tram will be replaced by a bus service U3E between Plieningen Garbe and Landhaus.
Please also notice the flyer in your conference binder.

Tickets public transport (VVS)

To travel between hotels, conference venue and City of Stuttgart,
there are three kinds of ticket:

We recommend the

3-Day-Ticket, 10,60 Euro

This ticket can be used to travel between conference venue, hotels and the City of Stuttgart, BUT NOT to/from the airport.

The ticket can be bought at these places:

→ Tourist-Information Center Flughafen Stuttgart, Terminal 3, Ebene 2
→ Hotel Neotel
→ Hotel Gloria
→ Conference secretariat

There are also one-way and four-trip tickets:

One-way ticket, 1-zone	2,10 Euro	Hotels – Conference Venue
One-way ticket, 2-zone	2,60 Euro	Airport Stuttgart – Hotels / Conference Venue City of Stuttgart – Hotels / Conference Venue
Four-trip ticket, 1-zone	7,90 Euro	
Four-trip ticket, 2-zone	9,90 Euro	

These tickets can be bought at ticket machines on platform. To select the appropriate ticket, you need to punch in a three-digit code. Use 001 and 002 for 1-zone and 2-zone tickets, respectively. There is also a button showing a ‘’4x’’ that needs to be pressed to select four-way ticket.

Taxi

You can order a taxi at:
Taxi-Auto-Zentrale Stuttgart e.G.
Phone: +49 (0) 711 55 10 000
For local calls use: 55 10 000
You may also ask the conference secretariat to help you order a taxi.

Information

Conference secretariat open*:*

Wednesday	08:00 – 18:00
Thursday	09:00 – 14:00
Friday	09:00 – 16:00

Internet access:

Internet access (WLAN) is available at the conference venue. For access details check the material in your conference binder or ask at the conference secretariat.

Local organisation:

Prof. Dr. Hans-Peter Piepho
Universität Hohenheim
Fachgebiet Bioinformatik - 340c -
Fruwirthstraße 23
70599 Stuttgart
Germany
Tel.: +49 (0) 711 459-22386
Fax: +49 (0) 711 459-24345
E-mail: piepho@uni-hohenheim.de

Acknowledgements

We thank the following supporters of our conference:

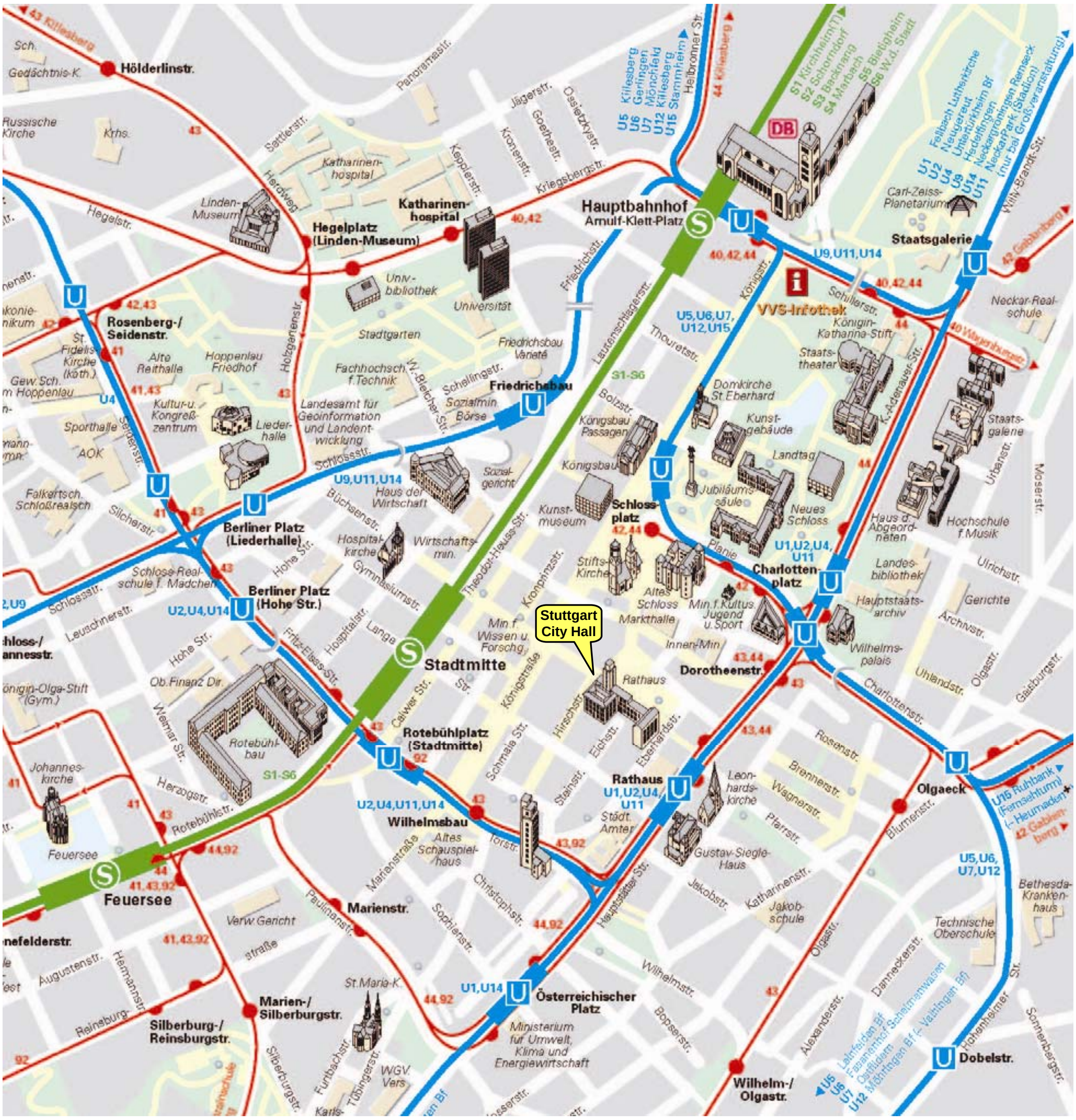
GenStat ASReml	**VSN International**
KWS Seeding the future since 1856	**KWS Saat AG**
strube innovativ aus Tradition	**Strube**
PIONEER A DuPont Business	**Pioneer Hi-Bred International, Inc.**
syngenta	**Syngenta Agro**
RIJK ZWAAN	**Rijk Zwaan**
Limagrain from earth to life	**Limagrain**
Saatzuchtgesellschaft M.B.H. NORDSAAT	**Nordsaat**

**We also thank the German Research Foundation (DFG)
for the financial support of our conference**

Social and cultural programme

Reception in the City Hall Stuttgart,
Wednesday 05 September 2012, 19:30 hrs

We are pleased to be able to invite you to the reception in the Stuttgart City Hall on Wednesday evening. While enjoying a “Viertele Wein und Brezel”, you will learn a little about the City of Stuttgart. After the reception you are free to have a walk through the City Centre and the beautiful “Green U” beginning at the neighbouring “Schlossplatz”, the central square of the city.



Instructions for public transport between Stuttgart City to the hotels can be found on the next page

Public transport from Stuttgart City (Hauptbahnhof, Schlossplatz, Charlottenplatz) to the hotels

→ Hotel Neotel:
U5 direction to Leinfelden Bf via Degerloch & Möhringen (tram stop 'Vaihinger Straße') **or**
U6 direction to Fasanenhof Schelmenwasen via Degerloch & Möhringen (tram stop 'Vaihinger Straße')

→ Hotel Flora:
U5 direction to Leinfelden Bf via Degerloch & Möhringen (tram stop 'Möhringen Bahnhof') **or**
U6 direction to Fasanenhof Schelmenwasen via Degerloch & Möhringen (tram stop 'Möhringen Bahnhof'), **or**
U12 direction to Möhringen Bf / Vaihingen Bf (tram stop 'Möhringen Bahnhof')

→ Hotel Gloria:
U5 direction to Leinfelden Bf via Degerloch & Möhringen (tram stop 'Möhringen Bahnhof'), then change to **U3** direction to Plieningen (tram stop 'Sigmaringer Straße') **or**
U6 direction to Fasanenhof Schelmenwasen via Degerloch & Möhringen (tram stop 'Möhringen Bahnhof'), then change to **U3** direction to Plieningen (tram stop 'Sigmaringer Straße) **or**
U12 direction to Möhringen Bf / Vaihingen Bf (tram stop ' Möhringen Bahnhof'), then change to **U3** direction to Plieningen (tram stop 'Sigmaringer Straße)

→ Conference venue:
U5 direction to Leinfelden Bf via Degerloch & Möhringen (tram stop 'Möhringen Bahnhof'),then change to **U3** direction to Plieningen (tram stop 'Plieningen Garbe') **or**
U6 direction to Fasanenhof Schelmenwasen via Degerloch & Möhringen (tram stop 'Möhringen Bahnhof'), then change to **U3** direction to Plieningen (tram stop 'Plieningen Garbe) **or**
U12 direction to Möhringen Bf / Vaihingen Bf (tram stop 'Bahnhof'), then change to **U3** direction to Plieningen (tram stop 'Plieningen Garbe')

Attention – in reference to U3

Due to reconstruction work until 8 September 2012, the tram line U3 will be out of service between Plieningen Garbe and Landhaus, the first stop before Plieningen Garbe. The tram will be replaced by a bus service U3E between Plieningen Garbe and Landhaus.
Please also notice the flyer in your conference binder.

Excursion to Ludwigsburg,
Thursday 06 September 2012, 14:00 hrs

Palace Museum (Schlossmuseum)

Visit the rooms were Duke Eberhard Ludwig, his son, Crown Prince Friedrich Ludwig and his wife Princess Henrietta Maria lived.

Guided Museum Tours

- Palace Theater Museum (Theatermuseum)
- Appartment of Duke Carl Eugen (Carl-Eugen-Appartment)
- Ceramics Museum (Keramikmuseum)
- Fashion Museum (Modemuseum)

Favorite Park

In the beautiful deer-park of 70-acre deer, mouflon and chital is preserved.
Go there for a walk and watch these animals.

Ludwigsburg City

The magnificent centre of the town is the Residental Palace with the porcelain factory which was founded by Duke Carl Eugen in 1758. Next to it, and just as imposing, are the hunting lodge and summer residence Favorite and the Lakeside Palace Monrepos. The 30 hectares gardens surrounding the residential palace are home to the garden show Blooming Baroque and the Fairy-Tale Garden. In the centre of the town you can have a look at the baroque living quarters and the birth houses of famous poets, go shopping and try out the local restaurants. Here you will find the tourist information and a city guide.

Conference Dinner at Schloss Hohenheim, Thursday 06 September 2012, 20:00 hrs

On Thursday evening we would like to invite you to take part in the Conference Dinner in “Schloss Hohenheim”

Enjoy baroque music and dance followed by an exquisite buffet in the impressive “Balkonsaal’’

and “Blauer Saal”

Committees

Scientific Committee

Fred van Eeuwijk (Chair)	Wageningen University, Wageningen, The Netherlands
Benjamin Stich	KWS SAAT AG, Einbeck, Germany
Laurence Moreau	INRA, Gif-sur Yvette, France
Charlotta Vaz Patto	Universidade Nova de Lisboa, Lisbon, Portugal
Pawel Krajewski	Polish Academy of Sciences, Poznan, Poland
Dietrich Borchardt	KWS SAAT AG, Einbeck, Germany
Hans-Peter Piepho	University of Hohenheim, Stuttgart, Germany

Local Organizing Committee

Hans-Peter Piepho (Chair)
Karin Hartung
Albrecht Melchinger
Karl Schmid
Jochen Reif

Local Team

Karin Hartung (Chair)
Zeynep Bekc-Akyildiz
Gerdi Frankenberger
Katrin Kleinknecht
Jens Möhring
Joseph Ogutu
Andrea Richter
Thomas Ruopp

Programme

(Pages 18 - 21)

Wednesday, 05 September 2012

08:00	Conference office open	

Beginning

09:00 – 09:10	Piepho, Hans-Peter	Local organizer, University of Hohenheim
09:10 – 09:20	van Eeuwijk, Fred A.	Head of section 'Biometrics in Plant Breeding'
09:20 – 09:30	Boller, Beat	President of Eucarpia

Topic 1 Genomic and marker assisted selection

Chair: Hans-Peter Piepho

09:30 – 10:15	Meuwissen, Theo	*(Invited)*	The accuracy of genomic selection	**T1**
10:15 – 10:35	Scutari, Marco		Efficient use of marker profiles in genomic selection	**T2**
10:35 – 10:55	Weber, Vanessa S.		Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments	**T3**
10:55 – 11:25	Coffee break / posters			
11:25 – 12:10	Sillanpää, Mikko	*(Invited)*	Bayesian Lasso-related methods for genomic predictions and QTL analysis using SNP data	**T4**
12:10 – 12:30	Bennewitz, Jörn		BayesD: Models for genomewide evaluations in segregating populations considering dominance	**T5**
12:30 – 13:30	Lunch			

Chair: Chris-Carolin Schön

13:30 – 14:15	Bernardo, Rex	*(Invited)*	Designing training populations for genomewide prediction	**T6**
14:15 – 14:35	Wimmer, Valentin		Inferences about the genetic architecture of complex traits from genome-based prediction models	**T7**
14:35 – 14:55	Malosetti, Marcos		REML implementations in GenStat for genomic prediction	**T8**
14:55 – 15:15	Iwata, Hiroyoshi		Genomic prediction of promising crosses based on genome-wide marker and phenotype data of parental varieties: a case study in Japanese pear *Pyrus pyrifolia*	**T9**
15:15 – 15:35	Charmet, Gilles		Accuracy of genomic prediction using simulated trait architecture on real wheat marker data	**T10**
15:35 – 16:05	Coffee break / posters			

Wednesday, 05 September 2012

Topic 1	Genomic and marker assisted selection		
		Chair: Laurence Moreau	
16:05 – 16:25	Bink, Marco C.A.M.	EpisMAI: A Bayesian approach to detection of epistatic QTL in multi-parent populations	**T11**
16:25 – 16:45	Hackett, Christine A.	Linkage analysis and QTL mapping in autotetraploids using SNP dosage data	**T12**
16:45 – 17:05	Maliepaard, Chris	A strategy for tetraploid genetic analysis using SNP markers	**T13**
17:05 – 17:25	De Silva, H. Nihal	A Microsoft® Excel® implementation of Bin Mapping in Kiwifruit	**T14**
17:25 – 17:45	Rasmussen, Søren K.	Gene discovery by GWAS for grain quality in barley and wheat	**T15**

Evening event

19:30	Reception in the City Hall Stuttgart

Thursday, 06 September 2012

Topic 2 **Phenotypic data: Experimental design and analysis for single and multiple experiments**

Chair: Fred van Eeuwijk

09:00 – 09:45	Williams, Emlyn	*(Invited)*	Modern experimental design – Some developments and applications	**T16**
09:45 – 10:05	Möhring, Jens		Efficiency of augmented p-rep designs in multi-environmental trials	**T17**
10:05 – 10:25	Gunjaca, Jerko		Comparison of blocking and spatial models	**T18**
10:25 – 10:55	Coffee break / posters			
10:55 – 11:15	Cullis, Brian		The design and analysis of multi-phase variety trials using mixtures of composite and individual replicate samples	**T19**
11:15 – 11:35	Gogel, Beverley		Partial compositing for the design and analysis of cereal tolerance trials in Australia: a new approach	**T20**
11:35 – 11:55	Eilers, Paul		Spatial P-splines and mixed models in agricultural trials	**T21**
11:55 – 12:15	Moder, Karl		Testing interaction in the linear model of a block design	**T22**
12:30 – 13:30	Lunch			
approx. 13:45	Departure excursion to Ludwigsburg			
17:45	Start return from Ludwigsburg			
approx. 18:30	Arrival at hotels			

Evening event

20:00	Dinner (Schloss Hohenheim)

Friday, 07 September 2012

08:30 – 09:10	Business Meeting of EUCARPIA Section "Biometrics in Plant Breeding''			
			Chair: Fred van Eeuwijk	
Topic 1	**Genomic and marker assisted selection**			
			Chair: Carlotta Vaz Patto	
09:10 – 09:55	Chapman, Scott	*(Invited)*	Interpreting effects of physiological GxE on marker and genomic selection	**T23**
09:55 – 10:15	Schnabel, Sabine		Modeling latent curves for genotype by environment interaction	**T24**
10:15 – 10:35	Hurtado López, Paula		Multi-environment QTL analysis of developmental traits in potato	**T25**
10:35 – 10:55	Wang, Huange		Inferring causal interrelationships among correlated phenotypes using QTL information	**T26**
10:55 – 11:25	Coffee break / posters			
Topic 3	**Implementation of breeding strategies in public and private sector programs**			
			Chair: Pawel Krajewski	
11:25 – 12:10	Gutteling, Evert	*(Invited)*	Use of quantitative methods in breeding programs	**T27**
12:10 – 12:30	Truntzler, Marion		Genomic selection models accounting for heterotic group specific linkage disequilibrium in maize	**T28**
12:30 – 13:30	Lunch			
Topic 4	**Analysis and use of high throughput data in plant breeding (omics, NGS, and phenotyping platforms)**			
			Chair: Dietrich Borchardt	
13:30 – 14:15	Stegle, Oliver	*(Invited)*	Exploiting NGS technologies for efficient and accurate genotype to phenotype mapping in plant systems	**T29**
14:15 – 14:35	Rathore, Abhishek		ISMU: An easy-to-use pipeline for identification of SNPs based on next generation sequencing (NGS) data	**T30**

Friday, 07 September 2012

Topic 4 **Analysis and use of high throughput data in plant breeding (omics, NGS, and phenotyping platforms)**

Chair: Dietrich Borchardt

14:35 – 15:05	Coffee break / posters			
15:05 – 15:25	Krajewski, Pawel		Data collection and processing in POLAPGEN-BD, a project on biotechnology for breeding cereals with increased resistance to drought	**T31**
15:25 – 16:10	Gianola, Daniel	*(Invited)*	”Predictomics”	**T32**

End of Conference

Part II

Abstracts for talks (T1 – T32)

The accuracy of genomic selection

Meuwissen, Theo[1]

1: Norwegian University of Life Sciences, Aas, Norway

Abstract: Genomic selection is a form of marker assisted selection applied on a genome-wide scale using high-density marker genotyping. Marker effects are estimated without testing for their statistical significance, which implies that also small effects are accounted for and thus all genetic variance can be addressed, i.e. no missing heritability. Genomic selection is currently widely used in dairy cattle, and other agricultural and aquacultural species are following their lead. It is thus important to understand the factors that underlie and determine the accuracy of genomic selection, which will help to design an efficient genomic selection scheme. Here the factors determining its accuracy are identified and quantified: 1) marker density; 2) size of the training population; 3) trait heritability; 4) genome size and structure; 5) historical effective population size; 6) relationship between training population and selection candidates; 7) number of genes and distribution of their effects; 8) method used for the estimation of marker effects. It is concluded that insufficient marker density results in a reduction of the genetic variance that can be explained by genomic selection. Factors 2) and 3) summarize the phenotypic information content. Factors 4) and 5) summarize the number of effects that need to be estimated. The number of genes is more important than the distribution of their effects and is only important when an estimation method is used that attempts to give extra weight to the biggest marker effects such as BayesB-type of methods. When the number of genes exceeds the effective number of segments in the genome, extra weighing of big marker effects is not effective because all markers are important. In this case the use of GBLUP is optimal.

Efficient use of marker profiles in genomic selection

Scutari, Marco[1]; Mackay, Ian[2]; Balding, David[3]

1: Genetics Institute, University College London (UCL), London, UK
2: National Institute of Agricultural Botany (NIAB), Cambridge, UK
3: Genetics Institute, University College London (UCL), London, UK

Abstract: We investigate two approaches to make a more effective use of the information contained in the marker profiles used for Genomic Selection (GS).

Some GS models, such as Ridge Regression and Random Regression-BLUP (Endelman, 2011) use all the available markers, while other models, such as LASSO and BayesB (Meuwissen et al., 2001), perform a feature selection in which a subset of informative markers is selected during model estimation. In the context of GS, feature selection is equivalent to assigning zero effects to non-informative markers; therefore, it is synonymous with variable selection and part of model selection. An alternative method of feature selection is provided by the use Markov blankets (Pearl, 1989), which also provide a theoretically sound framework for feature selection and causal modelling. We will explore the use of Markov blankets as an enhanced data pre-processing step removing makers that are non-informative for the trait under study before fitting a GS model. Both approaches can be beneficial for prediction, at the possible cost of losing some polygenic effects.

