III International Meeting on Plant Breeding

New Approaches on Plant Breeding: Insights into Artificial Intelligence

PROCEEDINGS

October – 2019

THE INTERNATIONAL MEETING ON PLANT BREEDING

As the world's population grows, different strategies must be taken by society to sustainably increase food and supplies production. Plant breeding can be considered one of the fundamental strategies for the development of more adapted and productive cultivars. New tools that aid genetic improvement are released every day, aiming to increase selection accuracy, decrease breeding cycle time and increase genetic gain and, consequently, productivity. Thus, discussions around these new tools are important to ensure it correct use and application within research and breeding projects.

The International Meeting on Plant Breeding is one of several events in the "Corteva Agriscience Plant Science Symposia Series". In this edition the topic addressed were artificial intelligence, which has gained the attention of researchers in recent years for allowing wide application in virtually all stages of plant breeding. It is possible to apply artificial intelligence to genotypic and phenotypic data, genomic prediction and mainly image processing. Considering this, the objective of the event was to promote the discussion of the application of artificial intelligence within the genetic improvement program.

GVENCK

The Genetics and Plant Breeding Group "Prof. Roland Vencovsky" (GVENCK) is composed by graduate and undergraduate students in Genetics and Plant Breeding at "Luiz de Queiroz" College of Agriculture (ESALQ/USP), under coordination of Professor Dr. José Baldin Pinheiro. Our mission is to integrate academics, professors and professionals with the goal of improving the training of future breeders and geneticists.

The main activities of the group are:
• Organization of scientific and training events;
• Promotion of discussion on relevant topics in genetic and plant breeding;
• Technical visits to companies and public research institutions;
• Promote the guidance of young talents under training from the "alumni voice", in which the alumni with consolidated careers will share professional experiences;
• Promote moments and opportunities for interaction between students, professors and researchers outside the university;
• Establishment of partnerships with companies and public institutions.

ORGANIZATION

Coordination
José Baldin Pinheiro
Rafael Massahiro Yassue
Ana Letycia Basso Garcia

Secretary
Alline Sekiya
Emanoel Sanches Martins
Jessica Aparecida Ferrarezi
Mariana Niederheitmann
Melina Prado
Pedro Augusto Medeiros Barbosa

Treasurer
Maiara de Oliveira
Patrícia Braga

Scientific
Albania José Patiño Torres
Jéssica Eliana Nogueira de Souza
Renata Flávia de Carvalho
Roberta Luiza Vidal
Talieisse Gomes Fagundes

Communications
Mariana Vianna
Bruna de Moura Lopes
Karina Midori Kobayashi Omosako
Marcos Antonio de Godoy Filho
Melissa Cristina de Carvalho Miranda
Yajahaira Nevenka Carbajal Gonzales

Logistics
Guilherme Kenichi Hosaka
Elesandro Bonhofen
Leonardo Fioravante Gotardi
Willian Giordani

ACKNOWLEDGEMENTS

PROGRAM:

October 1st, 2019 - Tuesday

07:30– 8:00 – Registration

08:00 – 08:30 – Opening Session – Tabare Abadie - Corteva Agriscience - USA

08:30 – 09:30 – Lecture 1: André Carlos Ponce de Leon Ferreira de Carvalho – "What is the relation between Big Data and Data Science? And how does Artificial Intelligence fits in this relation?"

09:30 - 09:50 – Coffee Break

9:50 – 10:50 – Lecture 2: Alencar Xavier – "Good learners, faster learning"

10:50 – 11:30 – Poster Session

11:30 – 13:30 – Break

13:30 – 14:30 – Lecture 3: Osval A. M. López – "Machine learning and deep learning methods for genomic prediction"

14:30 – 15:30 – Lecture 4: Katy Martin Rainey – "Phenomic Inference of Soybean Growth and Development"

15:30 - 16:00 - Coffee Break

16:00 – 17:00 – Lecture 5: Rodrigo G. Trevisan – "Deep learning for image-based high throughput plant phenotyping"

17:00 – 18:00 – Lecture 6: Vinícius Silva Junqueira – "How Data Science is changing the on the job experience in industry?"