GS models based on linear mixed models, such as RR-BLUP, can make use of a kinship matrix in the estimation of genetic effects (Crossa et al., 2011). There are many options for computing measures of genome-sharing for two individuals from their marker profiles, including allele-sharing (Heffner et al., 2009, Astle and Balding, 2009), allele correlation, methods that adjust for linkage disequilibrium (Yu et al., 2006) and haplotype-based methods (Lawson et al., 2012). We will investigate the advantages of different approaches.

We will illustrate the use of both approaches and examine their performance using several GS models and real-world data sets across animal and plant genetics.

References

Astle W, Balding D (2009). Population structure and cryptic Rrelatedness in genetic association studies. Statistical Science 24, 451-471

Crossa J, Pérez P, de los Campos G, Mahuku G, Dreisigacker S, Mogorokosho C (2011). Genomic Selection and prediction in plant breeding. Journal of Crop Improvement 25, 239-261

Endelman JB (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4, 250-255

Heffner EL, Sorrells ME, Jannink JL (2009). Genomic selection for crop improvement. Crop Science 49, 1-12

Lawson DJ, Hellenthal G, Myers S, Falush D (2012). Inference in population structure using dense haplotype data. PloS Genetics 8, e1002453

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829

Pearl J (1989). Probabilistic reasoning in intelligent systems. Morgan Kaufmann, San Francisco

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38, 203-208

Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments

Weber, Vanessa S.[1]; Atlin, Gary N.[2]; Crossa, Jose[2]; Hickey, John M.[2]; Jannink, Jean-Luc[3]; Sorrels, Mark[3]; Technow, Frank[1]; Riedelsheimer, Christian[1]; Melchinger, Albrecht E.[1]

1: Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany
2: International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico
3: Department of Plant Breeding and Genetics, Cornell University, Ithaca, USA

Abstract: Genomic prediction is expected to reshape plant breeding fundamentally in the near future. Here, marker effects estimated in 255 diverse maize (Zea mays L.) hybrids were used to predict grain yield, anthesis date and anthesis-silking interval within the diversity panel and 150 hybrids derived from five F2 populations. Whereas the prediction of testcross performance of F2 lines using marker effects estimated in the diversity panel resulted in poor prediction abilities, up to 25% of the genetic variance could be explained by cross validation within the diversity panel. Hybrids in the diversity panel could be grouped into eight breeding populations differing in mean performance. When validation was conducted separately for each breeding population, prediction ability was low. Our results suggest that prediction benefitted mostly from differences in performance and less from the relationship between the training and validation sets or linkage disequilibrium with causal variants underlying the predicted traits. Potential uses for genomic prediction in maize hybrid development are discussed emphasizing the need of (i) a clear definition of the breeding scenario in which genomic prediction should be applied, (ii) a detailed analysis of the population structure prior to performing cross validation, (iii) larger training sets and (iv) detailed information about the genetic relationship between training and validation sets.

Bayesian LASSO-related methods for genomic predictions and QTL analysis using SNP data

Sillanpää, Mikko J.[1]

1: Departments of Mathematical Sciences and Biology, University of Oulu, Oulu, Finland

Abstract: High-throughput laboratory techniques are producing vast amount of genomic marker data. Linear regression model is often considered to link study phenotypes and these marker measurements to each others for: (1) mapping trait-associated loci using multilocus association models, and (2) predicting genomic breeding values in plants. We consider recent LASSO-related methods for both of these tasks and cover some of our experiences with these methods. For example, Bayesian LASSO seems to work relatively well and provides tool that is competent to GBLUP for genomic breeding value estimation in presence of polygenic genetic architechture (e.g. Kärkkäinen and Sillanpää, 2012). BayesB does not seem to work as well in such a case. We also say something about recent extension of Bayesian LASSO which we call extended Bayesian LASSO (Mutshinda and Sillanpää, 2010). This method is related to so called adaptive LASSO where the tuning parameter has been made locus specific. The nice performance of the method is illustrated with few examples, where one considers epistasis (Li and Sillanpää, 2012).

References

Kärkkäinen HP, Sillanpää MJ (2012). Back to basics for Bayesian model building in genomic selection. Genetics (in press). doi: 10.1534/genetics.112.139014

Li Z, Sillanpää MJ (2012). Estimation of quantitative trait locus effects with epistasis by variational Bayes algorithms. Genetics 190, 231-249

Mutshinda CM, Sillanpää MJ (2010). Extended Bayesian LASSO for multiple quantitative trait loci mapping and unobserved phenotype prediction. Genetics 186, 1067-1075

BayesD: Models for genomewide evaluations in segregating populations considering dominance

Bennewitz, Jörn[1]; Wellmann, Robin[1]

1: Institute of Animal Husbandry and Breeding, University of Hohenheim, Stuttgart, Germany

Abstract : Genomic selection refers to the use of dense, genome-wide markers for the prediction of breeding values and subsequent selection of breeding individuals. It has become a standard tool in livestock and plant breeding for accelerating genetic gain. The core of genomic selection is the prediction of a large number of marker effects from a limited number of observations. Most influential methods were already published by Meuwissen et al. in 2001. Until now, the main research emphasis has been on additive genetic effects, for some good reasons. However, it is well known that dominance is the rule rather the exception in quantitative traits, see Bennewitz and Meuwissen (2010), and Wellmann and Bennewitz (2011) and references therein. If individual phenotypes are available, the inclusion of dominance effects could not only increase the accuracy of genomic selection, but predicted dominance effects could also be used to find mating pairs with good combining abilities by recovering inbreeding depression and utilizing possible overdominance in segregating populations. The aim of this study was to introduce Bayesian regression models for genomic evaluation of quantitative traits that account for dominance effects. These models are generalizations of existing Bayesian models that includes only additive effects. The proposed models differ in the way the dependency between additive effects, dominance effects and allele frequencies is modelled. Plausible informative priors are chosen which are in agreement with the genetic architectures of quantitative traits suggested in the literature. We call these generalizations the BayesD models, where D stands for dominance. Details of the models can be found in Wellmann and Bennewitz (2012). These models were validated using simulations, were the simulation protocol was designed to model the genetic architecture of the trait realistically, but without epistasis. Depending on the marker panel, the inclusion of dominance effects increased the accuracy of genetic values by about 17% and the accuracy of genomic breeding values by 2% in the offspring. Furthermore, it slowed down the decrease of the accuracies in subsequent generations. Once real genomic data from traits with precise information on the genetic architecture including dominance effects become available, it should be used to validate the proposed models. Software for the models can be obtained from the authors.

References

Bennewitz J, Meuwissen THE (2010). The distribution of QTL additive and dominance effects in porcine F2 crosses. J Anim Breed Genet 127, 171-179

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829

Wellmann R, Bennewitz J (2011). The contribution of dominance to the understanding of quantitative genetic variation. Genet Research 93, 139-154

Wellmann R, Bennewitz J (2012). Bayesian models with dominance effects for genomic evaluation of quantitative traits. Genet Research 94, 21-37

Lessons from genomewide prediction and selection in maize

Bernardo, Rex[1]

1: Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, USA

Abstract: Genomewide selection (or genomic selection) allows marker-based selection without QTL mapping. In genomewide selection, equations that predict genotypic value are first developed from phenotypic and marker data in a training population. The prediction equations are then used to assess genotypic values in a test population that has been genotyped but not phenotyped. I will review key lessons from applying genomewide selection in maize and other agronomic crops. First, genomewide selection is most advantageous when heritability is high in the training population but is low or zero in the test population. I will present several examples of how such situation can be achieved. Second, predictions in the context of plant breeding programs are usually most accurate with a simple model that assumes that each marker accounts for an equal proportion of the total genetic variance and that ignores epistasis. Third, traits differ in their prediction accuracies even when the marker density, population size, and heritability are kept constant. Empirical data are therefore needed to determine which traits are the most predictable and which traits are not. Fourth, for finding marker-trait associations, models that incorporate genomewide background effects are superior to composite interval mapping and to the QK model used for association mapping. Fifth, modeling breeding values in a manner that accounts for both marker effects and allele frequencies in the population has so far not been found useful. I will conclude by presenting future needs in applying genomewide selection in plants.

Inferences about the genetic architecture of complex traits from genome-based prediction models

Wimmer, Valentin[1]; Albrecht, Theresa[1]; Lehermeier, Christina[1]; Auinger, Hans-Jürgen[1]; Wang, Yu[1]; Knaak, Carsten[2]; Ouzunova, Milena[2]; Schön, Chris-Carolin[1]

1: Plant Breeding, Technische Universität München, Freising, Germany
2: KWS SAAT AG, Einbeck, Germany

Abstract: Genome-based prediction of genetic values is expected to accelerate crop improvement. To meet the challenges arising from advances in genotyping and sequencing technologies, statistical methods that can handle high-dimensional data have been developed and a series of parametric and non-parametric models for genomic prediction have been proposed. However, their respective properties are still not fully understood, causing considerable uncertainty about the choice of models for genomic prediction in experimental data sets. In simulation studies, methods applying variable selection such as BayesB have been reported to yield superior predictive abilities compared to ridge regression BLUP (RR-BLUP). We investigated predictive abilities of these two frequently used prediction models in experimental plant populations from maize (*Zea mays* L.), rice (*Oryza sativa* L.) and the model plant *Arabidopsis thaliana* (L.). Populations differed with respect to effective population size and the extent of linkage disequilibrium. Analyses were performed for quantitative traits of different genetic architecture which we characterized using a genome partitioning approach. For most traits under study results revealed a non-uniform distribution of locus effects across the genome and considerable deviations from the infinitesimal model. Despite substantial differences in their genetic architecture, predictive abilities obtained by RR-BLUP and BayesB did not differ significantly for these traits. From theory it is known that successful variable selection is only guaranteed if the true model is sparse relative to the number of observations being available, an assumption which is unlikely to be fulfilled for the given populations and traits. An empirical method is proposed for estimation of the number of non-zero coefficients to judge if variable selection can further advance prediction performance.

REML implementations in GenStat for genomic prediction

Malosetti, Marcos[1]; Boer, Martin P.[1]; van Eeuwijk, Fred A.[1]

1: Biometris, Department of Plant Science, Wageningen, The Netherlands

Abstract: The advent of molecular marker technology has equipped breeders with information that can be used to take breeding decisions. One typical example is that of QTL detection, where markers linked to QTLs can in turn be used to select superior parental lines. The main statistical problem in QTL mapping is that of model selection: how to select a handful of markers from a larger set, and use those as predictors in (usually) a linear model. However, the increasing number of markers has made this approach increasingly difficult.

Genomic prediction appears as an interesting alternative because it avoids the problem of model selection by using all markers in the model. The animal breeding field has been particularly active in this area, e.g. the seminal work by Meuwissen et al. (2001). However, the much larger number of predictors in relation to the number of observations calls for statistical methods that impose some penalty on the model parameters. The Bayesian framework has been particularly useful in this respect with methods as Bayes A and Bayes B as popular examples (Gianola et al., 2009). More recently, semi parametric models such as reproducing kernel Hilbert space regression models have been proposed (De los Campos et al., 2009). In this paper, we discuss implementations for genomic prediction within a mixed model formulation in a spirit similar to Piepho (2009). Essentially, the models discussed here use a function of a marker based relationship matrix to structure the distribution of random genotypic effects. We illustrate the case using as example a published wheat data set (Crossa et al., 2010) and discuss how the models can be fitted and their predictive performance compared by GPREDICTION, a procedure developed in GenStat.

References

Crossa J, De Los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun HJ (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186, 713-724

De Los Campos G, Gianola D, Rosa GJ (2009). Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. Journal of Animal Science 87, 1883-1887

Gianola D, De Los Campos G, Hill WG, Manfredi E, Fernando R (2009). Additive genetic variability and the Bayesian alphabet. Genetics 183, 347-363

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829

Piepho HP (2009). Ridge regression and extensions for genomewide selection in maize. Crop Science 49, 1165-1176

Genomic prediction of promising crosses based on genome-wide marker and phenotype data of parental varieties: a case study in Japanese pear *Pyrus pyrifolia*

Iwata, Hiroyoshi[1]; Hayashi, Takeshi[2]; Terakami, Shingo[3]; Takada, Norio[3]; Toshihiro, Saito[3]; Yamamoto, Toshiya[3]

1: Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan
2: National Agricultural Research Center, National Agriculture and Food Science Organization, Tsukuba, Ibaraki, Japan
3: National Institute of Fruit Tree Science, National Agriculture and Food Science Organization, Tsukuba, Ibaraki, Japan

Abstract: Genetic improvement of fruit trees is strongly hindered by their long lifespan, large plant size, an extended juvenile phase for seedling, and a marketable product that cannot be assessed until a seedling is physiologically mature. Genomic selection (GS) is an attractive technology to surmount the fruit tree breeding problems because it enables selection without field-testing and accelerates the selection process and reduces the progeny sizes and the costs of carrying individuals to maturity in the field. GS prediction models can be used not only for identifying the best trees in a segregating population but also for identifying promising crosses from the various combinations of parental varieties. Because the time and cost constraints are strong in fruit tree breeding programs, it is highly desirable to determine the promising crosses and the required sizes of segregating populations in a reasonable way. In the present study, we proposed a method for predicting the segregations of target traits in a progeny population based on the genome-wide marker and phenotype data of parental varieties. In the method, we combined segregation simulation and Bayesian regression via Markov chain Monte Carlo (MCMC) simulation. Segregations of genome-wide markers in a progeny population were simulated on the basis of phased haplotypes estimated for parental varieties, and the segregations of target traits were then predicted by plugging the MCMC samples of marker effects into regression equations based on the simulated marker segregation data. Through this method, we can calculate the posterior probability for obtaining trees that meet specified selection criteria and can use the probability to determine the promising crosses and the required size of a progeny population.

In the present study, we applied the method to data collected from parental varieties used in Japanese pear breeding programs. The data contained the genotypes of 394 bi-allelic genome-wide markers with known positions on a linkage map and the phenotypes of 12 agronomic traits for the 76 parental varieties. Because the traits were scored on ordinal categorical scales, we introduced unobserved continuous phenotypes as a latent dependent variable in the Bayesian regression as conducted by Iwata et al. (2009). Prediction accuracy (i.e., the correlation between predicted continuous genotypic values and observed ordinal scores) estimated via cross-validation among the parental varieties was the highest (0.73) in harvest time, high (0.6-0.62) in firmness of flesh and fruit skin color, and moderate (0.4-0.6) in acid content and fruit size. During the building process of prediction models, we recorded the MCMC samples of model parameters including marker effects. Plugging the MCMC samples into the prediction model, we could estimate the posterior probability for obtaining trees that meet specific selection criteria. Based on the probability, for example, we could rank all possible crosses on the probability of obtaining trees with “earlier harvest time” and “larger fruit size”. The proposed method will provide significant information for improving the efficiency of Japanese pear breeding by increasing the size of training data and the number of genome-wide markers.

References

Iwata H, Ebana K, Fukuoka S, Jannink JL, Hayashi T (2009). Bayesian multilocus association mapping on ordinal and censored traits and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. Theoret Appl Genet 118, 865–880

Accuracy of genomic prediction using simulated trait architecture on real wheat marker data

Charmet, Gilles[1]

1: INRA-UBP, UMR Genetics, Clermont-Ferrand, France

Abstract: With the expected development of thousands of molecular markers in most crops, the marker-assisted selection theory has recently shifted from the use of a few markers targeted in QTL regions (or derived from candidate genes) to the use of many more markers covering the whole genome. Provided that a sufficient level of linkage disequilibrium exists between the markers used for genotyping and the true genes underlying QTLs for complex traits, these genome-wide markers can be used to predict the true breeding value (Meuwissen et al., 2001). To be useful for breeding purposes, the accuracy of this Genome Estimate of Breeding Value (GEBV) should not be worse than the estimation based on phenotype, which is not always the best predictor of Breeding Value, particularly in the presence of GxE interactions. Moreover, since GEBV allows shorter selection cycles, an overall improvement of genetic progress per time unit is expected. (Heffner et al., 2009). We present a case study using DArT markers on the INRA wheat breeding programme, in an attempt to implement whole genome selection as an alternative to phenotypic selection. We used simulated data to assess the ability of different models to accurately predict the "true" (simulated) breeding value, as suggested by Iwata and Janninck (2011). The material consisted in 341 breeding lines of the INRA programme, genotyped with 2236 DArT markers. One hundred or 500 markers were assigned QTLs, then removed from the training dataset. QTL effects were drawn from either one uniform, one Gaussian or a mixture of two uniform distributions, with a random Gaussian noise added to achieve a trait heritability of 0.3 or 0.6. Epistatic interactions accounting for 30% of the phenotyping variance were also considered in addition to additive effects of similar variance. The following methods were used: Pedigree BLUP, Ridge Regression BLUP, Bayesian Ridge Regression and Bayesian Lasso, Random Kernel Hilbert Space (RKHS) and random forest regression. 80% of the data were used as training set to predict the target 20%, this being re-iterated 500 times for each parameter combination. Methods were compared for their ability to predict either simulated breeding value or phenotype, according to generic architecture (number of QTL, distribution of effects and trait heritability). Since the correlation of GEBV with TBV and phenotypes were high, such simulations can be useful for choosing the most stable prediction method(s) in real cases.

References

Heffner EL, Sorrells ME, Jannink JL (2009). Genomic selection for crop improvement. Crop Science 49, 1-12

Iwata H, Janninck JL (2011). Accuracy of genomic prediction in barley breeding programmes: A simulation study based on real single nucleotide polymorphism data of barley breeding lines. Crop Science 51, 1915-1927

Meuwissen THE, Hayes B, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829

EpisMAI: A Bayesian approach to detection of epistatic QTL in multi-parent populations

Bink, Marco[1]; Totir, Liviu Radu[2]; Boer, Martin P.[1]; van Eeuwijk, Fred A.[1]; Winkler, C.W.[2]; ter Braak, Cajo J.F.[1]

1: Biometris, Wageningen University and Research Centre, Wageningen, The Netherlands
2: Pioneer Hi-Bred International, A Dupont Company, Johnston, Iowa, USA

Abstract: Pairwise epistasis is the phenomenon that the allelic contribution of a QTL depends on the segregation status of another QTL. Consequently, epistatic effects may be left undetected - or classified as an additive effect, in a population where one but not both QTL are segregating. This detection or classification failure may severely affect the use of QTL in breeding. To increase the probability of detecting epistatic effects multi-parent populations may be used.

A Bayesian approach will be presented that allows detection of pairwise epistatic QTL that may (not) have additive main effects. Novel Markov chain Monte Carlo sampling algorithms were devised to explore models that are different in the number of QTL as well as the type of effects per QTL. A Poisson prior on the number of QTL and a truncated Geometric distribution is assumed for the number of effects per QTL. The priors on main and epistatic QTL effects were scale-invariant Normal distributions.

We will use a simulated dataset of two related populations (one parent in common) to illustrate the effect non-segregating QTL. In addition, we will report epistatic QTL from analyzing data from an eight-parental complex cross in Arabidopsis. We will discuss the application of the approach to SNP data from association panels.

Linkage analysis and QTL mapping in autotetraploids using SNP dosage data

Hackett, Christine A.[1]; McLean, Karen[2]; Glenn, Bryan J.[2]

1: Biomathematics and Statistics Scotland, Invergowrie, Dundee, UK
2: Cell and Molecular Sciences, James Hutton Institute, Invergowrie, Dundee, UK

Abstract: SNP genotyping of autotetraploid mapping populations not only provides large numbers of markers for map construction, but also additional information about SNP dosage, i.e. whether the genotype is AAAA, AAAC, AACC, ACCC or CCCC. Using such dosage information, the recombination frequency can be estimated more precisely than with presence/absence data alone. The dosage for parents and their F1 offspring must first be estimated from allele intensity ratios (known as theta scores), which can be done using normal mixture models. Previous methods for linkage analysis in tetraploid species have been extended to calculate recombination frequencies and most likely phases using allele dosage data by means of an EM algorithm. The JoinMap 4 software (Van Ooijen, 2006) can be used to order the SNPs using the pairwise recombination frequencies and lod scores, taking the map after two rounds of ordering as a basic map. The remaining SNPs have been placed in bins neighbouring the mapped SNPs. A hidden Markov model is used to estimate QTL genotype probabilities along each chromosome. A check of the linkage map can then be made by mapping the theta scores as quantitative traits to see whether their position and phase corresponds to that of the corresponding marker. These methods have been applied to map SNPs measured using the Potato SolCap 8300 Infinium Chip on potato genotypes Stirling and 12601ab1 and 190 offspring. This produced SNP maps for each chromosome with 82-168 mapped SNPs, together with up to 200 further SNPs placed into bins neighbouring each mapped SNP.