18:00 – 18:30 – "Roland Vencovsky" Award and Closing Remarks

19:00 – Official Event Dinner

October 2nd, 2019 - Wednesday: Minicourse I

8:00 – 18:00 - Osval A. M. López - "Machine learning methods for genomic prediction"

October 3rd, 2019 - Thursday: Minicourse II

8:00 – 18:00 - Rodrigo G. Trevisan - "A hands-on introduction to deep learning for image-based high throughput plant phenotyping"

SPEAKERS

Dr. Tabare Abadie has a Ph.D. in Plant Breeding and Quantitative Genetics from the University of Minnesota. His areas of interest include Plant Breeding, Higher Education, Career Development and Mentoring. It is currently a Corteva Fellow and his main responsibility is to lead the Plant Sciences Symposia Series. He is also currently serving as a member of the Board of Trustees of the Integrated Breeding Platform, and of the Agronomic Science Foundation (ASF), and as a member of external scientific advisory panels for USDA and NSF funded projects at academic organizations.

Dr. André Carlos Ponce de Leon Ferreira de Carvalho is a Professor at the Institute of Mathematical and Computer Sciences, University of São Paulo (ICMC-USP), São Carlos campus. Vice President of the Brazilian Computer Society (SBC). He holds a Bachelor degree (1987) and a Master's degree in Computer Science (1990) from the Federal University of Pernambuco, and a doctorate in Electronic Engineering from the University of Kent (1994). His main research interests are Machine Learning, Data Mining, and Data Science, with applications in many areas.
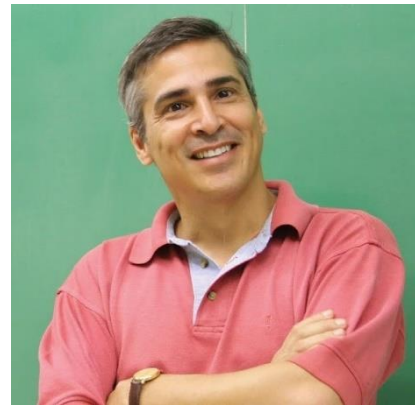
Dr. Alencar Xavier has his Ph.D. degree in soybean breeding and genetics statistics at Purdue University. His works on statistical genetics are focused on genomic-assisted breeding with emphasis on theoretical and computational aspects of data-driven plant breeding, such as modeling, prediction, and selection using various sources of information. His researches regard the development and implementation of new quantitative methods using mixed models, Bayesian methods and machine learning, along with high-performance computing. Currently, hi is a quantitative scientist at Corteva Agriscience.
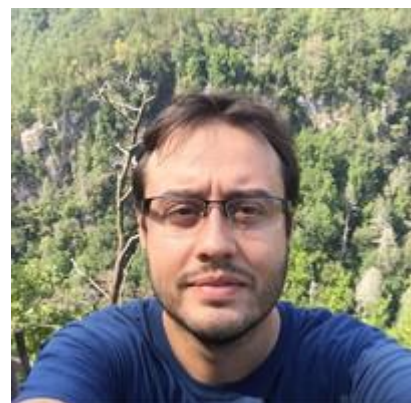
Osval A Montesinos-López has PhD. in statistics and biometry and is a professor at the Faculty of Telematic, Universidad de Colima in Mexico. Dr. Osval researches are currently in Biostatistics and Bioinformatics.

Dr. Katy Martin Rainey is an Associate Professor of Plant Breeding and Genetics in the Purdue University Department of Agronomy. She studies the genetics improvement of soybeans for increased yield and better quality using multidisciplinary approaches.
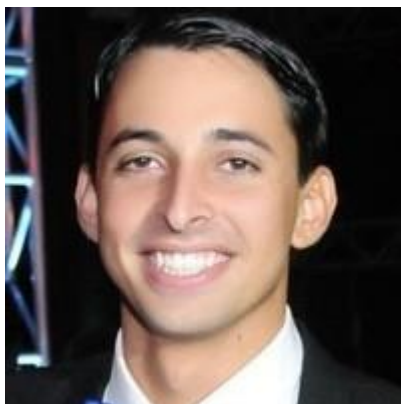
Rodrigo Trevisan is a graduate research assistant in Crop Science at the University of Illinois at Urbana-Champaign - USA. He is currently responsible for research and development of technology solutions for agribusiness at Smart Agri. Has experience in agricultural planning, precision agriculture, farm level experimentation, remote sensing, geographic information systems, data analysis, and artificial intelligence.