References

Van Ooijen JW (2006). JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma, Wageningen, Netherlands

A strategy for tetraploid genetic analysis using SNP markers

Maliepaard, Chris[1]; Voorrips, Roeland[1]; Smulders, René[1]

1: Wageningen UR Plant Breeding, Wageningen University and Research Centre, Wageningen, The Netherlands

Abstract: Genetic analysis using molecular markers is well established for diploid crops but until recently the tools for mapping and QTL analysis in tetraploids were not readily available. However, this is rapidly changing: next-generation sequencing enables the identification of large numbers of SNPs. Array hybridisation generates large SNP marker data sets; software for dosage scoring (fitTetra, Voorrips et al., 2011) allows efficient assignment of the tetraploid SNP genotypes of individuals. Simulation software (see poster abstract P27 by Voorrips & Maliepaard) allows the study of the consequences of different meiotic configurations and inheritance modes for genetic analysis.

We analysed simulated data of a tetraploid cross with respect to assigning markers to individual chromosomes and subsequently individual marker alleles to the eight different (homologous) haplogroups per chromosome. We estimated recombination frequencies under different assumptions regarding disomic or tetrasomic inheritance and investigated the possibilities for constructing tetraploid linkage maps, indicating map positions of markers on each of the haplogroups.

Here we will present the results of these analyses and we will outline a strategy for the construction of linkage maps and genetic analysis in polyploid crops using SNP markers, which in the future will enable QTL mapping in polyploid progenies.

The methods and mapping strategy that we developed will be applied to a SNP dataset obtained on a progeny of almost 240 individuals of a tetraploid potato cross with a 20K Infinium SNP array.

References

Voorrips RE, Gort G, Vosman B (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics 12, 172

A Microsoft® Excel® implementation of bin mapping in kiwifruit

De Silva, H. Nihal[1]; Hall, Alistair J.[2]; McNeilage, Mark A.[1]

1: The New Zealand Institute for Plant & Food Research Limited (PFR), Mt Albert Research Centre, Auckland, New Zealand
2: PFR, Palmerston North, New Zealand

Abstract: Linkage mapping typically requires a large mapping population of 100 – 400 individuals, depending on the population type, to position markers onto a map with adequate precision. When the objective is to construct a high density map, the genotyping effort that is needed for linkage mapping can be considerable. This is particularly so when there is a need to incorporate large numbers of functionally meaningful expressed sequence tag (EST) markers on to a framework map on an ongoing basis. Vision et al. (2000) introduced an alternative approach to reduce this genotyping effort and improve the mapping efficiency, namely bin or selective mapping. Traditional linkage mapping relies on random sampling of recombinants, with little prior knowledge of individuals forming the sample. Once a mapping population has been genotyped for a set of framework markers, information is available on individuals' recombination cross-over sites or breakpoints, and individuals in a mapping population are likely to differ in the number and distribution of these breakpoints in the framework linkage map. Given prior knowledge of breakpoints for individual progeny, the process of bin mapping involves the selection of a smaller sub-sample of more informative individuals from the mapping population, which have collectively an optimal distribution of breakpoints for subsequent placing of new markers. New markers are subsequently screened over the smaller bin sample, rather than the entire mapping population. Hence, bin mapping involves a two-step process: firstly, a bin set of individuals is selected, using the genotyping data and the map for a set of framework markers, and secondly, the bin set is genotyped for new markers and markers assigned to one of finite number of bins constructed on the framework map. The new markers are placed at lower precision cf. framework markers. Brown and Vision (1999) developed a software programme called MapPop, which is freely available for implementation of bin mapping.

The motivation for this study came from the need to implement bin mapping, as part of the New Zealand government-funded 'Accelerated Development of New Kiwifruit Cultivars' (ADNKC) research programme. We anticipated that modification of the MapPop algorithm might be necessary to suit this out-crossing species. We have developed an Excel® workbook following the algorithm described by Brown and Vision (1999), which includes macros that enables the user to examine breakpoints between framework markers in a mapping population dataset visually, as well as to select a smaller optimal sample of individuals based on different bin length criteria. We selected a bin sample size of 46 from the kiwifruit mapping population of 550 individual vines, which had been genotyped for 228 framework markers. We will present the draft Excel bin mapping workbook and discuss its application in relation to the kiwifruit project, as well as the potential for further development of the workbook.

References

Brown DG, Vision TJ (1999). A computationally novel way to place new markers onto genetic maps. Cornell University Technical Report CCOP-99-9: 14 pp

Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD (2000). Selective mapping: a strategy for optimizing the construction of high-density linkage maps. Genetics 155, 407-420

Gene discovery by GWAS for grain quality in barley and wheat

Rasmussen, Søren K.[1]; Shu, Xiaoli[1]; Belluci, Andrea[1]; Ingvardsen, Christina[1]; Backes, Gunter[1]; Torp, Anna Maria[1]

1: Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Copenhagen, Denmark

Abstract: The barley collection genotyped in the ExBarDiv project (Tondelli et al., in prep.) serves as the platform for genome-wide association studies (GWAS) of quality traits. Barley is grown worldwide primarily for feed and malt but in a few regions barley is still a staple food. Barley is also becoming important as a health promoting food like high-amylose barley. We are exploring the potential of using GWAS for gene discovery taking advantage of the fact that all the 7000 SNP markers are located in coding sequences and combine this with synteny to the rice and Brachypodium genomes. Recent results obtained for resistant starch (Xiaoli et al., submitted) and phytic acid (Ingvardsen et al., submitted) will be presented. While the barley grain is used for food and feed the straw may be used for bioethanol production. GWAS on barley straw is carried out in order to identify genes that control recalcitrance towards downstream bioconversion into bioethanol. Similar projects relevant also for breeding organic production (www.molbreed.life.ku.dk:8080/biobreed/) are pursued in hexaploid wheat (www.ikorn.life.ku.dk) genotyped with DArT markers. These markers offer a less direct path for gene identification as compared to SNP markers in barley. Phenotyping for quality traits e.g. grains and straw are generally laborious and expensive to carry out. Therefore medium through-put robotic systems are being implemented in order to increase accuracy and number of replica to strengthen the calculations in GWAS. Results from our first experience within GWAS in barley and wheat will be presented for discussion.

References

Ingvardsen CR, Backes G, Rasmussen SK: Association mapping for accumulation of phytic acid and Pi in the barley grain (submitted)

Tondelli A, Xu X, Schnaithmann F, Sharma R, Ingvardsen C, Comadran J, Thomas WTB, Russell J, Waugh R, Schulman AH, Pillen K, Kilian B, Rasmussen SK, Cattivelli L, Flavell AJ: Structural and temporal variation in the genetic diversity of a European collection of barley cultivars and utility for association mapping of quantitative traits (in prep)

Xiaoli S, Backes G, Rasmussen SK: GWAS of Resistant starch (RS) phenotypes in barley variety collection (submitted)

http://molbreed.life.ku.dk:8080/biobreed/

http://www.ikorn.life.ku.dk

Modern experimental design – some developments and applications

Williams, Emlyn[1]

1: Statistical Consulting Unit, The Australian National University, Canberra, Australia

Abstract: Since the “Design of Experiments” book by Fisher and the introduction of balanced incomplete block designs by Yates, there has been an explosion of ideas for the design and analysis of experiments.

Much of the development has been driven by the power and availability of computers, which have allowed the construction of experimental designs for a wide range of parameter values, and for their analysis using mixed-models.

This talk will discuss some of the developments in the construction of efficient experimental designs for use in practice. The software package CycDesigN for constructing experimental designs will be discussed.

Efficiency of augmented p-rep designs in multi-environmental trials

Möhring, Jens[1]; Piepho, Hans-Peter[1]

1: Institute of Crop Science, Department of Bioinformatics, University of Hohenheim, Stuttgart, Germany

Abstract: In plant breeding augmented designs with unreplicated genotypes are frequently used for early-generation testing. Unreplicated trials have the advantage to allow genotype testing in multi-environmental trials with limited amount of seed. Check plots within augmented designs allow the estimation of error variances and a connection of otherwise unconnected trials in METs. Cullis et al. (2006) proposed to replace these check plots by plots of non-check genotypes leading to partially replicated (p-rep) designs. Williams et al. (2011) apply this idea to augmented designs (augmented p-rep designs). While p-rep designs are increasingly used, a comparison of the efficiency of augmented designs and augmented p-rep designs in MET is lacking. We simulated genetic effects and allocated them depending on four designs to yield data of a triticale uniformity trial. We compared an augmented design and an augmented p-rep design with replicated and unreplicated designs of the same total size and the same number of genotypes. We further varied the error model (spatial and non-spatial) and the assumption of fixed or random genotype effects. We extended our simulation to include correlated genotype effects which are common in genomic selection. We found an advantage of augmented p-rep designs and of using random genotypic effects especially in case of correlated genotype effects. Spatial error models had minor effects and often converged to a model with independent errors.

References

Cullis BR, Smith AB, Coombes NE (2006). On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics 11, 381-393

Williams ER, Piepho HP, Whitaker D (2011). Augmented p-rep designs. Biometrical Journal 53, 19-27

Comparison of blocking and spatial models

Gunjača, Jerko[1]; Buhiniček, Ivica[2]; Jukić, Mirko[2]; Ikić, Ivica[2]; Šarčević, Hrvoje[1]

1: University of Zagreb, Faculty of Agriculture, Zagreb, Croatia
2: Bc Institute, Zagreb, Croatia

Abstract: Since the introduction of spatial analysis to field trials (Papadakis, 1937), concepts of blocking and spatial analysis have usually been opposed and mutually exclusive (Edmondson, 2005). Comparative analyses employing both types of models often tended to prove superiority of spatial models either in simulation (Baird and Mead, 1991) or real data analysis (Cullis and Gleeson, 1989). Uniformity trials reveal the impact of plot size and shape on the model efficiency (Wu and Dutilleul, 1999), but even for large scale forestry trials, despite the omnipresence of strong spatial trends, the importance of appropriate blocking scheme is stressed (Dutkowski et al., 2006). Further case studies involving larger array of experiments (Qiao et al., 2000) lead to the conclusion that spatial models should be considered as a useful addition rather than alternative to blocking (Piepho and Williams, 2010).

Data used in this study were collected from series of field trials performed within different project frameworks, targeting different objectives, carried out at various environments, and involving two major crops (maize and wheat). All trials were set up in latinized row-column design, using CycDesign software. First step in modelling strategy was reduction of the fixed part of the model, so three different basic fixed models were created by excluding latinized rows/columns and replicates: full (latinized), RCBD and CRD. Three baseline models were then complemented with various random parts, corresponding to different blocking and spatial structures. Fitted models were compared using the three criteria: average standard error of difference for the treatment comparison, likelihood ratio test for added random effects and AIC. Selection of the optimal model was further aided by surface plots of data, fits and residuals. Model were fitted using ASREML software (Gilmour et al., 2009) and visualized using R package ‘agridat’ (Wright, 2011).

References

Baird D, Mead R (1991). The empirical efficiency and validity of two neighbour models. Biometrics 47, 1473-1487

Cullis BR, Gleeson AC (1989). The efficiency of neighbour analysis for replicated variety trials in Australia. J Agric Sci 113, 233-239

Dutkowski GW, Costa e Silva J, Gilmour AR, Wellendorf H, Aguiar A (2006). Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. Can J For Res 36, 1851-1870

Edmondson RN (2005). Past developments and future opportunities in the design and analysis of crop experiments. J Agric Sci 143, 27-33

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009). ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK

Papadakis JS (1937). Méthode statistique pour des expériences sur champ. Bull Inst Amél Plantes á Salonique 23

Piepho HP, Williams ER (2010). Linear variance models for plant breeding trials. Plant Breed 129, 1-8

Qiao CG, Basford KE,·DeLacy IH,·Cooper M (2000). Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theor Appl Genet 100, 9-16

Wright K (2011). agridat: Agricultural datasets. R package version 1.3. http://CRAN.R-project.org/package=agridat

The design and analysis of multi-phase variety trials using mixtures of composite and individual replicate samples

Cullis, Brian[1,2]; Smith, Alison[1]; Butler, David[3]; Cavanagh, Colin[4]

1: Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia
2: Mathematics Informatics and Statistics, CSIRO, Canberra, Australia
3: Department of Agriculture, Fisheries and Forestry, Toowoomba, Australia
4: CSIRO Plant Industry and Food Futures Flagship, Canberra, Australia

Abstract: Accurate phenotypic information on quality traits is vital for successful variety selection in plant breeding programs and for genetic research including the identification of QTL and genomic selection. Many quality traits, such as milling yield, dough rheology and bread baking characteristics for wheat, are obtained from multi-phase experiments in which varieties are first grown in a field trial then processed further in a laboratory. Smith et al. (2006) showed the importance of using sound experimental design (including randomisation and replication) in all phases of the experiment followed by an appropriate statistical analysis that captures all sources of non-genetic variation and correlation. Replication in a laboratory phase involves splitting of the experimental units from the previous phase and testing as separate (duplicate) samples. Quality testing tends to be labour intensive and/or expensive so there is typically a limit to the total number of samples that can be tested. Fully replicated sampling strategies are prohibitively expensive and are unnecessary from a statistical perspective. Smith et al. (2006) propose an approach for multi-phase quality testing that involves partial replication. For example, for a two-phase experiment, a subset of the varieties is tested using multiple field replicate samples (the remainder being tested using a single replicate only), then a subset of the selected field plots are split to produce duplicate samples for the laboratory process.

The partial replication approach for multi-phase testing has been shown to work well but is wasteful in the sense that field plots of some varieties are completely ignored. Smith et al. (2011) consider a similar issue but in the context of grain quality traits that are derived using a single (field) phase alone. They propose that some varieties be tested using individual replicate samples and others using composite samples (formed for each variety by combining grain from the individual replicates for that variety). Smith et al. (2011) demonstrate that with such data it is still possible to fit mixed models that account for sources of non-genetic variation and spatial correlation from the field phase. In the current paper we consider the application of the Smith et al. (2011) approach for the first phase in a multi-phase quality experiment. Not only does this avoid discarding field plots but addresses the common problem of insufficient grain from individual replicate plots. Quality testing, particularly end-product testing, involves a minimum grain requirement that can sometimes only be met by compositing several replicate plots for a variety. The application of the Smith et al. (2011) approach in this setting is both a practical and efficient solution, provided that a moderate number of varieties may be tested as individual replicate samples. The need for duplication in the laboratory phases places an additional demand on grain since this requires the splitting of samples such that each sub-sample meets minimum requirements. In this paper we show that duplication may be achieved not only by the splitting of samples from individual field plots (where possible) but also by the splitting of composite samples. The concepts will be demonstrated using a milling yield example.

References

Smith A, Lim P, Cullis BR (2006). The design and analysis of multi-phase plant breeding experiments. Journal of Agricultural Science 144, 393-409

Smith AB, Thompson R, Butler DG, Cullis BR (2011). The design and analysis of variety trials using mixtures of composite and individual plot samples. Journal of the Royal Statistical Society Series C 60, 437-455

Partial compositing for the design and analysis of cereal tolerance trials in Australia: a new approach

Gogel, Beverley[1]; Smith, Alison[2]; Cullis, Brian[2,3]

1: University of Adelaide, Adelaide, Australia
2: University of Wollongong, Wollongong, Australia
3: CSIRO, Canberra, Australia

Abstract: The root lesion nematodes *Pratylenchus thornei* and *P. neglectus* are prevalent across the cereal growing regions in Australia. They are microscopic worms that invade the developing roots of seedlings, inhibit normal root growth and result in reduced uptake of nutrients and water. Large populations of these nematodes can result in major damage to crops and significant yield loss.

Designed field trials are conducted annually across Australia to assess the tolerance of cereal varieties to *Pratylenchus*, where tolerance is defined as the ability of the variety to grow and yield well in the presence of the nematode (Vanstone et al., 2008). Each variety is typically tested at both high and low levels of the nematode and the traits recorded include yield and a measure per plot of the nematode population both at seeding and immediately following harvest. Collecting soil samples and then processing them in the laboratory to obtain initial and final population levels is time consuming and costly. Typically measures are obtained for all plots, that is, for all replicates of all varieties at both the high and low population levels. A new sampling strategy has recently been proposed that involves a mixture of individual and composite plot samples (Smith et al., 2011). Specifically, the replicates for a subset of the varieties are processed individually while a composite sample across some or all of the replicates for the remaining varieties is processed. This approach allows an efficient mixed model analysis.

This talk will describe a trial that has been conducted as a part of the ongoing research into *Pratylenchus* being undertaken by the Molecular Diagnostics Group, South Australian Research and Development Institute (SARDI). The mapping of the composited data back to the full trial layout will also be discussed.

References

Smith AB, Thompson R, Butler DB, Cullis BC (2011). The design and analysis of variety trials using mixtures of composite and individual plot samples. Applied Statistics 60, 437-455

Vanstone VA, Holloway GJ, Stirling GR (2008). Managing nematode pests in the southern and western regions of the Australian cereal industry: continuing progress in a challenging environment. Australasian Plant Pathology 37, 220-234

Spatial P-splines and mixed models in agricultural trials

Eilers, Paul[1,2]; van Eeuwijk, Fred A.[2]

1: Biometris, Wageningen University and Research, Wageningen, The Netherlands
2: Erasmus University Medical Center, Rotterdam, The Netherlands

Abstract: Most agricultural trials are laid out on a field and so are influenced by spatial variation of soil properties. A classical way to compensate for it is to use blocking. More advanced is the use of error models with autoregressive correlation structures (along rows and columns of plots). In this contribution we propose to use spatial P-splines. They combine tensor products of B-splines with difference penalties on the rows and columns of the coefficient matrix.

The penalties can be transformed into a mixed model structure, which can be combined with mixed model components for varieties and separate row and column effects. When isotropy is assumed, or a fixed ratio for variability in row and column directions is given, the spatial penalties lead to one variance component. Unfortunately, the anisotropic case does not allow a similar analogy with two variance components, because the two penalties influence each other. Recent work with Dae-Jin Lee (CSIRO) and Maria Durbán (Carlos III University) has resulted in a decomposition with five variance components. It works well in this application and the five components have a clear interpretation. A straightforward EM procedure works well to estimate all variance in the model. For a practical illustration we use a large trial on growth of potatoes.

Testing interaction in the linear model of a block design

Moder, Karl[1]

1: Institute of Applied Statistics and Computing, University of Natural Resources and Applied Life Sciences, Vienna, Austria

Abstract: Block designs are often used designs to evaluate influences of a factor in the presence of some disturbance variables. Although this kind of design is widely used, it suffers from one drawback. As there is only one observation for each combination of a block and factor level it is not possible to test interaction effects because the mean square value for interaction has to serve as the error term. Although there are some attempts to overcome this problem, these methods, however, have not been adopted in practice and have not been broadly disseminated.

Many of these tests are based on nonlinear interaction effects (e.g. Tukey, 1949; Mandel, 1961; Johnson and Graybill, 1972). Others are based on the sample variance for each row in the block design (Millken and Rasmuson, 1977) with some modification by Piepho (1994). A somehow similar method was proposed by Kharrati-Kopaei and Saddooghi-Alvandi (2007). A review on such tests is given by Karabatos (2005) and Alin and Kurt (2006). Rasch et al. (2009) proposed the use of nonlinear regression which is fitted to Tukey's model and is tested by the likelihood ratio test.

Here a new model is introduced to test interaction effects in block designs. It is based on an additional assumption regarding the columns of the block design which is intuitive and common in Latin Squares. The application of this model is very simple and a test on interaction effect is very easy to calculate based on the results of an appropriate analysis of variance.

The method as such is applicable for fixed effect models as well as for mixed and random effect models.

References

Alin A, Kurt S (2006). Testing non-additivity (interaction) in two-way anova tables with no replication. Stat in Medicine 15, 63-85

Johnson DE, Graybill FA (1972). An analysis of a two-way model with interaction and no replication. Journal of the American Statistical Association 67, 862-869

Karabatos G (2005). Additivity Test. In Everitt BS, Howell DC (Eds.) Encyclopedia of Statistics in Behavioral Science: 25-29. Wiley

Kharrati-Kopaei M, Saddooghi-Alwandi SM (2007). A new method for testing interaction in unreplicated two-way analysis of variance. Communications in Statistics-Theory and Methods 36, 2787-2803

Mandel J (1961). Non-additivity in two-way analysis of variance. Journal of the American Statistical Association 56, 878-888

Millken GA, Rasmuson D (1977). A heuristic technique for testing for the presence of interaction in nonreplicated factorial experiments. Australian Journal of Statistics 19, 32-38

Piepho HP (1994). On tests for interaction in a nonreplicated two-way layout. Australian Journal of Statistics 36, 363-369

Rasch D, Rusch T, Simeckova M, Kubinger K, Moder K, Simecek P (2009). Tests of additivity in mixed and fixed effect two-way anova models with single sub-class numbers. Statistical Papers 50, 905-916

Tukey JW (1949). One degree of freedom for non-additivity. Biometrics 5, 232-242

Interpreting effects of physiological GxE on marker and genomic selection

Chapman, Scott[1]

1: CSIRO, St. Lucia, Australia

Abstract: Following its theoretical and practical demonstration in animal models, genomic selection (GS) is strongly advocated for use in crop plants. Two major issues are that (1) breeders frequently wish to use specific markers associated with known genes, in addition to markers used for GS; and (2) that G x E effects may potentially confound GS in complex target environments. In the literature, much of the emphasis has been on the sampling and composition of the training population, and considerations of the accuracy of prediction from this population. For complex traits and/or target environments, the sampling and composition of environments in the training step are also critical. In sorghum, we now have QTL data from NAM-like BC1 populations for multiple growth traits – flowering time, tillering, transpiration efficiency and root angle. These traits have been incorporated into a physiological model to simulate yield outcomes for millions of genotype by environment combinations in Australian drought-prone regions, while tracking the allelic composition of genotypes for all traits. This is allowing us to construct different types of training populations, sample different types of environments during training, and to compare the outcomes of different combinations of phenotyping and GS strategies in simulated breeding programs. Results from that work will be presented.

Modeling latent curves for genotype by environment interaction

Schnabel, Sabine[1,2]; van Eeuwijk, Fred A.[1,2]; Eilers, Paul[1,3]

1: Biometris, Wageningen University and Research Centre, Wageningen, The Netherlands
2: Centre for BioSystems Genomics, Wageningen, The Netherlands
3: Department of Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands

Abstract: To study the adaptive behaviour of plant genotypes, sets of genotypes are often studied in a series of field and/or greenhouse trials under different conditions. A typical result of such experiments is a genotype by environment (GxE) table of means. The order of the environments is often unknown or unclear. Generally, this kind of table cannot be fitted well by an additive model with effects for G and E due to the omnipresence of GxE interaction. Most methods to describe tables of GxE means use centered data: row-wise, column-wise or double-centered.

In the literature one finds several approaches which are essentially all based on the addition of multiplicative components to additive genotype and environments effects. References to such models go back as far as Fisher and MacKenzie in 1923. Estimating the underlying order of the environments can be seen as a seriation problem. We propose to order the environments based on smooth latent curves. In the context of a GxE table the resulting latent gradient can be interpreted as an unknown environmental characteristic.

This is a new approach to model genotype by environment interaction for plant breeding. The result is an order of the environments which was initially unknown. These complete data can be used in further analyses to gain more insight about the relationship of the phenotypic trait and the environmental conditions. We are interested in the genotype-specific curves. Characteristics of the fitted curves over environments might be associated with QTLs. The model can also be extended to accommodate missing data through a weighting scheme. Additionally, in a second step the order of the genotypes per environment and over the series of environments can be evaluated using expectile curves. This can provide guidelines for breeders identifying high performing genotypes to be used for future breeding programmes.

References

Fisher RA, MacKenzie WA (1923). Studies in variation II. The manorial response in different potato varieties. Journal of Agricultural Science 13, 311–320

Multi-environment QTL analysis of developmental traits in potato

Hurtado López, Paula[1,2]; Schnabel, Sabine[2,4]; Maliepaard, Chris[1]; Eilers, Paul[2,3]; Visser, Richard[1,4]; van Eeuwijk, Fred A.[2,4]

1: Wageningen UR Plant Breeding, Wageningen University and Research Center, Wageningen, The Netherlands
2: Biometris–Applied Statistics, Wageningen University, Wageningen, The Netherlands
3: Department of Biostatistics. Erasmus Medical Center, Rotterdam, The Netherlands
4: Centre for BioSystems Genomics, Wageningen, The Netherlands

Abstract: The presence of genotype by environment and QTL by environment interactions play an important role in the expression of complex traits involved in plant development under field conditions. To understand the genetic basis of developmental processes in potato, multiple experiments have been done under different day length conditions.

In this study, field trials including ~200 genotypes from a diploid backcross mapping population were planted under 3 contrasting day length settings, in Ethiopia (short), The Netherlands (long) and Finland (very long). Flowering, haulm senescence and plant height were evaluated in time series during the growing season to have a better understanding of potato development and adaptation under contrasting environments. Instead of describing the developmental trait as a function of time (days after planting), we transformed the time units into photo-thermal units accounting for temperature and photoperiod in each environment to facilitate the comparability across locations.

The phenotypic time series data were analyzed in a smoothed generalized linear model (Hurtado et al., 2011), so fitted curves were obtained per genotype. Curve characteristics such as onset, progression rate, inflection point and end of the developmental processes under study, described development as a continuous process in time. The curve parameters, having a single observation per genotype, were treated as phenotypic traits in a multi-environment QTL analysis enhancing the power of QTL detection. A multi-trait QTL analysis was also performed in each location to understand the genetic basis of trait correlations and co-location of QTLs. QTL by environment interactions were observed at different development stages of the population, as well as common QTLs across locations. Pleiotropic QTLs were also identified in each location facilitating the understanding of the common genetic bases driving development in potato.

References

Hurtado PX, Schnabel SK, Zaban A, Vetelainen M, Virtanen E, Eilers PHC, van Eeuwijk FA, Visser RGF, Maliepaard C (2012). Dynamics of senescence-related QTLs in Potato. Euphytica 183, 289-302

Inferring causal interrelationships among correlated phenotypes using QTL information

Wang, Huange[1]; Paulo, Joao[1,4]; Jansen, Johannes[1,4]; Bovy, Arnaud[2,4]; Wubs, Maaike[1]; Heuvelink, Ep[3]; Alimi, Nurudeen[1]; Bink, Marco[1]; ter Braak, Cajo[1]; van Eeuwijk, Fred A.[1,4]

1: Biometris–Applied Statistics, Wageningen, The Netherlands
2: Plant Breeding, Wageningen, The Netherlands
3: Horticultural Production Chains, Wageningen University, Wageningen, The Netherlands
4: Centre for BioSystems Genomics, Wageningen, The Netherlands

Abstract: In the QTL analysis of multiple correlated traits it is of interest to organize the traits in directed networks and assess possible causal directions. Current methods mostly yield undirected networks due to inherent limitations of network structure learning algorithms. For example, approaches based on partial correlations don't include assessment of causal directions at all, while likelihood-based scoring metrics for network model selection lead to collections of networks that are likelihood equivalent, meaning that certain edges cannot be directed. Neto et al. (2008) proposed a method for estimating causal directions in phenotypic networks in segregating populations by introducing QTL information. We developed a generalization that is also based on the addition of QTL information to phenotypic networks using a different recursive procedure to infer causal direction for undirected edges. We evaluate and compare the performance of our algorithm to the Neto et al. (2008) algorithm via simulations. Results show that our novel algorithm leads to more reliable orientation outcomes. We apply our algorithm to two real-world data sets in tomato and pepper.

References

Neto EC, Ferrara CT, Attie AD, Yandell BS (2008). Inferring causal phenotype networks from segregating populations. Genetics 179, 1089-1100

Use of quantitative methods in breeding programs

Gutteling, Evert[1]

1: Rijk Zwaan Breeding B.V. Fijnaart, The Netherlands

Abstract: In breeding programs one of the important methods of trait mapping remains classical QTL-analysis (mapping in biparental populations). The progress made in molecular biology and sequencing over the last couple of years allows to take these studies a step further. This presentation will illustrate how an integrative approach is being used to study the genetic basis of complex traits in a commercial cucumber variety. Additionally, the possibilities of new breeding techniques are exemplified.

Genomic selection models accounting for heterotic group specific linkage disequilibrium in maize

Truntzler, Marion[1]; Totir, Liviu Radu[1]; Habier, David[1]

1: Pioneer Hi-Bred International – A DuPont Business, Johnston, IA, USA

Abstract: In maize elite breeding programs, hybrids are created out of two or more heterotic groups (HGs) to exploit heterosis. Genomic selection methodology is available to predict hybrid performance of double haploid (DH) parents, which allows distinguishing these DHs within their bi-parental families by utilizing information from linkage disequilibrium (LD) and co-segregation. Our goal is to capture not only additive allele effects, but also dominance effects to exploit specific combining ability for prediction. The standard dominance model fits one effect for homozygous genotypes and one for heterozygous genotypes at each SNP irrespective of population origin of alleles. However, that model does not account for LD differences in HGs. Studying a two locus model of one SNP and one QTL reveals that selecting for hybrid performance in recurrent selection may be inefficient or even lead the population in the wrong direction. To account for differences in both allele frequencies and LD in HGs, we propose to fit one effect for each of the four SNP genotypes of hybrids, meaning that the two heterozygous genotypes are distinguished by population origin of their alleles. The objective of this study was to compare accuracies within and across bi-parental populations obtained from the standard dominance model and the HG specific model by using simulated and real data. Both models were analyzed by BayesB. For real data and even for extreme simulation scenarios in which both HG were genetically independent, the HG specific model shows only small or no benefit in accuracy compared to the standard dominance model. In simulations with only additive QTL effects the HG specific model was inferior due to the higher number of effects estimated. Additional simulations were conducted in which the training data set consisted of 1,000 individuals from 10 independent populations, and the validation data set only contained individuals from one of the populations. Results showed a persistent and relatively high accuracy due to LD ten generations after training. In conclusion, SNP effects capture not only additive-genetic relationships between close relatives in prediction, but are also able to determine from what population the validation individuals descended and to exploit LD from their own population. Thus, extrapolating from a simple two locus model is not sufficient to predict the usefulness of a statistical model for genomic selection, but a more profound understanding is necessary.

Exploiting NGS technologies for efficient and accurate genotype to phenotype mapping in plant systems

Stegle, Oliver[1]

1: Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Abstract: Combining NGS technology and modelling at a systems level to dissect the causes of transcriptome variability. Molecular traits vary between samples and dissecting the precise origin of this variation is an important step towards understanding the implications of cellular variation for higher-level traits.

We have studied the mRNA variability in 19 accessions of Arabidopsis thaliana at an unprecedented level of detail (Gan et al., 2011) and have extended this analysis to larger samples sets of over 200 lines.

As a first step, accurate quantification of molecular phenotypes is needed. To this end, we will discuss computational approaches to quantify whole-gene expression levels and splicing phenotypes from RNA-Seq datasets.

Second, building on quantitative readouts of the molecular state, we discuss novel statistical approaches to dissect the causes of molecular variability. These models (Stegle and Parts, 2012) allow for attributing the overall gene expression variability to genetic factors, environmental effects and their interactions. Genetic factors may either act in cis or trans and differ in effect sizes for single locus effects and larger indels.

Finally, population structure and subtle environmental factors may confound such analysis if not appropriately taken into account within the model.

By means of this tight integration of computational genomics and statistics analyses, we are able to derive a comprehensive picture of the heritable and environmental component of molecular traits, attributing more than 50% of gene expression variation to individual causes.

References

Gan X, Stegle O, et al. (2011). Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature 477, 419-423

Stegle O, Parts L (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nature Protocols 7, 500-507

ISMU: An easy-to-use pipeline for identification of SNPs based on next generation sequencing (NGS) data

Rathore, Abhishek[1]; Varshney, Rajeev K.[1]; Azam, Sarwar[1]; Shah, Trushar[1]; BhanuPrakash, A.[1]

1: International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India

Abstract: SNP identification based on NGS data requires complex bioinformatics analysis including mapping/alignment of short reads, consensus calling and detection of variants. These steps require one to either use some commercial softwares or to identify appropriate open-source tools, understand their command line options and also their host operating system. To overcome these difficulties and provide the biologists an easy-to-use integrated platform for identification of SNPs based on NGS data, the ISMU (Integrated SNP Mining and Utilization) pipeline has been developed. This pipeline encompasses powerful open-source mapping softwares like Maq, NovoAlign and SOAP and offers the options to the users to modify/select appropriate parameters for SNP discovery. The ISMU pipeline requires the users to feed only raw sequences reads from any NGS platform and a reference genome/transcriptome. As a result of analysis, the pipeline generates alignment file of NGS reads for genotypes, list of SNPs between genotypes in text and gff3 format and analysis results are automatically visualized in popular assembly visualization software Tablet. The pipeline has been developed in perl-cgi for 64 bit linux/unix based system and has been tested on RHEL 5 and Fedora 13. The current version of the pipeline is available at http://hpc.icrisat.cgiar.org/NGS/ as well as a standalone tool.

As a next phase to ISMU we are adding analysis capabilities also to the pipeline. Currently we are working on implementation of modules for Marker Assisted Recurrent Selection (MARS), Genome Wide Selection (GWS) and Genotyping by Sequencing (GBS) to the pipeline.

Data collection and processing in POLAPGEN-BD, a project on biotechnology for breeding cereals with increased resistance to drought#

Krajewski, Paweł[1]; Sawikowska, Aneta[1]; Kaczmarek, Zygmunt[1]; Ćwiek, Hanna[1]; Frohmberg, Wojciech[1]; and Consortium POLAPGEN[2]

1: Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland
2: POLAPGEN Consortium for Applied Genetics and Genomics, Poznań, Poland

Abstract: The subject of the POLAPGEN-BD project is drought resistance in cereals, investigated in spring barley treated as both a model and economically important plant. Increasing desiccation of the environment, reflecting water deficit in soil, requires tools that would enable the breeders to carry out selection of genotypes resistant to drought. The project consists of 23 research tasks, carried out under POLAPGEN Consortium whose partners are 10 research units and 2 breeding companies. The tasks were formulated in such a way that a wide range of characteristics determining plants' resistance to drought is studied. All the tasks are realized on the same plant material, which integrates many research teams and enables to evaluate interdependence of various parameters. A systems approach is achieved by adopting a model of tolerance of plants to drought stress containing ecophysiological, morphological, anatomical, metabolic, proteomic, and molecular levels considered in the context of genetics. The project will allow to set new markers for resistance to drought, both molecular and morphological, as well as new methods of evaluation of resistance based on physiological and physical indicators. Data obtained in the experiments will also help to create the ideotype of resistant varieties, characterized by the complex of characteristics and properties at the whole plant level and at the cellular and molecular level.

In this presentation the main features of the solutions applied in the POLAPGEN-BD project in the area of data analysis will be presented. The Biometry Group at Institute of Plant Genetics of the Polish Academy of Science is responsible for coordination of collection, processing, statistical analysis and integration of phenotypic and genotypic data. To achieve this task, the group constructed an infrastructure based on publically available or commercial tools and tools that were specially shared with the project by the developers. Data are submitted by the experimenters in special exchange formats and then processed in Genstat (http://www.vsni.co.uk) and Java scripts for input to the database, being an implementation of the Germinate MySQL (http://bioinf.scri.ac.uk) schema. Information stored in the database is subjected to standardization based on existing ontologies. Some observations need preprocessing, the most interesting example being metabolomic data obtained by GC-MS and UPLC/UV protocols, for which signal identification pipelines were designed. The database can be accessed by all partners via the interface constructed using Biomart software (http://www.biomart.org). Statistical analysis is done in Genstat or R scripts with direct connection to the database or acting on downloaded text files. Specially challenging is the analysis of observations done in multi-factorial designs by high-throughput technologies like microarrays or chromatography/mass spectrometry.

Project no. WND-POIG.01.03.01-00-101/08 carried out under Innovative Economy Programme 2007-2013, Action 1.3, Subaction 1.3.1. within the subject „Biological progress in agriculture and environment protection" (http://www.poig.gov.pl/). Some elements of the presentation refer to results obtained in collaboration with FP7 Capacities-Infrastructures project transPlant RI-283496 (http://transplantdb.eu).

PREDICTOMICS: from pedigrees and DNA to complex phenotypes

Gianola, Daniel[1,2,3]; Perez, Paulino[1]; Tusell, Llibertat[1]

1: Department of Animal Sciences, University of Wisconsin, Madison, USA
2: Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, USA
3: Department of Dairy Science, University of Wisconsin, Madison, USA

Abstract: Large amounts of genomic data (e.g., SNPs) are available in animals, humans and plants. An important issue, e.g., in animal breeding or personalized medicine, is that of prediction of complex traits, that is, "Predictomics". Animal breeders exploit this type of data together with pedigrees, but typically use prediction methods that conform to the assumption of additive inheritance. Knowledge from regulatory and systems biology suggests that interaction and non-linearity are pervasive: complex pathways involve many enzymes (genes) and Michaelis-Menten kinetics; if one enzyme "fails", the reaction stops. On the other hand, linear regression models used by breeders assume constant increments, without allowance for general interactions (statistical interactions are linear) or for adaptive, data driven, relationships. Although linear Bayesian regression models address the "small *n,* large *p*" problem, these are obviously wrong mechanistically. Hence, the claim that such methods help to understand "genetic architecture" is arguably inaccurate. If epistatic systems appear to behave as additive, how can one understand complexity from this pseudo-topology? Predictive models can be strengthened by use of non-parametric and machine learning methods; these have been used with success in the study of complex systems. This presentation reviews some experiences with reproducing kernel Hilbert spaces regressions (RKHS) and Bayesian neural networks (BNN), and suggests how Fisher's infinitesimal model can be extended to the non-linear domain. Results from many plant data sets indicate that a strong linear learner, the Bayesian Lasso, is clearly defeated by BNN and by RKHS. We also review an application in pigs combining kernel averaging with Bayesian model averaging. We tentatively conclude that there is no uniformly best predicting machine; that the relative performance of various Bayesian linear regression models reflects robustness rather than "genetic architecture", and that prediction of complex traits requires more flexible methods, although linear predictors are often good enough for the narrow objective of artificial breeding.

Part III

Abstracts for posters (P1 – P31)

Poster number	First author	Title
P1	Adetunji, Ibraheem Olalekan	Whole-genome genetic diversity and linkage disequilibrium analysis in sugar beet
P2	Alimi, Nurudeen Adeniyi	Mixed model multi-trait multi-environment QTL analyses; an application in pepper
P3	Alves, Mara Lisa	Building the basis for population breeding: Genetic structure and diversity of maize populations in Portugal
P4	Alves, Mara Lisa	Tracking down the changes in maize participatory breeding: Portuguese “Amiúdo” and “Castro Verde” populations
P5	Dehghani, Hamid	SAS program for estimating nonparametric measures stability
P6	Estaghvirou, Sidi Ould Boubacar	Evaluation of approaches for estimating prediction accuracy in genomic selection
P7	Ishiguro, Seiya	Evaluation of approaches for estimating prediction accuracy in genomic selection
P8	Kleinknecht, Kathrin	Comparison of the performance of BLUE and BLUP for zoned Indian maize data
P9	Lehermeier, Christina	Genomic prediction of testcross performance in multi-line crosses of maize (Zea mays L.)
P10	Lillemo, Morten	QTL mapping in difficult datasets with epistasis – a comparison of common mapping softwares using powdery mildew resistance in wheat as a case
P11	Lootens, Peter	A medium-throughput procedure to estimate growth parameters of field-grown forage grass plants
P12	Maenhout, Steven	Progeno: an integrated software system for phenotypic and genomic selection
P13	Mega, Ryosuke	ABA catabolism in spikelets at the booting stage is promoted under cold condition in cold-tolerant rice (*Oryza sativa* L.)