Dr. Vinícius Junqueira holds a degree in veterinary medicine from the University of Brasilia (2010), a master's degree in Genetics and Breeding at the Federal University of Viçosa (2014) and a doctorate in Genetics and Breeding at the Federal University of Viçosa (2018). He was a visiting researcher at the University of Georgia (2016). He is currently Data Science Supervisor - Bayer Corporation.

<h1 style="text-align: center">SUMMARY</h1>

COMPARISON OF TWO APPROACHES FOR CONTIG ORDERING IN GENOME ASSEMBLIES

Marcos Antonio de Godoy Filho[1]; Lucas Borges dos Santos[2]; João Paulo Gomes Viana[3]; Wei Wei[3]; Steven J. Clough[3,4]; and José Baldin Pinheiro[1]

[1]Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP. [2]Universidade Estadual de Campinas. [3]University of Illinois at Urbana-Champaign, [4]US Dept of Agriculture, Urbana, IL. E-mail: marcos.a.godoy.f@usp.br.

Soybean is the main commodity in Brazil, in 2018, soybean and its derivatives accounted for $40.9 billion in exports. However, there are some obstacles to obtaining higher yield, and among the difficulties, the attack of insects stands out, especially by the complex of stink bugs. An alternative to the use of insecticides is the development of resistant cultivars. The assembly and comparison of contrasting genomes can be a good approach to identify polymorphisms between them, and allows integration with differentially expressed gene data and QTLs for a better understanding of the genetic architecture of the trait. However, genome assembly and ordering are challenges considering the high complexity of soybean genome. For a better understanding of the genetic architecture involved in the resistance of soybean to the stink bug complex, genomes of soybean cultivar IAC-100 (tolerant) and cultivar CD-215 (susceptible) were sequenced. For this, we used the Chromium Genome technology, and sequencing was performed on the Illumina NovaSeq 6000 plataform. The draft genome assembly was performed using the software Supernova 2.1.1, which for the IAC-100 generated an assembly with 1.051MB with a scaffold N50 of 4.02 Mb. For the cultivar CD-215, an assembly of 1.055MB was generated, with an N50 scaffold of 4.28 MB. Two approaches for ordering and breaking chimeric contigs were tested, using Ragoo based on reference genome of Willians 82 and Cromonomer based on consensus genetic map of soybean. Contig ordering and breaking proved to be highly efficient in increasing the quality of genome assembly, with Ragoo improving the N50 scafolds up to 53.5MB in IAC100 and 53.8MB for CD215, plus break 44 chimeric contigs in assembly of IAC100 and 49 in CD215. This result for N50 scaffold is better than all soybean genomes available in NCBI, in addition CG content is close to the other assemblies available (~31%) and 75% of assembly of both genomes are only 15 scaffolds. The Cromonomer boosted N50 scafolds to around 35-37MB but found fewer chimeric contigs, however, it should be noted that just over 2700 SNPs lined up against assemblies and was used for ordering using the genetic map. The comparative genomics through genome assembly in soybeans opens the door to several other analyzes and studies in order to understand the genetic architecture of bed bug complex tolerance. The combination of supernova and Raggo softwares can be a powerful tool for assembling genomes that already have a reference, with Chromonomer software being a great alternative for species without assembled genomes but with good genetic maps.
Keywords: Assembly; Soybean; Genome; Ordering; Stinck Bugs.

FACTOR ANALYSIS APPLIED TO MAIZE ADAPTABILITY STUDIES AND ENVIRONMENTAL STRATIFICATION

Maria Angélica Marçola[1]; Yasmin Moura Araújo[2]; Deoclécio Domingos Garbuglio[3]; Gabriela Inocente[4]; Pedro Mário de Araújo[5]

[1]Instituto Agronômico do Paraná (IAPAR); [2] Agronômico do Paraná (IAPAR); [3]Instituto Agronômico do Paraná (IAPAR); [4]Universidade Estadual de Londrina (UEL); [5]Instituto Agronômico do Paraná (IAPAR). E-mail: maria_marcola@hotmail.com.