P14	Onogi, Akio	Prediction accuracy of rice agronomic traits using genome-wide markers and functional nucleotide polymorphisms under single- and multi-trait modeling
P15	Ota, Yuya	The comparative analyses on the seed performances in the rice hybrids with different combinations
P16	Safner, Toni	Genetic structure of Spanish pea landrace collection using ISSR markers
P17	Sato, Yutaka	Correlation of gene expression with seedling vigor under cold conditios in rice (*Oryza sativa* L.)
P18	Satovic, Zlatko	Association mapping of essential oil components in Dalmatian sage (*Salvia officinalis* L.)
P19	Sawikowska, Aneta	Comparison of classical and functional principal components applied to chromatographic data
P20	Schrag, Tobias A.	Comparing SNP, AFLP, and SSR markers for genetic diversity analysis of elite European maize inbred lines with focus on ascertainment bias
P21	Stankovic, Goran	F_2 Population and genetic gain for resistance to maize ear rot
P22	Studnicki, Marcin	Comparing efficiency of sampling strategies to establish a representative in phenotypic-based genetic diversity core collection of orchardgrass (*Dactylis glomerata* L.)
P23	Teyssèdre, Simon	Comparison of genomic selection models in maize
P24	Thorwarth, Patrick	A comparison of models for the prediction of genotypic values in self fertilizing plants
P25	van Berloo, Ralph	In silico defined breeding strategies applied in practice: Prospects and lessons learned
P26	van Heerwaarden, Joost	Genomic prediction under local adaptation
P27	Voorrips, Roeland E.	PedigreeSim: simulation of diploid and tetraploid meioses and pedigrees
P28	Welham, Sue	A menu-based pipeline for trial analysis and QTL detection
P29	Yabe, Shiori	Potential of genomic selection for mass selection breeding in allogamous crops for traits requiring selection before or after pollination
P30	Zinn, David	The challenges of organizing the data of a breeding program

Whole-genome genetic diversity and linkage disequilibrium analysis in sugar beet

Adetunji, Ibraheem Olalekan[1]; Willems, Glenda[1]; Bürkholz, Alexandra[1]; Boer, Martin[2]; Malosetti, Marcos[2]; Tschoep, Hendrik[1]; Horemans, Stefaan[1]; van Eeuwijk, Fred A.[2]

1: SESVanderHave NV/SA, Tienen, Belgium
2: Biometris, Wageningen, The Netherlands

Abstract: Genome-wide association studies (GWAS) are widely applied not only in model organisms but also increasingly in crop species, as they provide an efficient means of identifying genes underlying important agricultural traits. The analysis of genetic diversity and linkage disequilibrium (LD) are prerequisites of GWAS, especially in non-model organisms for which research in this area has generally been lacking. In this work a sugar beet (*Beta vulgaris*) germplasm collection comprising 234 elite breeding lines and 99 non-elite beet accessions was genotyped with 498 single nucleotide polymorphism (SNP) markers, of which 454 were retained for subsequent analyses. The genetic architecture of the elite breeding lines (calculated using STRUCTURE and a principal coordinate analysis (PCoA)) revealed the presence of two distinct groups, which correspond to the seed parent and the pollen parent breeding pools. No distinct subpopulation structure was identified for the non-elite accessions. LD patterns on all nine chromosomes of the sugar beet genome were highly affected by population structure in the elite lines but not in the non-elite beet accessions. After correcting for genetic relatedness, LD decayed within a distance of less than 2 to 5 cM on all chromosomes in the elite lines. When genetic relatedness was not taken into account, LD decay ranged from 7 to 8 cM on chromosomes 1, 2, 6 and 7, and did not decay for other chromosomes. Fifty percent of the 454 SNPs showed significant differences in allele frequencies between the seed parent and the pollen parent pool. These SNPs were distributed over all nine chromosomes and are indicative of the breeding history of the elite sugar beet lines used in this study. Strong selective pressure provides at least a partial explanation for the long range LD as observed in the elite lines when genetic relatedness was not accounted for. The results of the LD decay, which lies within the range of values identified in previous studies, show the importance of taking genetic relatedness into account when estimating LD, especially in crop species. Furthermore, in regions that have undergone strong selection, such as on chromosome 3 (on which important disease resistance loci are located), GWAS might fail to identify causal loci even if densely covered.

Mixed model multi-trait multi-environment QTL analyses; an application in pepper

Alimi, Nurudeen Adeniyi[1,2]; Bink, Marco[1]; Malosetti, Marcos[1]; Voorrips, Roeland[2]; Palloix, Alain[2]; van Eeuwijk, Fred A.[1]

1: Wageningen UR, Biometris, Wageningen, The Netherlands
2: INRA PACA, Montfavet Cedex, France

Abstract: Plant breeding experiments often involve the collection of information on a number of correlated traits. Proper QTL mapping can then show whether the correlation is due to pleiotropic QTL(s) or genetic linkage. A further complexity in QTL analysis is introduced by a requirement to map QTLs for multiple environments. Mixed model QTL approaches offer a suitable framework for handling such complexities. Multi-trait multi-environment (MTME) QTL models help to identify the genomic regions responsible for genetic correlations, whether caused by pleiotropy or genetic linkage, and can show how genetic correlations depend on the environmental conditions. We used a MTME mixed model QTL analysis to analyse the genetic basis of yield and component traits in pepper as part of the EU-SPICY project. A RIL population (n=149) from an intraspecific cross between 'Yolo Wonder' (large – fruited bell pepper) and 'Criollo de Morelos 334' (small-sized hot pepper) was phenotyped at two locations during two seasons. The marker data comprised 455 markers on 12 chromosomes. The results from our MTME analysis revealed a number of QTLs influencing yield and component traits. Some of the QTLs were pleiotropic, with patterns consistent with trait's genetic correlations. Both consistent and environment-specific QTLs were identified. As an example, the number of QTLs for yield detected in each environment vary from seven to nine. Three of the QTLs were consistently picked up in the four environments with the other QTLs being environment-specific. Total explained trait variance by the QTLs varied between 37% and 54%. Comparing the QTL results from MTME and those from the initial STSE (single trait single environment) revealed up to a doubling of the number of QTL identified and the variance explained by these QTL. These results confirm the usefulness of sophisticated mixed model QTL methodology to increase our understanding of complex traits and our ability to use QTL in genome-assisted breeding.

References

Alimi NA, Bink MCAM, Dieleman JA, van Eeuwijk FA, et al. (2012). Genetic and QTL analyses of yield and a set of physiological traits in pepper. Submitted.

Alimi NA, Bink MCAM, Malosetti M, van Eeuwijk FA, et al. (2012). Multi-environments and multi-traits QTL analyses of pepper traits using mixed model approach. To be submitted.

Boer MP, Wright D, Feng LZ, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007). A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177, 1801-1813

Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA (2008). A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (Zea mays L.). Euphytica 161, 241-257

Van Eeuwijk FA, Bink M, Chenu K, Chapman SC (2010). Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology 13, 193-205

Building the basis for population breeding: Genetic structure and diversity of maize populations in Portugal

Alves, Mara Lisa[1]; Brites, Cláudia[2]; Paulo, Manuel[2]; Afonso, Ana Catarina[1]; Mendes-Moreira, Pedro[1,2]; Šatović, Zlatko[3]; Vaz Patto, Maria Carlota[1]

1: Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal
2: Escola Superior Agrária de Coimbra. Departamento de Ciências Agronómicas, Coimbra, Portugal
3: Faculty of Agriculture, Department of Seed Science and Technology, University of Zagreb, Zagreb, Croatia

Abstract: Maize was introduced in Portugal in the XVI century and many traditional landraces have been developed since then, adapted to specific regional growing conditions, as well as farmers needs. Nowadays, the enduring landraces are mainly flint type open pollinated varieties (OPV) with technological ability for production of the traditional maize leavened bread called “broa” (Brites et al., 2010). Only a few preliminary studies have been developed on the characterization of the breeding potential of this germplasm (Vaz Patto et al., 2007, 2009).

Under the SOLIBAM European project - Strategies for Organic and Low-input Integrated Breeding and Management, we thoroughly analysed the phenotypic and molecular diversity of 35 Portuguese maize populations (enduring traditional landraces and participatory breed OPVs). Our objectives were:

1) Analyse the phenotypic and molecular diversity within and among the maize populations

2) Test for deviations from Hardy-Weinberg equilibrium, at individual *loci*

3) Test for linkage disequilibrium, between pairs of *loci*

Thirty individuals per population were fingerprinted with 20 microsatellite markers uniformly distributed throughout the genome. The phenotypic HUNTERS evaluation and yield were measured on a plot basis on the multi-location trials (6-9 locations).

The studied populations will be classified into distinct clusters using three different multivariate approaches (principal component, cluster and discriminating analysis). Analysis of molecular variance will be used to test the existence of significant differences, at the genetic level, between the identified phenotypic clusters.

This study will allow us to characterize the genetic structure of these populations, establishing the basis for an efficient use of this germplasm in breeding programs, and to evaluate the suitability of these populations for association mapping studies.

References

Brites C, Trigo MJ, Santos C, Collar C, Rosell CM (2010). Maize-Based Gluten-Free Bread: Influence of Processing Parameters on Sensory and Instrumental Quality. Food and Bioprocess Technology 3, 707-715

Vaz Patto MC, Mendes-Moreira P, Carvalho V, Pêgo S (2007). Collecting maize (Zea mays L. convar. mays) with potential technological ability for bread making in Portugal. Genetic Resources and Crop Evolution 54, 1555-1563

Vaz Patto MC, Alves ML, Almeida NF, Santos C, Mendes-Moreira P, Šatović Z, Brites C (2009). Is the bread making technological ability of Portuguese traditional maize landraces associated with their genetic diversity? Maydica 54, 297-311

Tracking down the changes in maize participatory breeding: Portuguese “Amiúdo” and “Castro Verde” populations

Alves, Mara Lisa[1]; Belo, Maria[2]; Carbas, Bruna[3]; Brites, Cláudia[4]; Paulo, Manuel[4]; Spencer, Graciano[4]; Mendes-Moreira, Pedro[1,4]; Brites, Carla[3]; Bronze, Maria do Rosário[1,2,5]; Pego, Silas[6]; Šatović, Zlatko[7]; Vaz Patto, Maria Carlota[1]

1: Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal
2: Faculdade de Farmácia, Universidade de Lisboa, Av. das Forças Armadas, Lisboa, Portugal
3: Instituto Nacional de Recursos Biológicos, I.P, L-INIA, Unidade de Tecnologia Alimentar , Oeiras, Portugal
4: Escola Superior Agrária de Coimbra. Departamento de Ciências Agronómicas, Coimbra, Portugal
5: Instituto de Biologia Experimental e Tecnológica, Oeiras, Portugal
6: Fundação Bomfim, Braga, Portugal
7: Faculty of Agriculture, Department of Seed Science and Technology, University of Zagreb, Zagreb, Croatia

Abstract: Portugal has a long term participatory plant breeding (PPB) programme running since 1984 – the VASO project. This project aims on increasing yield and improving quality of local maize landraces known for their ability for bread production, while respecting traditional agriculture, accepting low input and intercropping, and favouring diversity in a way that local genetic resources could be competitive and maintained on-farm.

Our objectives were to evaluate the impact of the VASO PPB approach at molecular and phenotypic level. For that we analysed the genetic, agronomic and quality evolution of two regional OPV populations that integrated this project: “Amiúdo”, a yellow flint early maturity variety (FAO 200), adapted to stress conditions (aluminium toxicity and water deficit); and “Castro Verde”, a orange flint late maturity variety (FAO 600), characterized for having taller plants (Mendes-Moreira, 2006). Both populations were expected to be potentially interesting for food and “Castro Verde” for a double food and feed use.

Using 20 SSR uniformly distributed throughout the genome, we genotyped 30 individuals from three selection cycles of “Castro Verde” population (representing 15 years of on-farm mass selection) and three selection cycles of the “Amiúdo” population (representing 25 years of on-farm mass selection) comparing the latter with the third cycle of the breeder’s recurrent selection by S1 lines. Yield, technological (e.g. flour viscosity), nutritional (fibre, fat and protein) and organoleptic (carotenoids, phenolics and antioxidant activity) quality traits were also studied on the same selection cycles. Multi-location trials (6-9 locations) were established for yield evaluation, being the samples for quality evaluation collected from one only location.

Allelic richness and gene diversity will be assessed for each cycle, and genic and genotypic differentiation among cycles tested. Regression analysis on the yield response to selection will be discussed. AMOVA will be used to partition the total genetic variance according to the available phenotypic data, as well as to study the genetic diversity within and among different selection cycles.

This analysis is fundamental to clarify if this PPB program is being successful in conserving diversity and improving quality and yield on these two OPVs as it was already shown for the “Pigarro” OPV’s genetic diversity conservation, also breed under this approach (Vaz Patto et al., 2008).

References

Mendes-Moreira P (2006). Participatory maize breeding in Portugal. A case study. Acta Agronomica Hungarica 54, 431-439

Vaz Patto MC, Mendes-Moreira P, Almeida NF, Šatović Z (2008). Genetic diversity evolution through participatory maize breeding in Portugal. Euphytica 161, 283-291

SAS program for estimating nonparametric measures stability

Dehghani, Hamid[1]; Akbarpour, Omid Ali[1]; Kang, Manjit[2]

1: Faculty of Agriculture, Department of Plant Breeding, Tarbiat Modares University, Tehran, Iran
2: Vice Chancellor, Punjab Agricultural University, Ludhiana, India

Abstract: To estimate phenotypic stability of genotypes, several nonparametric methods have been proposed on the basis of genotype ranks from different environments. Non-parametric measures of stability play an important role in plant breeding, especially in genotype-performance trials conducted in the final stages of breeding programs. In this paper, SAS code is provided to compute non-parametric statistics, such as $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$, $S_i^{(6)}$, $NP_i^{(1)}$, $NP_i^{(2)}$, $NP_i^{(3)}$ and $NP_i^{(4)}$, stratified ranking technique and Kang's rank-sum. Because computations of these statistics can be cumbersome, an easy-to-use computer program was needed. Such a program was developed using SAS software to compute these nonparametric statistics. The program deals with two-way data with K genotypes and n environments.

Evaluation of approaches for estimating prediction accuracy in genomic selection

Estaghvirou, Sidi Ould Boubacar[1]; Ogutu, Joseph[1]; Piepho, Hans-Peter[1]

1: Institute of Crop Science, Department of Bioinformatics, University of Hohenheim, Stuttgart, Germany

Abstract : Genomic selection (GS) is getting increasingly used to make selection decisions in plant and animal breeding programs. GS is involves predicting genomic breeding values using molecular markers covering the whole genome. The accuracy of genomic selection is often measured by the correlation between the observed phenotypic and the predicted genomic breeding values (predictive ability), because the true breeding value is generally unknown in practice.

Predictive ability is often divided by the square root of heritability to obtain an approximate measure of accuracy called predictive accuracy. We assessed the reliability of estimates of predictive accuracy in genomic selection in plant and animal breeding using simulation and estimated parameters for the simulation model based on real datasets.

We consider five different approaches for estimating heritability, two of which are new and are described here for first time. Overall, predictive accuracy was estimated using seven different methods, two of which are direct and involve the use of simulated true breeding values whereas the remaining five methods involve first computing predictive ability and then dividing it by the square root of heritability, computed using each of the five different methods.

Predictive accuracies for five of the seven methods were assessed using 5-fold cross-validation each replicated five times. Each of the seven estimates of predictive accuracy was compared with the simulated accuracy as the benchmark.

Evaluation of the genome-wide expressions against the cool stress at the rice booting stage

Ishiguro, Seiya[1]; Ogasawara, Kei[1]; Sato, Yutaka[2]; Kishima, Yuji[1]

1: Laboratory of Plant Breeding, Graduate School of Agriculture, Hokkaido University, Sapporo, Japan
2: National Agricultural Research Center for Hokkaido Region, Sapporo, Japan

Abstract: The cool weather causes the floral impotency in rice, of which damage is a serious problem for the rice cultivation in high latitude regions. The damages vary among the rice cultivars, implying that the floral impotency due to cool weather is a genetic character. In the rice anther, a number of the genes affected by low-temperature have been detected, though the complexity of the gene networks is yet an obstacle to be studied for cool tolerance. In our study, we focus on repetitive sequences, which are a major component of the genome organization, and not associated with such signal transductions. Because the repeat sequences are normally unrelated to the gene networks, the expressions of these sequences may be utilized as an indicator of perception of the environmental changes without associations with the complicated regulations of various gene expressions. About 32,000 repetitive sequences were mined from the rice databases and employed for micro-array analyses. We used five lines differing in the degree of cooling tolerance at the booting stage. The data showed that oscillations of the genome-wide expressions in the rice anther transcripts were detected when the rice plants at the booting stage were exposed to low temperature. Interestingly, the degrees of oscillations of the genome-wide expressions were different among the five lines and coordinated with the degree of the pollen sterility. The changes of the genome-wide expressions in the sensitive lines at the low temperature were greater than those in the tolerant lines. These results led us to consider the oscillations of the genome-wide expressions as an indicator for cool stress in the rice anther. This work was supported by the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN).

Comparison of the performance of BLUE and BLUP for zoned Indian maize data

Kleinknecht, Kathrin[1]; Möhring, Jens[1]; Singh, Krishan Pal[2]; Zaidi, Pervez H.[3]; Atlin, Gary N.[4]; Piepho, Hans-Peter[1]

1: Institute of Crop Science, Bioinformatics Unit, University of Hohenheim, Stuttgart, Germany
2: Directorate of Maize Research, Pusa Campus, New Delhi, India
3: CIMMYT- ARMP, C/o ICRISAT, Patancheru, Hyderabad, India
4: CIMMYT, Mexico, D.F., Mexico

Abstract: The maize growing area in India is divided into five zones for maize testing and evaluation. During triannual testing of maize genotypes in official trials within the All-India Coordinated Maize Improvement Program (AICMIP) a high number of entries are rejected each year. Thus, only a very low number of entries is carried forward to the advanced stage of performance testing. The subdivision of the breeding sites into zones means that the amount of data per zone is limited. Hence, the question arises how to select the best genotypes per zone and how information can be borrowed across zones to improve the accuracy of selection within zones. To address this problem, we compared the performance of Best Linear Unbiased Prediction (BLUP) using the correlation of genetic effects between zones with Best Linear Unbiased Estimation (BLUE) based on data per zone. In both cases data were analysed using a mixed model. We used simulations to calculate correlations between the true simulated values and the predicted genotype values obtained by BLUE and BLUP using the same models. Both the data structure and the variance components used in simulations were based on the analysis of 40 triannual series of four different maize maturity groups. BLUP outperformed BLUE in 38 out of 40 series and on average over all series. An advantage of BLUP was observed for varying genetic correlations between zones. We conclude that the use of BLUP enhanced the estimation accuracy in zoned AICMIP maize testing trials and can be recommended for future use in these trials.

Genomic prediction of testcross performance in multi-line crosses of maize (Zea mays L.)

Lehermeier, Christina[1]; Bauer, Eva[1]; Schönleben, Manfred[1]; Walter, Hildrun[1]; Bauland, Cyril[2]; Camisan, Christian[3]; Campo, Laura[4]; Meyer, Nina[5]; Ranc, Nicolas[6]; Schipprack, Wolfgang[7]; Altmann, Thomas[8]; Flament, Pascal[3]; Melchinger, Albrecht E.[7]; Menz, Monica[6]; Moreno-González, Jesús[4]; Ouzunova, Milena[5]; Charcosset, Alain[2]; Schön, Chris-Carolin[1]

1: Plant Breeding, Technische Universität München, Freising, Germany
2: INRA, UMR de Génétique Végétale, Gif-sur-Yvette, France
3: France Limagrain Verneuil Holding, Riom, France
4: Centro Investigacións Agrarias Mabegondo (CIAM), La Coruña, Spain
5: KWS SAAT AG, Einbeck, Germany
6: Syngenta S.A.S., Saint-Sauveur, France
7: Plant Breeding, Universität Hohenheim, Stuttgart, Germany
8: Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Abstract: Advances in high throughput marker technologies have stimulated genomic prediction in plant breeding. As has been shown in the literature, genomic prediction can be performed with high accuracy in large biparental populations due to a high degree of relatedness between individuals, balanced allele frequencies, absence of population structure and consistent linkage phase between markers and quantitative trait loci. However, in elite breeding programs selection is performed among large numbers of crosses with small to moderate family sizes and varying degrees of relatedness between crosses.

We investigated different genomic prediction strategies using progenies from eleven flint and ten dent biparental families that were phenotyped as testcrosses in six and four environments, respectively. All 1655 doubled-haploid lines were genotyped with the MaizeSNP50 BeadChip from Illumina. Predictive abilities were assessed by genome-based best linear unbiased prediction and cross-validation. Our analyses focused on three quantitatively inherited traits: dry matter yield, dry matter content and plant height. We compared the predictive abilities within and across biparental families, as well as across heterotic groups. By categorizing the marker loci according to their physical position on the chromosomes we analyzed the effect of different linkage disequilibrium levels on the prediction of testcross performance.

Prediction across families yielded similar predictive abilities compared to those obtained within biparental families. As expected, predictive abilities were low for crosses without connectivity due to pedigree. Our results suggest that depending on the trait of interest, the genetic material and the breeding design, flexible solutions for implementing genomic prediction in breeding will be required.

QTL mapping in difficult datasets with epistasis – a comparison of common mapping softwares using powdery mildew resistance in wheat as a case

Lillemo, Morten[1]

1: Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, Aas, Norway

Abstract: During the last couple of decades, QTL mapping has become a routine task in most genetic studies on quantitative traits in plants, and researchers have a multitude of QTL mapping softwares to choose from. As many software packages have user-friendly interfaces and presents nice graphs and tables, very little knowledge in statistics and mathematics is required for using them. Although there has been an evolution in the algorithms used to map QTL in terms of better performance, ability to account for the effects of multiple QTL in the same model, inclusion of epistasis and QTL-by-environment interactions etc., additive models are assumed. The objective of the present study was to compare the performance of different QTL mapping algorithms in identifying QTL for powdery mildew resistance in a wheat Doubled Haploid (DH) population exhibiting strong epistatic interactions. A data set comprising powdery mildew data from three environments on 93 DH lines from the Arina x NK93604 population was used. The population has previously been used for mapping QTL for Fusarium Head Blight (FHB) resistance (Semagn et al., 2007) and has a linkage map of 624 markers covering all 21 chromosomes (Semagn et al., 2006). The parents are believed to carry different partially defeated race specific resistance genes to powdery mildew and histograms of the phenotypic data of the DH lines showed bimodal distributions with about a quarter of the lines being near-immune, which is a strong indication of digenic epistasis. The following mapping algorithms and software packages were tested and compared on the data set: Simple interval mapping and composite interval mapping using PlabQTL (Utz and Melchinger, 2003) and MapQTL 6 (Van Ooijen, 2009), inclusive composite interval mapping using QTL IciMapping (Li et al., 2007) and a mixed linear mapping approach using QTLNetwork v 2.0 (Yang et al., 2007). A stunning result from this work was that the different mapping approaches yielded very different results and only a relatively small proportion of the genetic variation could be explained by the identified QTL despite high within environment heritabilities and good marker coverage of most chromosomes. It is concluded that standard QTL mapping approaches are not suitable for handling this type of data sets.

References

Li HH, Ye GY, Wang JK (2007). A modified algorithm for the improvement of composite interval mapping. Genetics 175, 361-374

Semagn K, Bjørnstad A, Skinnes H, Marøy AG, Tarkegne Y, William M (2006). Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49, 545-555

Semagn K, Skinnes H, Bjørnstad A, Marøy AGM, Tarkegne Y (2007). Quantitative Trait Loci controlling Fusarium Head Blight resistance and low deoxynivalenol content in hexaploid wheat population from 'Arina' and NK93604. Crop Sci 47, 294-303

Utz HF, Melchinger AE (2003). PLABQTL: A computer program to map QTL, Version 1.2. Institute of plant breeding, seed science and population genetics, University of Hohenheim, Stuttgart, Germany

Van Ooijen JW (2009). MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands.

Yang J, Zhu J, Williams RW (2007). Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23, 1527-1536

A medium-throughput procedure to estimate growth parameters of field-grown forage grass plants

Lootens, Peter[1]; Ruttink, Tom[1]; Rohde, Antje[1]; Carré, Serge[2]; Combes, Didier[2]; Barre, Philippe[2]; Roldán-Ruiz, Isabel[1]

1: ILVO – Plant Sciences Unit, Melle, Belgium
2: INRA – UR4 P3F, BP 6, Lusignan, France

Abstract: Association and QTL mapping studies in agricultural crops require phenotypic characterization of large, replicated collections of plants. For the screening of traits such as (re)growth potential or persistency of perennial forage grass species, phenotyping of field-grown adult plants is required. Here we present a simple and robust medium-throughput procedure to estimate growth parameters of *Lolium perenne* (perennial ryegrass) plants, based on digital image analysis.

The purpose of this study was to describe the (re)growth potential of a *L. perenne* association mapping population, grown at two locations (Melle-Belgium and Lusignan-France) during two seasons. The plants were subjected to a mowing regime that simulated pasture exploitation. Parameters derived from top-view and side-view images were used to describe plant volume, habitus and geometry in ways that single manual measurements of plant height or diameter alone could not achieve. For the top-view images we combined information extracted from pictures taken at bi-monthly intervals (to capture ground coverage changes) and pictures taken at weekly intervals (to capture leaf elongation rate). Dedicated analysis procedures were developed to analyze time series of pictures and to extract relevant growth dynamics parameters.

Here we will present a methodology to overcome technical problems associated to the use of digital images taken under non-standardized open-air conditions, and the image analysis procedures developed.

Progeno: an integrated software system for phenotypic and genomic selection

Maenhout, Steven[1,2]; De Baets, Bernard[1]; Haesaert, Geert[2]

1: Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Gent, Belgium
2: Department of Plant Production, University College Ghent, Gent, Belgium

Abstract: As the price tag of high-density genotyping comes down, genomic selection gradually finds its way into breeding programs of various plant and animal species. Widespread adoption of this selection approach is, however, hindered by a lack of user-friendly software that allows practical breeders to distil reliable breeding values from their available phenotypic and genotypic data collections. The increasing density of commercial SNP chips and the computational challenges this imposes on existing software solutions forms a second hurdle which is generally difficult to overcome without considerable concessions with respect to the dimensionality of the initial problem.

These two obstacles have been the key motivation for the development of the Progeno software system. In its core, Progeno is a linear mixed model computing engine which was written from scratch to solve problems involving millions of unbalanced phenotypic observations in combination with dense molecular marker profiles containing tens, if not hundreds of thousands of individual marker scores. It offers ample model flexibility including a wide range of possibilities for imposing a predefined structure on the variance of random effects and residuals including unstructured, compound symmetry, autoregressive, anisotropic and many other variance and correlation structures. Pedigree and molecular marker information can be integrated in the variance structure of random effects using various approaches ranging from a classic numerator relationship matrix to more recent advances such as Reproducing Kernel Hilbert Space Regression. Unknown variance parameters are estimated from the data by means of Average Information Restricted Maximum Likelihood. The computational workload can be spread over multiple processors and flexible out-of-core storage techniques avoid having to compromise with respect to data quantity or model complexity due to computer memory limitations.

The Progeno computing engine allows to integrate all the available phenotypic and molecular data that has been gathered over many years of breeding. For each phenotypic trial, the linear mixed model formulation that produces the best fit to the data is automatically selected using a subset of a predefined selection of candidate model terms and residual variance structures. The specific model details of each trial are combined in an all-encompassing meta model after which all variance parameters are re-estimated producing the optimal model for the available data. Outlier detection and cross-validation routines allow to identify suspicious observations and validate the quality of the resulting model respectively.

The resulting prediction models can be valorised by breeders by means of a user-friendly web application that allows to explore raw, corrected and aggregated trial results and to obtain breeding values for individuals with phenotypic observations as well as pedigree- or genomic-based predictions for unobserved and non-existent individuals or crosses. The system also generates optimal cross-advice when provided with a set of candidate parental individuals and a weighing of the relative importance of each trait.

The industrial applicability of the presented Progeno technology has been validated in commercial breeding programmes of maize, winter oilseed rape, potatoes, pigs and orchids and proof-of-concept projects for various other species are ongoing.

ABA catabolism in spikelets at the booting stage is promoted under cold condition in cold-tolerant rice (*Oryza sativa* L.)

Mega, Ryosuke[1]; Meguro, Ayano[1]; Sato, Yutaka[1]

1: National Agricultural Research Center for Hokkaido Region, Sapporo, Japan

Abstract: Rice is a staple food for more than half of the human population in the world. The growth of rice is hampered by various problems such as low temperature, fungal infection, insect damage and drought. The reproductive stage is particularly sensitive to abiotic stress damage (cold, drought, heat). Low temperature at the reproductive stage is a yield-limiting factor in all temperate rice-growing areas of the world, and an estimated 7 million hectares worldwide are prone to damage by cold. Irreversible pollen sterility is mainly caused by exposure to low temperatures at the booting stage in rice.

To identify the genes involved in cold tolerance at the booting stage in rice, we investigated gene expression in spikelets in the cold-sensitive cultivar 'Toyohikari', the cold-tolerant cultivar 'Hayayuki' and nearly isogenic lines (NILs) that were developed by recurrent back-crossing of 'Toyohikari' to 'Hayayuki'. The results showed that low temperature induced the expression of ABA biosynthetic genes in 'Toyohikari' but not in 'Hayayuki' or NILs. In contrast, ABA catabolic genes were induced by cold treatment in 'Hayayuki' and NILs but not in 'Toyohikari'. Also, expression of an ABA responsive gene increased with ABA biosynthetic genes in 'Toyohikari'. Under cold conditions, endogenous ABA content of the spikelets increased in 'Toyohikari', while the content remained similar to the control condition in 'Hayayuki' and NILs. These results suggested that cold-sensitive rice cultivars tend to accumulate endogenous ABA by cold treatment and to be sensitive to ABA, while cold-tolerant rice cultivars tend to catabolize endogenous ABA. Furthermore, *in situ* hybridization showed that an ABA catabolic gene is expressed distinctively in the tapetum at the vacuolated pollen stage rather than at the young microspore stage. Our results indicate that effective ABA catabolism in the tapetum might contribute to improvement in cold tolerance at the booting stage in rice.

This work was supported by the Programme for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry.

Prediction accuracy of rice agronomic traits using genome-wide markers and functional nucleotide polymorphisms under single- and multi-trait modeling

Onogi, Akio[1]; Osama, Ideta[2]; Kaworu, Ebana[3]; Takuma, Yoshioka[4]; Masanori, Yamasaki[4]; Iwata, Hiroyoshi[1]

1: Laboratory of Biometry and Bioinformatics, Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan
2: National Agricultural Research Center for Western Region, National Agriculture and Food Research Organization, Tsukuba, Ibaraki, Japan
3: Genetic Resources Center, National Institute of Agrobiological Sciences,Tsukuba, Ibaraki, Japan
4: Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University, Kobe, Japan

Abstract: Prediction of phenotypes or breeding values is an important issue in plant and animal breeding, and recently, genome-wide markers are often utilized for this purpose. In this study, we investigated the prediction ability based on genome-wide markers and/or functional nucleotide polymorphisms (FNPs) in four agronomic traits of rice, *Oryza sativa* L. ssp. *japonica*. The four traits were days to heading (DH), culm length (CL), panicle length (PL) and number of panicles (PN). The FNPs previously identified for DH and CL were expected to affect the other traits through pleiotropy or physiological relationships among the phenotypes of traits. We focused on whether the prediction accuracies could be improved by 1) including FNPs in the models in addition to genome-wide markers, and by 2) predicting multiple traits simultaneously. The results of leave-one-out cross-validation show that for DH, CL and PL, prediction using single-trait modeling was the most accurate when using both FNPs and genome-wide markers, and for PN, the most accurate when using only genome-wide markers. When these four traits were predicted simultaneously under multi-trait modeling, the prediction accuracies of DH and CL were equivalent to those under single-trait modeling. Interestingly, the accuracy of PL increased, while that of PN decreased, under multi-trait modeling. When a subset of the traits was predicted simultaneously, the combinations of DH, CL and PL and of PL and PN slightly improved the accuracy of PL. On the other hand, the accuracies of PN decreased in any combinations that we tested. These results suggest that 1) the inclusion of FNPs in a prediction model in addition to genome-wide markers is effective to increase prediction accuracy, and that 2) multi-trait modeling doesn't always result in improving prediction accuracy.

The comparative analyses on the seed performances in the rice hybrids with different combinations

Ota, Yuya[1]; Ogasawara, Kei[1]; Kishima, Yuji[1]

1: Laboratory of Plant Breeding, Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan

Abstract: Rice (*Oryza sativa*) is an autogamous crop, but F1 hybrids are also available in the practical production. However, little is known about detailed traits involved in vigor in the rice hybrids. Particularly, F1 seed performances have not been investigated so far. We have produced F1 hybrids between Nipponbare and a series of differentially related lines. Here, we compared seed weight, embryo length and germination performances. Nipponbare was used in all the crosses as one of the parents with the 11 lines consisting of four japonica lines (A58, T65, Koshihikari and Kitaake), three indica lines (IR36, Kasalath and #108), three *O. rufipogon* lines (W107, W593 and W630) and one *O. glaberrima* (WK18). In every combination, the reciprocal cross was carried out, so the maternal effect was examined in a total of 22 crossed lines. In the seed weight, most of the hybrids did not exceed the values of both the parents. The lower seed weight in the rice hybrids relative to the parents implies the suppression of the endosperm due to imprinting of double fertilization. While the embryo size showed vigor in the hybrids, which were associated with cytoplasmic effect, the hybrids between Nipponbare and the japonica lines had the larger embryo when Nipponbare was used as pollen, whereas in the hybrids with the other combinations, the larger embryos appeared with Nipponbare used as the seed parent. The shoot length was observed for three days after onset of germination. The hybrids crossed with the japonica or indica lines did not elongate the shoot beyond the lengths of both the parents. However, the hybrids with *O. rufipogon* or *O. glaberrima* revealed superior shoot length compared with the parents in the case where Nipponbare was used as seed parents compared with the parent. These results demonstrated that the seeds in the rice hybrids were unable to show universal performances in terms of hybrid vigor, but the superior performances over the parents occurred dependent on particular conditions. We suggest that full heterosis characters in rice hybrids are not yet exerted in the juvenile stage during the rice plant development.

Genetic structure of Spanish pea landrace collection using ISSR markers

Safner, Toni[1]; Caminero Saldaña, Constantino[2]; De la Rosa, Lucia[3]; Lázaro, Almudena[4]

1: University of Zagreb, Faculty of Agriculture, Zagreb, Croatia
2: Unidad de Cultivos Herbáceos, Instituto Tecnológico Agrario de Castilla y León (ITACyL), Valladolid, Spain
3: Centro Nacional de Recursos Fitogenéticos (CRF-INIA), Alcalá de Henares, Madrid, Spain
4: Departamento de Investigación Agroalimentaria IMIDRA - Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario, Alcalá de Henares, Madrid, Spain

Abstract: Landraces or traditional varieties are local populations, maintained and multiplied by farmers. They are simultaneously adapted to local environmental conditions through the process of natural selection, and continuously selected for desirable traits (Zeven, 1998). Their dissemination is not necessarily limited to a particular region, and successful landraces are propagated across natural and political borders through seed exchange and trade. However, increased seed exchange can put the adaptation of landraces at risk (Parzies et al., 2004) Therefore, identification of patterns of adaptation and magnitude of seed exchange traceable in genebank collections are of the crucial importance for developing of appropriate conservation strategy.

The Spanish pea landraces collection is conserved in CRF-INIA and ITACyL includes 439 accessions collected in Spain, in a wide geographic area of different ecological conditions. Information of accessions within collection includes passport and characterization data (available at www.inia.es). A subset of 289 accessions (250 of them are analyzed in this contribution) has been also characterized by means of ISSR markers by using 10 ISSR which have produced a total of 112 polymorphic bands. The number of bands per primer ranges from 7 to 15. For measuring inter-populational diversity, DNA was extracted and amplified following the methodology previously published. The diversity of the Spanish landraces has been previously studied by different authors (Lázaro & Aguinagalde, 2006; Martín-Sanz et al., 2011).

Classification of accessions based on molecular data was carried out by both excluding spatial information (to detect genetic stratification independent of geographic distribution) and including it (to detect assumptive pattern of regionalization). The extent of seed exchange was investigated by a correlogram between geographic distance and genetic coancestry.

Both Bayesian model-based clustering (STRUCTURE, Pritchard et al., 2000) and distance-based clustering (UPGMA on Jaccard's distance matrix) detected two distinct clusters. Neither of inferred clusters was confined to a specific geographic region nor related to any of environmental covariates. Furthermore, spatial model-based clustering using TESS (Durand et al., 2009) failed to detect any spatial cluster of accessions. A correlogram based on geographic distance classes and pairwise kinship coefficients indicates that seed exchange was practically unlimited throughout Spain, because only the most distant accessions were detected to be less related than random.

References

Durand E, Jay F, Gaggiotti OE, François O (2009). Spatial inference of admixture proportions and secondary contact zones. Molecular Biology and Evolution 26, 1963-1973

Lazaro A, Aguinagalde I (2006). Genetic variation among Spanish pea landraces revealed by Inter Simple Sequence Repeat (ISSR) markers: its application to establish a core collection The Journal of Agricultural Science 144, 53-61

Martin-Sanz A, Caminero C, Jing R, Flavell AJ, Perez de la Vega M (2011). Genetic diversity among Spanish pea (*Pisum sativum* L.) landraces, pea cultivars and the World *Pisum* sp. core collection assessed by retrotransposon-based insertion polymorphisms (RBIPs). Spanish Journal of Agricultural Research 9, 166-178

Parzies HK, Spoor W, Ennos RA (2004). Inferring seed exchange between farmers from population genetic structure of barley landrace Arabi Aswad from Northern Syria. Genet Resour Crop Ev 51, 471-478

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

Zeven AC (1998). Landraces: A review of definitions and classifications. Euphytica 104, 127-139

Correlation of gene expression with seedling vigor under cold conditios in rice (*Oryza sativa* L.)

Sato, Yutaka[1]; Ohashi, Mihoko[1]; Iwata, Natsuko[2]

1: National Agricultural Research Center for Hokkaido Region, Sapporo, Japan
2: Agricultural Research Institute, HOKUREN Federation of Agricultural Cooperatives, Hokkaido, Japan

Abstract: Vigorous rice growth during the seedling stage under cold conditions is an important trait for stable seedling establishment in the direct seeding method in temperate areas and at high altitudes in tropical and sub-tropical areas. In order to identify the genes involved in seedling vigor under cold conditions, comprehensive gene expression analysis was performed with recombinant inbred lines (RILs) derived from a cross between cultivars with high seedling vigor and low seedling vigor. A gene for which expression correlated with seedling vigor under cold conditions was indentified.

A highly vigorous Italian rice cultivar 'Italica Livorno', and a Japanese rice cultivar with low seedling vigor, 'Hayamasari', were used for evaluation of seedling vigor under cold conditions. Seeds were soaked in water at 28 ºC for 3 days and sown in soil in 20-cm-diameter plastic containers placed in growth chambers under a 16-hour photoperiod at 15, 17 and 19 ºC, respectively. Shoot length was measured at 7, 9, 11 and 14 days after sowing. Seedling vigor under cold conditions in RILs was evaluated by measuring shoot length of seedlings grown for 9 days in the growth chamber under a 16-hour photoperiod at 17 ºC. Microarray analysis was performed with RNAs isolated from shoots of 'Italica Livorno', 'Hayamasari', two highly vigorous RILs and two RILs with low vigor grown at 17 ºC for 9 days. Highly expressed genes in the highly vigorous cultivars and RILs were selected and then analyzed by quantitative real-time PCR using TaqMan probes with 22 highly vigorous RILs and 20 RILs with low vigor. Genes involved in seedling vigor under cold conditions were identified on the basis of significance level of the correlation coefficient between gene expression levels and shoot length at 17 ºC.

Shoot length in 'Italica Livorno' was significantly longer than that in 'Hayamasari' grown for 5 days at 19 ºC, for 7 days at 17 ºC and for 9 days at 15 ºC. Thereafter, the differences in shoot length of the two cultivars became large at each temperature. Seedling vigor under cold conditions of 108 RILs derived from a cross between 'Italica Livorno' and 'Hayamasari' was evaluated after growth at 17 ºC for 9 days. Thirty-one RILs were found to be significantly more vigorous than 'Hayamasari', and 33 RILs were significantly less vigorous than 'Italica Livorno'. Forty-four RILs showed seedling vigor between those of 'Italica Livorno' and 'Hayamasari'. A comprehensive microarray survey with two highly vigorous RILs, two RILs with low vigor, 'Italica Livorno' and 'Hayamasari' showed that 7 genes were expressed at higher levels in the highly vigorous RILs and the cultivar than in those with low vigor. The expression levels of these 7 genes were analyzed by quantitative real-time PCR using TaqMan probes with 22 highly vigorous RILs and 20 RILs with low vigor. The results showed that the expression level of 1 of the 7 genes, a cytochrome P450 gene, was significantly correlated with seedling vigor under cold conditions.

This work was supported by the Programme for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry.

Association mapping of essential oil components in Dalmatian sage (*Salvia officinalis* L.)