For adaptability and stability studies, maize cultivars are evaluated in different regions with edaphoclimatic diferences in order to explore the effects of genotype x environment interaction (GxE). In this context, among univariate methods, linear regression has been widely used, but is not informative if linearity fails and tends to simplify response models. Among the multivariate techniques, factor analysis show promising results, being used in plant breeding relatively recently, both in environmental stratification and genotype adaptability studies. Thus, the objectives of this work were to evaluate the adaptability and stability of genetically modified maize cultivars and to perform environmental stratification in the IAPAR maize evaluation network, using the factor analysis technique. The trials were carried out 2018/2019 crop, in the location: Londrina, Santa Tereza do Oeste, Santa Helena, Palotina, Cambará, Foresta, Campo Mourão, Pato Branco, Guarapuava and Ponta Grossa, evaluating 13 maize cultivars. The experimental design was randomized complete blocks with three replications per evironment, and the plots consisted of two rows of 5 meters length, 80 cm between rows and 20 cm between plants. Through the yeld means, were identified the cultivars FS 481 PW and 2B500 PW, which showed wide broad adaptability to different environments clutered into 4 final factors. Cultivars R9789-VIP3 and R9080-PRO2 presented low adaptability in factor sets F1xF2, F1xF3 and F1xF4, which clusted 53.8%, 38.4% and 38.4% of the environments, within each combination of factors, respectively. The other cultivars presented specific adaptability to certain factor combinations. The factor analysis technique is efficient for the identification of genotypes with different responses to GxE interaction, as well as for grouping of environments.
Keywords: *Zea mays L.;* GxE interaction; Multivariate analysis; Heatmaps.

GENOMIC X ENVIROTYPING KERNELS DRIVE TO A BETTER PREDICTION AND UNDERSTANDING OF MAIZE YIELD PLASTICITY

Germano Martins F. Costa-Neto[1]; Giovanni Galli[1]; Roberto Fritsche-Neto[1]

[1]Luiz de Queiroz College of Agriculture, University of São Paulo (ESALQ/USP), E-mail: germano.cneto@usp.br.

Whole-genome prediction (WGP) is guided from several genotype-to-phenotype (G-to-P) modeling strategies integrated into a time-saving, cost-effective breeding pipeline. However, over multi-environment analysis, the genotype by environment interaction (GE) and their unpredictable impact on phenotypic plasticity may limit the range of applications of WGP-based breeding. One of the strategies used to solve this gap relies on the integrative envirotyping-based data with WGP modeling, which are focused on emulating the latent genotypic-specific reaction norms across the mean environmental gradient. Here we integrate explicit envirotyping data with WGP by emulating the implicit envirotype-to-phenotype (E-to-P) dynamics driving GE effects and yield plasticity over a target population of environments (TPE) in maize. From available phenotypic data of an optimized training population set (N = 50% of genotypes), we obtained the empirical environmental relationship kernel ($*K_E$), which is trait-specific for grain yield. Then, partial least squares regression with envirotyping data (e.g., daily weather, nitrogen level) were used to translate the trait-specific E-to-P dynamics present in $*K_E$ into a realized environmental kernel K_E. One advantage is that we can fit an approximation of the infinitesimal environment effect and compute its impact on the overall environmental quality driving K_E for a specific TPE. Finally, genomic best-unbiased predictions (GBLUP) were carried out, incorporating additive and dominant effects with the K_E-based reaction norm. As a proof-of-concept, we computed the predictive ability (r), selection coincidence (SC) and root mean squared error prediction (RMSE) front to the benchmark additive-dominant single-variance GE deviation GBLUP, over 5-folds of leave-one-environment and 3 prediction scenarios: (i) new GE (nGE); (ii) new genotypes (nG) and (iii) new environments (nE). We observed lower r gains in the scenarios nGE (+5%) and nE (+3%), but an increased by 10% for nG, indicating that the modeling focused on target trait and TPE may improve prediction for new genotypes within the multi-environment testing network. However, SC% values were increased for all scenarios, from 10% (nGE) to 15% (nG), and RMSE values reduced by up 391%. In addition, we identified the orthogonal impact of the environmental sources onto GE variation, such as solar nitrogen level (20%), radiation across crop development (50%), wind speed (35%) and air temperature at initial growing stages (30%). Our results indicate that genomic $\times$ envirotyping based kernels better emulates the genotypic ranks for target environments. Therefore, the quantitative and qualitative differences between the genotypic-specific plastic responses may be better modeled and exploited for the whole TPE. Finally, the genomic $\times$ envirotyping modeling expands the WGP spectrum of applications, such as for yield adaptability screening or for design future breeding strategies.

Keywords: genomic selection; reaction norm; multi-environment; hybrid testing.

IMPACTS OF PAST AND FUTURE CLIMATE CHANGE ON GENETIC DIVERSITY OF TWO *Eremanthus* SPECIES

Lucas Fernandes Rocha[1,2]; Isaias Emilio Paulino do Carmo[1,3]; Joema Souza Rodrigues Póvoa[1]; Dulcinéia de Carvalho[1]; Evandro Vagner Tamburissi[2,4]

[1]Departamento de Ciências Florestais, Universidade Federal de Lavras, Brazil. [2]Departamento de Ciência Florestal, Universidade Estadual Paulista "Júlio de Mesquita Filho", Brazil. [3]Institute of Biosystems Engineering, Poznan University of Life Sciences, Poland. [4]Departamento de Engenharia Florestal, Universidade Estadual do Centro-Oeste, Brazil. E-mail: lucasrochamg@gmail.com.

Phylogeographic patterns of endemic species is a critical key to understand its adaptation to future climate change. Here, based on chloroplast DNA (cpDNA), we analyzed the genetic diversity of two endemic and endangered tree species from the Brazilian savanna and Atlantic forest (*Eremanthus erythropappus* and *Eremanthus incanus*). Additionally, we applied the climate-based ecological niche modeling (ENM) to evaluate the impact of the Quaternary climate (Last Glacial Maximum ~ 21 kyr BP and Mid-Holocene ~ 6 kyr BP) on the current haplotype distribution. Moreover, we modeled the potential effect of future climate change on the species distribution in 2070 for the most optimistic and pessimistic scenarios. One primer/enzyme combination (SFM/HinfI) revealed polymorphism with very low haplotype diversity, showing only three different haplotypes. The haplotype 1 has very low frequency and it was classified as the oldest, diverging from six mutations from the haplotypes 2 and 3. The *E. erythropappus* populations are structured and differ genetically according to the areas of occurrence. In general, the populations located in the north region are genetically different from those located in the center-south. No genetic structuring was observed for *E. incanus*. The ENM revealed a large distribution during the past and a severe decrease in geographic distribution of *E. erythropappus* and *E. incanus* from the LGM until present and predicts a drastic decline in suitable areas in the future. This reduction may homogenize the genetic diversity and compromise a relevant role of these species on infiltration of groundwater.

Keywords: ecological niche modeling; genetic diversity; climate change; chloroplast DNA.

KINSHIP COEFFICIENTS IN FOREST SPECIES

Antônio Higo Moreira de Sousa[1]; Lucas Fernandes Rocha[1]; Evandro Vagner Tambarussi[1,2]

[1]Departamento de Ciência Florestal, Universidade Estadual Paulista "Júlio de Mesquita Filho", Brazil. [2]Departamento de Engenharia Florestal, Universidade Estadual do Centro-Oeste, Brazil. E-mail: anthigo@gmail.com.