Šatović, Zlatko[1]; Liber, Zlatko[2]; Jug-Dujaković, Marija[3]; Radosavljević, Ivan[2]; Greguraš, Danijela[2]; Ristić, Mihailo[4]; Pljevljakušić, Dejan[4]; Dajić-Stevanović, Zora[5]; Gunjača, Jerko[1]

1: University of Zagreb, Faculty of Agriculture, Zagreb, Croatia
2: University of Zagreb, Facutly of Science, Zagreb, Croatia
3: Institute for Adriatic Crops and Karst Reclamation, Split, Croatia
4: Institute for Medicinal Plant Research "Dr. Josif Pančić", Belgrade, Serbia
5: University of Belgrade, Faculty of Agriculture, Belgrade, Serbia

Abstract: The mixed-model association mapping approach using kinship coefficients (K-matrix) and the population structure information (Q-matrix) was applied to identify molecular markers associated with eight major essential oil compounds in Dalmatian sage (*Salvia officinalis* L.). Twenty-five natural populations originating from Croatia (23) and Bosnia and Herzegovina (2), each consisting of 20 to 25 plants, were genotyped using AFLP and SSR markers. Four AFLP primer combinations yielded 559 polymorphic markers while by using 8 SSRs a total of 161 alleles were identified. Essential oils were analyzed using analytical gas chromatography (GC/FID) and gas chromatography-mass spectrometry (GC/MS) techniques. Out of 62 compounds detected, eight were found in concentrations higher than 5% in at least one sample (cis-thujone, camphor, trans-thujone, 1,8-cineole, β-pinene, camphene, borneol and bornyl acetate).

Data on each of eight compounds were subjected to mixed-model analysis using a range of K, Q and K + Q models based on different methods for construction of K and Q matrices. Pairwise kinship coefficients were based either on AFLP data (Hardy, 2003), or SSR data (Ritland, 1996; Loiselle et al., 1995) and estimated using the SPAGEDI software (Hardy and Vekemans, 2007). The population structure matrix (Q) based on SSR data was obtained by (a) STRUCTURE (Pritchard et al., 2000), (b) principal components analysis (PCA; Patterson et al., 2006), (c) principal co-ordinate analysis (PCoA) based the on proportion-of-shared-alleles distance matrix, and (d) factorial correspondence analysis (FCA) as implemented in Genetix 4.05 (Belkhir et al., 2004). For each variable (essential oil compound) the models were fitted in ASReml 3 (Gilmour et al., 2009) and compared using BIC. In most cases, K+Q models showed better goodness of fit than the other models. The models including K-matrix based on AFLP data generally fitted better than models including K-matrix based on SSR data. The best fitting models for different compounds varied in the Q part. However, "Hardy + STRUCTURE" model was chosen best in four out of eight cases.

References

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009). ASReml user guide release 3.0. VSN International Ltd., Hemel Hempstead, UK

Hardy OJ (2003). Estimation of pairwise relatedness between individuals and characterisation of isolation by distance processes using dominant genetic markers. Molecular Ecology 12, 1577-1588

Hardy OJ, Vekemans X (2007). SPAGeDi 1.2: a program for Spatial Pattern Analysis of Genetic Diversity - User's manual. Laboratoire Eco-éthologie Evolutive, Université Libre de Bruxelles, Belgium

Loiselle BA, Sork VL, Nason J, Graham C (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany 82, 1420-1425

Patterson N, Price AL, Reich D (2006). Population structure and eigenanalysis. PLoS Genet 2:e190

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

Ritland K (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res Camb 67, 175-185

Comparison of classical and functional principal components applied to chromatographic data

Sawikowska, Aneta[1]; Madrigal, Pedro[1]; Piasecka, Anna[1]; Kuczyńska, Anetta[1]; Ogrodowicz, Piotr[1]; Mikołajczak, Krzysztof[1]; Kachlicki, Piotr[1]; Krajewski, Pawel[1]

1: Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

Abstract: Due to a fast technical progress, chromatography is now applied on a large scale in research on biological systems, in particular in studies aimed at determination of concentration of metabolites in plants. The current approaches in metabolomics are usually untargeted, which means that differences between plant samples with respect to the whole spectrum of observed chemical compounds are studied. For such applications, two modes of data processing can be applied: a method based on signal detection, and a method based on whole chromatographic profiles. We analyse a set of chromatograhic data obtained using ultra performance liquid chromatography with the ultraviolet detection (UPLC/UV) in POLAPGEN-BD, a project aimed at finding biotechnological tools for breeding cereals with increased resistance to drought. We show differences and similarities between (a) applying classical multivariate techniques, in particular principal components, to the data set obtained by signal detection and (b) analysing the chromatographic profiles by the functional principal components technique. Relations between the two methods in terms discrimination of samples, ease of interpretation and visualisation are discussed.

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" (www.polapgen.pl). The work of P. Madrigal was supported by FP7 Marie Curie ITN SYSFLO, project number 237909 (www.sysflo.eu).

Comparing SNP, AFLP, and SSR markers for genetic diversity analysis of elite European maize inbred lines with focus on ascertainment bias

Schrag, Tobias A.[1]; Frascaroli, Elisabetta[2]; Melchinger, Albrecht E.[1]

1: Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany
2: Department of Agroenvironmental Sciences and Technologies, University of Bologna, Bologna, Italy

Abstract: Recent advances in high-throughput sequencing have triggered a shift toward single nucleotide polymorphism (SNP) markers, particularly for species with substantial genomic resources. Very large numbers of SNP markers are now available for detailed analysis of genome structure, genome wide association studies and precision breeding, especially for those species for which high density genotyping arrays are commercially produced (Ganal et al., 2011). However, a systematic bias (ascertainment bias) can be introduced if SNPs are ascertained in a small panel of genotypes and then used for characterizing a larger population (Clark et al., 2005; Hamblin et al., 2007; Moragues et al., 2010; Rafalski, 2011). The objective of this study was to evaluate a potential ascertainment bias of the Illumina MaizeSNP50 array (Ganal et al., 2011) with respect to elite European maize (*Zea mays* L.) dent and flint inbred lines. For this purpose, we compared the genetic diversity among these materials based on amplified fragment length polymorphisms (AFLPs), simple sequence repeat (SSRs), SNPs of the MaizeSNP50 array (SNP-A) and two subsets of it, i.e., Panzea (SNP-P) and Syngenta markers (SNP-S). We evaluated major allele frequency, allele number, gene diversity, modified Roger´s distance (MRD), molecular variance (AMOVA) and revealed mild ascertainment bias in SNP-A, compared to AFLPs and SSRs. The bias affected all genetic parameters, especially for European flint lines analyzed with markers SNP-S, specifically developed to maximize differences among North American dent germplasm (Ganal et al., 2011). However, the principal coordinate and Procrustes analyses for the different marker systems resulted in similar graphical representations of the structure, both in the dent and flint populations, and thus the bias did not substantially alter the relative distances between inbred lines within groups. Our results are in accordance with those of other researches (Hamblin et al. 2007; Jones et al., 2007; Lu et al. 2009; Hübner et al., 2012). For these reasons we conclude that the SNP markers of the MaizeSNP50 array can be employed for breeding purposes in the investigated material. However, attention should be paid in case of comparisons between genotypes belonging to different heterotic groups. In this case, it is advisable to prefer a marker subset with potentially low ascertainment bias, like in our case the SNP-P marker set.

References

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005). Ascertainment bias in studies of human genome-wide polymorphism. Genome Res 15, 1496-1502

Ganal MW, Durstewitz G, Polley A, Berard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J, Le Paslier MC, McMullen MD, Montalent P, Rose M, Schön CC, Sun Q, Walter H, Martin OC, Falque M (2011). A large maize (Zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28 334

Hamblin MT, Warburton ML, Buckler ES (2007). Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 2: e1367

Hübner S, Günter T, Flavell A, Fridman E, Graner A, Korol A, Schmid KJ (2012). Islands and streams: clusters and gene flow in wild barley populations from the Levant. Molecular ecology DOI: 10.1111/j.1365-294X.2011.05434.x

Jones E, Sullivan H, Bhattramakki D, Smith J (2007). A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize Zea mays L. Theor Appl Genet 115, 361–371

Lu Y, Yan J, Guimarães CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong T, Crouch JH, Xu Y (2009). Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120, 93–115

Moragues M, Comadran J, Waugh R, Milne I, Flavell AJ, Russell JR (2010). Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. Theor Appl Genet 120, 1525–1534

F_2 Population and genetic gain for resistance to maize ear rot

Stankovic, Goran[1]; Delic, Nenad[1]; Stankovic, Slavica[1]; Levic, Jelena[1]

1: Maize Research Institute, Zemun Polje, Belgrade-Zemun, Serbia

Abstract: The utilisation of genetically divergent initial material for the development of genotypes resistant to diseases, especially to root, stalk and ear rot, is emphasised in actual maize breeding programmes. As growing several millions plants in one combination of the F_2 population is impossible, but theoretically desirable when parents differ in a greater number of genes, it is necessary in practice to grow as many plants as possible in order to find out desirable gene recombinations. This experiment was conducted to study the effects of test-cross population size on genetic gain.

The investigation was carried out with the F_2 population derived from the cross of the inbred lines L982 and L588. The F_1 generation of crosses of 300 used plants to the inbred line L1325 was studied in the plot set according to the nested design with replications within sets, with 20 crosses within 15 sets, and three replications in two locations in two years. Plants were inoculated at 5 days post-silk emergence with inoculum prepared as previously described (Raid et al., 1992). Severity of ear rot symptoms was evaluated using a 7-scale rating scale. Genetic gain was estimated for the F_2 population of 100, 200, and 300 plants by the application of half-sib recurrent selection at the selection intensities of 5%, 10%, 20%.

The greatest genetic gain for all three selection intensities was detected in the population of 300 test hybrids. Differences in mean estimates for all levels of selection intensity in the population size of 200 and 300 were significant and highly significant, pointing to an advantage of higher intensity of selection. However, differences between selection intensities of 5% and 10% were not determined in the population of 100 test hybrids. Since a lower selection intensity is more favourable in genetically narrow-base populations, because of genetic variability conservation, the selection intensity of 10% can be recommended in this case.

Comparing efficiency of sampling strategies to establish a representative in phenotypic-based genetic diversity core collection of orchardgrass (*Dactylis glomerata* L.)

Studnicki, Marcin[1]; Madry, Wieslaw[1]; Schmidt, Jan[2]

1: Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences – SGGW, Warsaw, Poland
2: Botanical Garden, Plant Breeding and Acclimatization Institute - National Research Institute, Bydgoszcz, Poland

Abstract: The establishing of a core collection that represents genetic diversity of the entire collection with minimum loss of its original diversity and minimum redundancies is an important problem for gene-bank curators and crop breeders. In this paper we assessed the representativeness of the original genetic diversity in core collections consisting of one tenth of the entire collection, obtained according to 23 sampling strategies. The study was done using Polish orchardgrass *Dactylis glomerata* L. germplasm collection as a model. The representativeness of core collections was validated by the difference of means (MD%) and difference of mean squared Euclidean distance ($\overline{dD}$%) for studied traits in core subsets and the entire collection. This way we compared the efficiency of simple random and 22 (20 *cluster-based* and 2 direct cluster-based) stratified sampling strategies. Each *cluster-based* stratified sampling strategy is a combination of two clustering, five allocation and two methods of sampling in a group. We used accession genotypic predicted values for 8 quantitative traits, tested in field trials. A sampling strategy is considered more effective in establishing core collections if means of the traits in a core are maintained at the same level as means in the entire collection (mean of MD% in simulated samples is close to zero) and, simultaneously, when overall variation in a core collection is greater than in the entire collection (mean of $\overline{dD}$% in simulated samples is greater than that obtained for the simple random sampling strategy). Both cluster analyses (UPGMA and Ward) were similarly useful to construct those sampling strategies capable to establish representative core collections. Among the allocation methods relatively most useful for constructing efficient samplings were proportional and D2 (including variation). Within the Ward-clusters random sampling was better than cluster-based sampling, but not within the UPGMA-clusters.

Comparison of genomic selection models in maize

Teyssèdre, Simon[1]; Chauvet, Stéphanie[1]; Karaman, Zivan[1]

1: Limagrain Europe, Biostatistics Unit, Chappes, France

Abstract: The objective of this study was to compare several statistical models for genomic prediction using empirical maize data which included both unrelated panels and connected segregating populations.

The panel data included lines from two distinct heterotic groups and for each group one trait was studied. On these data, a set of 262 and 532 lines were phenotyped and genotyped with ~55K SNP markers using Illumina Maize SNP50 BeadChip.

The segregating populations data (referred to as MARS data) came from 38 datasets of different sizes, including on average 352 lines coming from 9 biparental populations and genotyped with an average of 376 SNP markers. These datasets included material from diverse genetic origins and involved more than 300 parental lines and close to 15,000 individuals that were evaluated in field trials and genotyped. A total of 34 traits were studied on these data.

Compared methods were Ridge-Regression (RR), Bayes A (BA), Bayes B (BB), Bayes C Pi with Pi random (BCPI), and Bayesian Lasso (BL). Comparison criteria were: 1) overall predictive ability of the models assessed by the R^2 between observed and predicted values after 10-fold cross-validation process, 2) the predictive ability by biparental population (for MARS data), 3) the regression coefficients of the observed on predicted values and 4) the time required for computations.

The predictive abilities (cross-validated R^2) varied a lot depending on the dataset and the trait, but reached the values close to 0.7. Except for BB, there were no differences between methods on observed predictive abilities (all were equal at +/- 1% with a tendency for RR to give slightly better results), and very high correlations of predicted values between methods were observed (close to 0.98). Bayes B, which assumes that only 10% of markers have an effect, produced slightly lower predictive abilities. The same results were observed for biparental populations. One trait on panel data, possessing a major QTL known from previous studies, contradicted these findings. For this trait, results with BA and BB were slightly better than with other methods. Concerning the regression coefficients, RR, BL and BCPI methods were more accurate. Finally, concerning the computation time, RR was much faster than the other (Bayesian) methods (with a scale ranging from 16 to 44)

To conclude, we found that the RR model was powerful and fast. Bayesian methods appeared to be useful only in the situation where a trait is governed by just a few strong effects QTL. These results on empirical maize datasets agree with the results of previously published papers on comparison between methods.

A comparison of models for the prediction of genotypic values in self fertilizing plants

Thorwarth, Patrick[1]; Enders, Matthias[2]; Ordon, Frank[2]; Schmid, Karl[1]

1: Institute of Plant Breeding, Seed Science and Population Genetics, Department of Crop Biodiversity and Breeding Informatics, University of Hohenheim, Stuttgart, Germany
2: Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Resistance and Stress Tolerance, Quedlinburg, Germany

Abstract: The genomic prediction of genotypic values is a valuable tool for animal and plant breeding. To evaluate the prediction ability for the genomic estimation of genotypic values in self-fertilizing plant species, we perform a cross-validation study in the *Arabidopsis thaliana* MAGIC population consisting of 527 individuals genotyped with 1260 SNP-Marker and of a winter barley (*Hordeum vulgare L.*) population consisting of 113 individuals genotyped with 6808 SNP-Marker. We compare the prediction ability of three estimation methods: GBLUP, RRBLUP and Bayesian LASSO. The traits under observation are rosette diameter for *Arabidopsis thaliana* and thousand kernel weight for barley. Additionally we perform an association study and compare significant SNPs detected in the association study to the SNP effects calculated with RRBLUP and Bayesian LASSO. Considering the prediction ability no significant differences for the three estimation models in the two populations, respectively, are detected. For the *A. thaliana* population the prediction ability with GBLUP/ RRBLUP is 0.42 and with Bayesian LASSO 0.43. In the barley population prediction abilities of 0.78 and 0.76 are reached.

References

Meuwissen THE et al. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829

Kover PX et al. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genetics 5, e1000551

Pérez P et al. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian Linear Regression package in R. The Plant Genome 3, 106-116

Rode J et al. (2011). Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare L.*). Molecular Breeding (online)

Wimmer V et al. (2012). Synbreed: A framework for the analysis of genomic prediction data using R. Bioinformatics 19, 889-890

In silico defined breeding strategies applied in practice: Prospects and lessons learned

van Berloo, Ralph[1]; Antonise, Rudie[1]; Buntjer, Jaap[1]

1: Keygene NV, Wageningen, The Netherlands

Abstract: In recent years, in collaboration with partners in plant breeding industry, Keygene has developed statistical and mathematical methodology that is aimed at the genome-wide detection of genomic regions (genes/QTLs) underlying traits of interest, but, in addition, aims to construct the most efficient breeding strategies to *use* the detected regions and accumulate genomic fragments originating from different sources into new (pre) breeding germplasm. We are currently in a phase of implementation of these methods in the breeding practice in collaboration with different breeding companies. The feedback from companies is very helpful to steer further methodology development, recognize possible pitfalls and identify 'white spots' in the methodology relevant to the breeding practice. Next to this, we gained experience in making a case for genomic breeding strategy methods with practical crop breeders. In this presentation we want to briefly introduce the breeding strategy toolkit we developed, discuss the insights we gained from interaction with plant breeding practice and present an outlook for future applications.

Genomic prediction under local adaptation

van Heerwaarden, Joost[1]; Malosetti, Marcos[1]; Boer, Martin P.[1]; van Eeuwijk, Fred A.[1]

1: Biometris, Wageningen University, Wageningen, The Netherlands

Abstract: The increasing availability of high density genotyping data in crop plants has led to the expectation that genotype-phenotype correlations may be extrapolated to unrelated individuals from diverse breeding pools. The reliability of phenotypic predictions in such pools depends on both the genetic architecture of the targeted traits and on the capacity to distinguish linkage to QTL from spurious associations due to cryptic relatedness. Both these aspects are affected by population genetic processes such as selection and drift, but the importance of evolutionary forces in determining the success genomic prediction has not been well studied. The role of selection is of particular interest since local adaptation to different growing environments is probably a key source of genetic diversity in agronomic traits. It has been hypothesized that, depending on genetic architecture, local adaptation may cause trait differentiation in the absence of strong frequency differences at underlying QTL, making it harder to detect genomic associations. We present results of a simulation study on the evolutionary aspects of genomic prediction. We address issues related to the accuracy and stability of genomic prediction under different genetic architectures and levels of admixture.

PedigreeSim: simulation of diploid and tetraploid meioses and pedigrees

Voorrips, Roeland E.[1] ; Maliepaard, Chris A.[1]

1: Wageningen University & Research Center – Department of Plant Breeding, Wageningen, The Netherlands

Abstract: While the genetics of diploid inheritance are well studied and software for linkage mapping, haplotyping and QTL analysis are available, for tetraploids the available tools are limited. In order to develop such tools it would be helpful if simulated populations based on a variety of models of the tetraploid meiosis would be available.

Here we present PedigreeSim, a software package that simulates meiosis in both diploid and tetraploid species and uses this to simulate pedigrees and cross populations. For tetraploids a variety of models can be used, including both bivalent and quadrivalent formation, varying degrees of preferential pairing of hom(oe)ologous chromosomes, different quadrivalent configurations and more. We show that simulation of quadrivalent meiosis results in double reduction and recombination between more than two hom(oe)ologous chromosomes.

This is the first public simulation software that implements all features of meiosis in tetraploids. It allows to generate data for tetraploid and diploid populations, and to investigate different models of tetraploid meiosis. The software and manual are available for free from
http://www.plantbreeding.wur.nl/uk/software/PedigreeSim.html

A menu-based pipeline for trial analysis and QTL detection

Welham, Sue[1]; Murray, Darren[1]; Boer, Martin[2]; Malosetti, Marcos[2]; Jansen, Johannes[2]; Thissen, Jac[2]; Keizer, Paul[2]; Payne, Roger[1]; van Eeuwijk, Fred A.[2]

1: VSN International Ltd, Hemel Hempstead, United Kingdom
2: Biometris, Wageningen UR, Wageningen, The Netherlands

Abstract: A menu-based pipeline for analysis of field trials and detection of QTL and QTL-by-environment (QTL×E) interactions has been constructed within the GenStat® statistical package (VSN International, 2012). It can be applied to the standard biparental populations, full-sib families of outbreeders (CP) and association mapping populations (AMP). The system uses a two-stage approach for analysis. At the first stage, mixed model analysis is applied to individual field trials to obtain adjusted genotype means. At the second stage, QTL detection can be applied to adjusted means for one trait from a single trial, or QTL×E interactions can be investigated using several environments. Alternatively, several traits from a single trial can be analysed together, as a multi-trait analysis. QTL detection is implemented via mixed models using the methods of Malosetti et al. (2004) and Boer et al. (2007), scanning for either single QTL or selecting a model for multiple QTL.