Understanding kinship between individuals in natural populations offers useful information that can be assessed based on estimators of molecular data. However, each estimation method is based on assumptions and is often restricted for forest species. Thus, modeling the data is an important step that is almost always neglected. This study aims to evaluate the efficacy of different kinship estimators for *Acrocomia aculeata* (Jacq.) Lodd, ex Mart., *Hymenaea stigonocarpa* Mart. Ex Hayne, and *Dipteryx alata* Vogel., and four different simulated populations. From the estimated mean coancestry ($\bar{\theta}$), the error estimates were calculated assuming that analyzed individuals were perfect half-siblings ($\bar{\theta}=0,125$). Estimates of kinship ranged according to the species and method used, generating different error values for each estimator. A correlation was observed only between estimators that used the same estimation method or similar assumptions. Simulated populations showed more accurate estimates and lower error values compared to actual data. The error values of the estimators demonstrate that the application of estimators to infer a certain degree of kinship can generate biased results and lead to inefficient decision making. Thus, the use of complementary information associated with kinship is necessary, such as analysis of the reproduction system and genetic structure of the population, enabling more precise inferences of the kinship between evaluated individuals.
Keywords: kinship estimators; forest species; genomic selection; simulations; mixed reproductive system.

MACHINE LEARNING-ASSISTED IMAGE ANALYSIS TO QUANTIFY FOLIAR DISEASE SEVERITY FROM TOMATO BACTERIAL SPOT

Mariana Niederheitmann[1]; Jéssica E. Nogueira de Souza[1]; Daniel P. Longatto[1]; Maria Carolina Quecine Verdi[1]; Fernando A. Piotto[2]; David M. Francis[3]

[1]'Luiz de Queiroz' College of Agriculture, Genetics Department, University of São Paulo, Brazil. [2]'Luiz de Queiroz' College of Agriculture, Crop Science Department, University of São Paulo, Brazil. [3]Ohio Agricultural Research and Development Center, Horticulture and Crop Science Department, The Ohio State University, Wooster, USA. E-mail: mariana_nieder@usp.br.

Foliar disease severity in plants is often assessed using rating or diagrammatic scales. Scaled evaluations may not fully capture genetic signals and the nonparametric features may limit their effective use in genome-wide association studies and fine mapping approaches, which assume a continuous distribution. Image analysis, utilizing segmentation based on machine learning algorithms, provides a straightforward tool for obtaining quantitative information of disease severity. Foliar symptoms of bacterial spot (BS), a major disease in many tomato-growing regions worldwide, include lesions at the margin of leaves and water-soaked, brown, angular spots. These patterned symptoms contrast with leaf and background color, which can be recognized by pixel classification algorithms. In this study, various methods used for image segmentation were tested to assess BS symptoms in 466 leaf images from 20 tomato genotypes displaying variable disease severities and previously rated with a commonly used scale. The leaves were classified and measured for disease severity using the Trainable Weka Segmentation and Color Pixel Counter plugins, respectively, both available in the free and open environment software ImageJ/Fiji. Methods were assessed based on computational time, correlations with rating data, and ability to partition variance into genetics (heritability). Results revealed high heritability (up to 35,0%) and significant correlation of some of the tested methods with disease severity observed in genotypes with known resistance. The most predictive methods were those in which the training image was DSCN9921, which was likely more representative for classifying the other images. The number of selections for background, healthy tissue and diseased tissue were not significantly associated with improved heritability or correlation values, though for some analysis 16-18 appear optimal. Adding the number of training algorithms beyond 3 did not improve prediction and indicated to affect running time. These results point out the potential of such automated tools for determination of disease severity in genetics and crop improvement studies.
Keywords: pixel classification; *Xanthomonas perforans*; plant breeding; *Solanum lycopersicum*.

MULTI-LAYER PERCEPTRON (MLP) NETWORKS AND DIALLEL ANALYSIS APPLIED TO MAIZE POPULATION SELECTION

Gabriela Inocente[1]; Deoclécio Domingos Garbuglio[2]; Pedro Mario de Araújo[3]; Maria Angélica Marçola[4]

[1]Universidade Estadual de Londrina (UEL); [2]Agronomic Institute of Paraná (IAPAR); [3]Agronomic Institute of Paraná (IAPAR); [4]Agronomic Institute of Paraná (IAPAR) E-mail: gab_inocente@hotmail.com.