Preliminary analysis of individual field trials is facilitated by menus for spatial analysis which provide diagnostic plots to aid selection of an appropriate variance model for each trait. Once a model has been selected, the trait data and model can be loaded into the QTL system and re-analysed to produce estimates of heritability, adjusted means and weights for use in the second stage analysis Welham et al. (2010). For multiple trials, modelling of genotype-by-environment (G×E) variation is guided by a menu that fits a range of covariance models and selects the best based on information criteria (AIC or BIC). Weights from the first stage can be used to incorporate within-trial precision in the estimation procedure.

Marker data and a genetic map required for QTL detection at the second stage can be loaded in several standard formats. A genetic map can be formed for inbred or CP populations, using the method of Jansen et al. (2001) and Jansen (2005) and various diagnostic tools are available.

The system allows QTL detection for F2, BC, DH, RIL or CP populations. Genetic predictors, contrasts between conditional QTL genotype probabilities given marker information, calculated using Hidden Markov Models (Jiang and Zeng, 1997), are evaluated at marker positions and at user-selected intervals. These genetic predictors are then incorporated in a mixed model which can include across-trial covariance and weights from the first stage. A single menu guides the user through an initial scan for single QTL (or QTL×E), then allows selected QTL to be used as cofactors in subsequent multi-QTL models. Back-selection is used to select a final model with multiple QTL.

QTL detection for AMP allows modelling of linkage disequilibrium and several methods for specification of population sub-structure (specification of groups, kinship matrix or eigen-analysis). Again, a menu guides the user through models for single and multiple QTLs, for either a single or multiple trials.

References

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007). A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177, 1801-1813

Jansen J (2005) Construction of linkage maps in full-sib families of diploid outbreeding species by minimizing the number of recombinations in hidden inheritance vectors. Genetics 170, 2013-2025

Jansen J, de Jong AG, van Ooijen JW (2001). Constructing dense genetic linkage maps. Theor Appl Genet 102, 1113-1122

Jiang C, Zeng ZB (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101, 47-58

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004). Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137, 139-145

VSN International (2012). GenStat for Windows 15th Edition. VSN International Ltd, Hemel Hempstead, UK

Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR (2010). A comparison of analysis methods for late-stage variety evaluation trials. Australian & New Zealand Journal of Statistics 52, 125-149

Potential of genomic selection for mass selection breeding in allogamous crops for traits requiring selection before or after pollination

Yabe, Shiori[1]; Ohsawa, Ryo[2]; Iwata, Hiroyoshi[1]

1: Laboratory of Biometry and Bioinformatics, Graduate School of Agricultural and Life Sciences, The University of T okyo, Tokyo, Japan
2: Laboratory of Plant Breeding, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Abstract: Mass selection is an important method for genetic improvement of allogamous crops. General steps of mass selection include: (1) individual plants are rejected or selected on the basis of their phenotype observed in a field experiment, (2) the offspring of all selected plants are grown in bulk, and these steps are repeated. Mass selection is simple but inefficient in general mainly because of inaccurate single-plant selection. Selection based on whole-genome markers, i.e., genomic selection (GS), might improve the efficiency of mass selection by improving the accuracy of single-plant selection and/or by increasing the number of selection cycles per unit time. In this study, we performed breeding simulations to assess the efficiency of mass selection with GS in annual allogamous crop breeding. To simulate the low levels of linkage disequilibrium (LD) in a breeding population, we assumed a linkage equilibrium in an initial breeding population. A major concern of the present study was to compare the efficiency of GS breeding between two types of target traits: one was a trait expressed before pollination and the other was a trait expressed after pollination. For the former type, we can select the best plants for the next generation before pollination. For the latter type, we cannot select pollen parents because we can observe phenotypes only after pollination and fertilization.

First, we simulated the case in which a target trait requires “selection before pollination (SBP)”. In the simulations, we compared GS breeding with phenotypic selection (PS) and conventional marker-assisted selection (MAS) breeding. In GS breeding, we updated a prediction model in the first cycle of each year with phenotype and marker-genotype data of a breeding population (we called this cycle as genomic and phenotypic selection: GPS) and, following the GPS, we performed up to two cycles of GS per year with the updated prediction model (we called these cycles as GS cycles). Results showed that GS breeding attained higher genetic gain than PS and MAS breeding. GS breeding with a larger population size and a larger number of cycles attained higher genetic gain except when the population size was as small as 50.
Second, we simulated the case in which a target trait requires “selection after pollination (SAP)”. In this case, we could select only seed parents at GPS cycles, but could select both seed and pollen parents at GS cycles. The efficiency of GS breeding in a trait requiring SAP was as high as in a trait requiring SBP, while the efficiency of PS breeding got significantly worse in a trait requiring SAP. Thus the relative efficiency of GS over PS was much higher in a trait requiring SAP than in a trait requiring SBP. By analyzing each simulation process carefully, we found that the genetic values of QTL haplotypes derived from pollen parents were predicted accurately in GS cycles immediately after GPS cycles. Because the pattern of LD in chromosomes derived from pollen parents did not change largely between GPS and GS cycles, chromosomes that were unselected at GPS cycles could be selected accurately in the succeeding GS cycle.

We compared the genetic gain per unit cost of GS over PS at the 6^{th} year of selection for both types of traits. The cost efficiency of GS was higher in a trait requiring SAP than in a trait requiring SBP, and was comparable to PS when genotyping cost was about three times and twice higher than phenotyping cost for traits requiring SBP and SAP, respectively. Thus, GS is anticipated to improve the efficiency of mass selection breeding in allogamous crops, especially when a target trait requires SAP. Selection experiments would be necessary for the next step toward the practical use of GS in allogamous crop breeding.

The challenges of organizing the data of a breeding program

Zinn, David[1]; Pultrini, Pascal[1]; Duminil, Tristan[1]; Bardet, Sébastien[1]; Royer, Florence[2]; Royer, Frédéric[2]

1: Doriane sas, Nice, France
2: Biosearch Data Management, Nice, France

Abstract: In the last two decades many fields have been facing new challenges regarding the rapid evolution of computer science and biotechnologies. Plant breeding, both for fundamental research and the industry, has benefited from the availability of cheap personal computers and private genotyping laboratories, thus entering the age of Biotechnology Assisted Breeding. These new assets imply, however, higher expectations, especially for data management.

Research foundations and major seed companies have invested in their own IT departments or have subcontracted developers to tailor software that fits their specific needs. This, however, has left smaller teams or companies without dedicated software, forcing them to rely on office suites and analyser software to manage their data and work flow.

LABKEY™, data management software resulting of nearly 30 years of effort in relationship with a large number of breeders of very various crops, aims to be both flexible and oriented towards plant breeding. This has been possible mainly for two reasons: The first reason is that regardless of all the specificities of the different crops, plant breeders perform many similar tasks: line crossing, gene marking, trait observing, progeny selecting. The second reason is that breeders are willing to invest in tools which adapt to their ever evolving needs regarding: material coding, activity reporting, data mining.

We would like to share the intellectual process performed by our teams, facing the challenges of creating such a tool and giving examples on how it can match different breeder needs. We will detail the issues, describe solutions and evaluate their qualities and limitations.

The first technical issue is building a structure able to store and access a large quantity of heterogeneous data. This can be addressed by many solutions including flat files, classical relational databases or object oriented relational databases. Another technical issue is managing the structure itself as the stored data may evolve and change. This can be addressed by several solutions including untyped storage and dynamic structures. Last but not least: configuring the output to make the data available to the user in a useful way, using solutions like object oriented interfaces, configuration wizards and script writing.

Concerning breeder needs, we highlight the management of starting crosses. The issue being that one needs to be able to explore the data of all the potential parents (traits, genes…) and plan ahead, perform, and track the crosses in real time. Secondly, we detail the offspring generational selection process. One needs to take into account observations (yield, visual traits…) and laboratory results (resistance genes, chemical composition…) to eliminate the offspring which does not carry the desired parent traits. Finally, we underline the specificities of the multi-location trials of the selected crops. Designed to evaluate, in a quantitative way, the inherited qualities of a crop (adaptation to environmental stresses, consistency of agronomic qualities), the trials must comply with statistical constraints to have significant results.

Through cooperation with our industrial and academic partners, we search for new and better ways to address these agronomic data managing issues. We do this keeping in mind a challenging principle: designing software as adaptable as both the living material it tracks and the innovating minds of the researchers we serve.

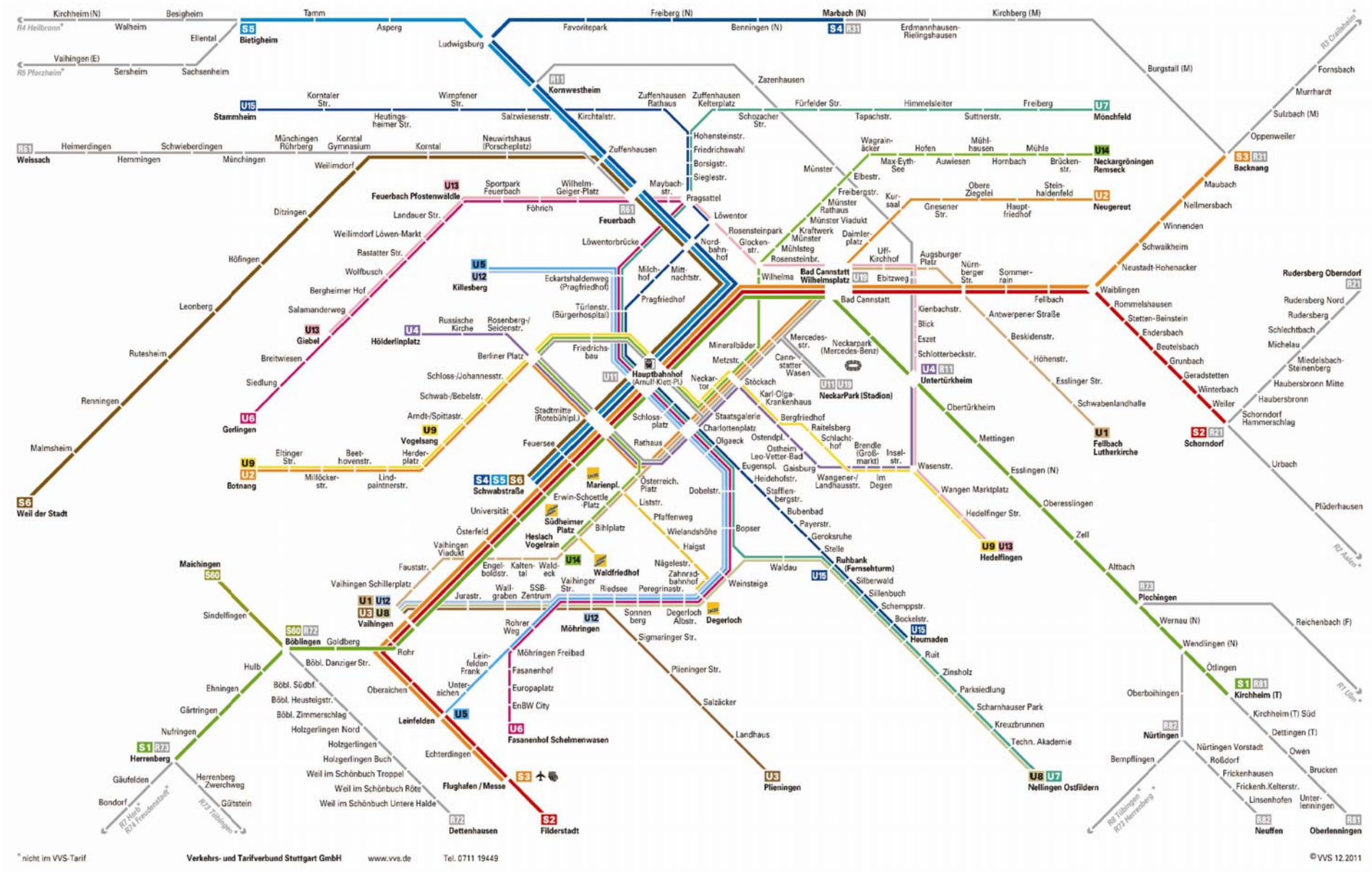
References

Batnini H, Sillon JF, Bardet S, Royer Fl, Royer Fr (2009). LABKEY: Using artificial intelligence for marker assisted breeding. EUCARPIA XIV Meeting of the Biometrics in Plant Breeding Section, Dundee 2-4 September: p.31

Sillon JF, Royer Fl, Royer Fr (2006). Management of research department data. EUCARPIA XIII Meeting of the Biometrics in Plant Breeding Section, Zagreb: 30 August - 1 September: p.62

Part IV

List of contributors


Kirchheim (N)
Besigheim
R4 Heilbronn*
Walheim
Ellental
S5
Bietigheim
Vaihingen (E)
R5 Pforzheim*
Sersheim
Sachsenheim
Tamm
Asperg
Ludwigsburg
Freiberg (N)
Favoritepark
Benningen (N)
Marbach (N)
S4 R81
Kirchberg (M)
Erdmannhausen-Rielingshausen
Burgstall (M)
R3 Crailsheim*
Fornsbach
Murrhardt
Sulzbach (M)
Oppenweiler
S3 R81
Backnang
Maubach
Neilmersbach
Winnenden
Schwaikheim
Neustadt-Hohenacker
Waiblingen
Rommelshausen
Stetten-Beinstein
Endersbach
Beutelsbach
Grunbach
Geradstetten
Winterbach
Weiler
S2 R81
Schorndorf
Rudersberg Oberndorf
R21
Rudersberg Nord
Rudersberg
Schlechtbach
Michelau
Miedelsbach-Steinenberg
Haubersbronn Mitte
Haubersbronn
Schorndorf Hammerschlag
Urbach
Plüderhausen
R3 Aalen*
R11
Kornwestheim
Zazenhausen
U15
Stammheim
Kornaler Str.
Wimpfener Str.
Heutingsheimer Str.
Salzwiesenstr.
Kirchtalstr.
Zuffenhausen Rathaus
Zuffenhausen Kelterplatz
Schozacher Str.
Fürfelder Str.
Himmelsleiter
Freiberg
Tapachstr.
Suttnerstr.
U7
Mönchfeld
R51
Weissach
Heimerdingen
Hemmingen
Schwieberdingen
Münchingen
Münchingen Rührberg
Korntal Gymnasium
Korntal
Neuwirtshaus (Porscheplatz)
Zuffenhausen
Hohensteinstr.
Friedrichswahl
Borsigstr.
Sieglestr.
Pragsattel
Löwentor
Rosensteinpark
Nordbahnhof
Löwentorbrücke
Glockenstr.
Münster
Elbestr.
Freibergstr.
Münster Rathaus
Münster Viadukt
Kraftwerk Münster
Mühlsteg
Rosensteinbr.
Daimlerplatz
Wagrainäcker
Max-Eyth-See
Hofen
Auwiesen
Mühlhausen
Hornbach
Mühle
Brückenstr.
U14
Neckargröningen Remseck
Kursaal
Obere Ziegelei
Steinhaidenfeld
U2
Neugereut
Gnesener Str.
Hauptfriedhof
Uff-Kirchhof
Augsburger Platz
Nürnberger Str.
Sommerrain
Ebitzweg
Wilhelma
Bad Cannstatt Wilhelmsplatz
U10
Bad Cannstatt
Felibach
Kienbachstr.
Blick
Eszet
Schlotterbeckstr.
Antwerpener Straße
Beskidenstr.
Höhenstr.
Esslinger Str.
Schwabenlandhalle
U1
Fellbach Lutherkirche
Weilimdorf
Ditzingen
Höfingen
Leonberg
Rutesheim
Renningen
Malmsheim
S6
Weil der Stadt
U13
Feuerbach Pfostenwäldle
Landauer Str.
Weilimdorf Löwen-Markt
Rastatter Str.
Wolfbusch
Bergheimer Hof
Salamanderweg
U13
Giebel
U6
Gerlingen
Sportpark Feuerbach
Föhrich
R51
Feuerbach
Wilhelm-Geiger-Platz
Maybachstr.
U5
U12
Killesberg
Eckartshaldenweg (Pragfriedhof)
Türlenstr. (Bürgerhospital)
Milchhof
Mittnachtstr.
Pragfriedhof
Russische Kirche
Rosenberg-/Seidenstr.
U4
Hölderlinplatz
Berliner Platz
Friedrichsbau
Hauptbahnhof (Arnulf-Klett-Pl.)
U11
Mineralbäder
Metzstr.
Mercedesstr.
Neckarpark (Mercedes-Benz)
Cannstatter Wasen
U11 U19
NeckarPark (Stadion)
U4 R11
Untertürkheim
Obertürkheim
Mettingen
Esslingen (N)
Oberesslingen
Zell
Altbach
R73
Plochingen
Wernau (N)
Reichenbach (F)
R1 Ulm*
Wendlingen (N)
Ötlingen
S1 R81
Kirchheim (T)
Kirchheim (T) Süd
Dettingen (T)
Owen
Brucken
Unterlenningen
R81
Oberlenningen
Oberboihingen
R82
Nürtingen
Nürtingen Vorstadt
Roßdorf
Frickenhausen
Frickenh.Kelterstr.
Linsenhofen
R82
Neuffen
Bempflingen
Rd Tübingen*
R73 Herrenberg*
Schloss-/Johannesstr.
Schwab-/Bebelstr.
Arndt-/Spittastr.
U9
Vogelsang
U9
U2
Botnang
Eltinger Str.
Beethovenstr.
Herderplatz
Millöckerstr.
Lindpaintnerstr.
Siedlung
Breitwiesen
Schloss-platz
Stadtmitte (Rotebühlpl.)
Rathaus
Feuersee
Marienpl.
Österreich. Platz
Liststr.
Pfaffenweg
Wielandshöhe
Haigst
Nägelestr.
Zahnradbahnhof
Weinsteige
Dobelstr.
Bopser
Waldau
U15
Ruhbank (Fernsehturm)
Silberwald
Sillenbuch
Schemppstr.
Bockelstr.
U15
Heumaden
Ruit
Zinsholz
Parksiedlung
Scharnhauser Park
Kreuzbrunnen
Techn. Akademie
U8 U7
Nellingen Ostfildern
Keckarstr.
Karl-Olga-Krankenhaus
Staatsgalerie
Bergfriedhof
Charlottenplatz
Raitelsberg
Olgaeck
Ostendpl.
Schlachthof
Ostheim
Leo-Vetter-Bad
Brendle (Großmarkt)
Inselstr.
Eugenspl.
Gaisburg
Heidehofstr.
Wangener-/Landhausstr.
Im Degen
Wasenstr.
Wangen Marktplatz
Hedelfinger Str.
U9 U13
Hedelfingen
Stafflenbergstr.
Bubenbad
Payerstr.
Geroksruhe
Stelle
Schwabstraße
S4 S5 S6
Erwin-Schoettle-Platz
Universität
Südheimer Platz
Heslach Vogelrain
Bihlplatz
Österfeld
Vaihingen Viadukt
Faustsr.
Vaihingen Schillerplatz
U1 U2
U3 U8
Vaihingen
Engelboldstr.
Kaltental
Waldeck
U14
Waldfriedhof
Vaihinger Str.
Riedsee
Peregrinastr.
Wallgraben
SSB-Zentrum
Jurastr.
Degerloch Albstr.
Degerloch
Sonnenberg
U12
Möhringen
Rohrer Weg
Möhringen Freibad
Fasanenhof
Europaplatz
EnBW City
U6
Fasanenhof Schelmenwasen
Sigmaringer Str.
Plieninger Str.
Salzäcker
Landhaus
U3
Plieningen
Maichingen
S60
Sindelfingen
S60 R72
Böblingen
Goldberg
Rohr
Hulb
Ehningen
Gärtringen
Nufringen
S1 R73
Herrenberg
Gäufelden
Bondorf
Herrenberg Zwerchweg
Gültstein
R7 Horb*
R14 Freudenstadt*
R73 Tübingen*
Böbl. Danziger Str.
Böbl. Südbf.
Böbl. Heusteigstr.
Böbl. Zimmerschlag
Holzgerlingen Nord
Holzgerlingen
Holzgerlingen Buch
Weil im Schönbuch Troppel
Weil im Schönbuch Röte
Weil im Schönbuch Untere Halde
Oberaichen
Unteraichen
Leinfelden
U5
Leinfelden Frank
Echterdingen
S3
Flughafen / Messe
R72
Dettenhausen
S2
Filderstadt
* nicht im VVS-Tarif
Verkehrs- und Tarifverbund Stuttgart GmbH
www.vvs.de
Tel. 0711 19449
© VVS 12.2011