In the maize breeding programs the search for commercial hybrids may initially be concentrated on the identification of base populations for inbred lines extraction. These populations are direct to crosses in which the best interpopulation hybrids can be used as an indication to justify the extraction of lines with good allelic complementarity. Thus, the present work aimed to evaluate a complete diallel, according to the methodology diallel analysis proposed by Eberhart and Gardner (1966) with adaptation of Morais et al. (1991), between 12 maize populations and predict, through a multi-layer perceptron (MLP) neural network model, selection patterns based on mean components. Twelve intercrossed maize populations were evaluated at the IAPAR (Agronomic Institute of Paraná) experimental stations in three locations: Londrina, Guarapuava and Ponta Grossa, Paraná, during the 2005/2006 crop. The experimental design was randomized complete blocks with two replications per location. The experimental plot was composed of a row with five meters long and five plants per meter after thinning. To identify the population base, only the variable grain yield (GY) was evaluated. It was found that the implemented MLP architecture (one input layer, three hidden layers and one output layer) and other associated parameters, allowed the development of a generalist model of genotype classification (Classes A, B, C and D), based on diallel model (General Combining Ability and Specific Heterosis). Thus the MLP neural network model was efficient in predicting selection classifications from mean components obtained from a complete diallel, regardless of the evaluation environment.
Keywords: *Zea mays* L., Artificial Neural Networks, Intermediate Hybrids, Diallel Crosses.

SOYBEAN LINEAGES TOLERANT TO ASIAN SOYBEAN RUST

Moreno, B.M[1]; Martins, G.M.[1]; Stella, A.A.[1]; Nekatschalow; M.C.[1]; Vello, N.A.[1]; Pinheiro, J.B[1].

[1] Genetics Department, University de Sao Paulo - "Luiz de Queiroz" College of Agriculture USP/ESALQ. E-mail: bruna.marques.moreno@usp.br.

Asian soybean rust is one of the most important diseases that affect soybean production around the world. The season 18/19, FRAC identified a sensibility reduction in sporus collected and after exposed to compounds such as triazole and triazolinthione. As soybean occupies around 36 million hectares, and the pathogen Phakopsora pachyrhizi shows high rates of mutation; using tolerant cultivars can be promising disease management. The main objective was to evaluate and identify in 256 experimental soybean lineages in $F_{5:8}$, tolerance to soybean rust using two treatments (one of them controlling the disease complex of the end of the cycle and the soybean rust, and the other one, controlling the end-cycle diseases and without the control) using the rust effect approach. The experiment was conducted in Piracicaba/SP, in the summer season of 18/19, using Federer blocks as experimental design. Each treatment had two repetitions, to evaluate 256 experimental lineages and 3 controls (IAC100, BRS133 and CB 07-958-B, commercial cultivars). Each block had 1 m^2. The plants were evaluated in plant height, the number of days to maturity, yield, and weight of 100 seeds. In addition, the rust effect was estimated based on the difference between the treatment with the rust control and the treatment without the rust control.The mean yield of lineages with the rust control was 3.827,47 kg.ha[-1], and without the control was 3.255,61 kg.ha[-1]. The weight of 100 seeds was 17,26 g and 13,82 g respectively. The plant height, 97 cm, and 93 cm. And, the plant cycle, 136 days and 127 days. The mean rust effect calculated for yield was 0,571 kg.ha[-1], for the weight of 100 seeds was 3,44 g, height 4 cm and cycle 8,7 days. In conclusion, the plant evaluation has shown that 22 lineages present desirable yield, even when exposed to the disease, and have good agronomic traits. Then, these lineages were chosen as the best lineages to continue the breeding program. This lineages are USP 17-42.101, USP 17-42.121, USP 17-42.421, USP 17-42.111, USP 17-42.073, USP 17-42.043 and USP 17-42.223. And the most tolerant experimental lineages to soybean rust are USP 17-42.143, USP 17-42.101, USP 17-42.121, USP 17-42.284, USP 17-42.133, USP 17-42.102, USP 17-42.231, USP 17-42.592, USP 17-42.034, USP 17-42.222, USP 17-42.213, USP 17-42.074, USP 17-42.413, USP 17-42.402 e USP 17-42.103.

Keywords: *Glycine max; Phakopsora pachyrhizi*, Genetic tolerance; Rust effect.

STUDY OF *AUSTROPUCCINIA PSIDII* DURING INFECTION IN *EUCALYPTUS GRANDIS* REVEALS MEIOTIC-SPECIFIC FEATURES OF THE PATHOGEN

J. A. Ferrarezi[1]; C. A. A. Hayashibara[1]; T. B. de Lima[1]; M. da S. Lopes[1]; M. C. Quecine[1]

[1]Department of Genetics, College of Agriculture "Luiz de Queiroz", University of Sao Paulo, Piracicaba-SP, 11 Padua Dias Avenue, 13418-900, Brazil. E-mail: jessica.ferrarezi@usp.br.

The basidiomycete fungus *Austropuccinia psidii* is the causal agent of myrtle rust, which may infect approximately 270 Myrtaceae species. Among those, *Eucalyptus* spp. stand out as one of the most economical important genus attacked by this pathogen worldwide. *A. psidii* is a biotrophic pathogen and little is known about its biological and genetic aspects. We investigated the presence of meiotic genes in the pathogen's genome and transcriptome, and monitored meiosis-related structures by inoculating urediniospores of *A. psidii* in *E. grandis* and analysing a 52-day-long time course. Comparative analysis of 136 genes present in *Ustilago maydis,* a model basideomycete to sexual cycle*,* with *A. psidii* and other 4 rust species resulted in 116 genes common for all compared genomes. We found 123 meiotic genes in *A. psidii* draft genome, including cell cycle control, condensation and chromosome structure maintaining, DNA replication, chromatids maintaining protein, DNA and incorrect pairing repair, double strand break formation, strand invasion, synapse and recombination. Furthermore, genes related to sporulation and teliospore germination were found in *A. psidii*'s genome. Spore samples were collected at 0, 14 and 28 days after inoculation (DAI) for microscopic and gene expression analysis. We investigated the expression of 3 mitotic genes and 3 meiotic genes. We found teliospores during interaction with *E. grandis* and it is an indicator of the starting-point of the pathogen's meiosis. Relative expression ratios of the mitotic-related genes DNA helicase and G2-specific cyclin *cyb* were significantly down-regulated, while two meiotic genes – CDK assembly factor and checkpoint DNA exonuclease *rad1* - were significantly up-regulated in samples of spores collected 28 DAI in *E. grandi*s compared to 14 DAI, corroborating with the difference observed in teliospore abundance during time course monitoring. This clear evidence supports the hypothesis that *A. psidii* is able to undergo meiosis in teliospores produced during interaction with *E. grandis*. We conclude that *A. psidii* teliospores are produced in very low rates at early stages of rust in *E. grandis* and expressively increases around 50 DAI, basidiospores were observed after teliospore germination *in vitro*, as never described before in literature.
Key words: gene expression; myrtle rust; meiosis; mitosis; teliospore.

WITHIN-FAMILY GENOMIC SELECTION IN SUGARCANE FOR BROWN RUST RESISTANCE VIA MACHINE LEARNING

Alexandre Aono[1]; Ricardo Pimenta[1]; Melina Mancini[1]; Cláudio Cardoso-Silva[1]; Fernanda dos Santos[2]; Luciana Pinto[2]; Reginaldo Kuroshu[3]; Anete Pereira de Souza[1]

[1]Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, Brazil. [2]Centro de Cana, Instituto Agronômico de Campinas, Ribeirão Preto, Brazil. [3]Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo, São José dos Campos, Brazil. E-mail: alexandre.aono@gmail.com.

Sugarcane production is a major national source of income due to its bioenergetic potential. One of the most common pathogens affecting this crop is brown rust (*Puccinia melanocephala*), which has spread worldwide in growing areas generating considerable economic impact. Scientific advances concerning the genetic resistance of sugarcane to this disease have been limited because of its large, complex and polyploid genome. Besides, these studies face the lack of appropriate methodologies, softwares and genomic references. A full-sib progeny with 144 individuals from Agronomic Institute of Campinas sugarcane breeding program was used to develop a method for the identification of varieties resistant to brown rust. The progeny was genotyped using Genotyping-by-Sequencing. High quality reads were mapped to 7 genomic references. Five tools were used for SNP calling; the final SNP set was obtained from the intersection criteria of at least two of them. Due to sugarcane's aneuploidies, the reference allelic proportions of SNPs were used for analyses. Phenotypic data from 5 years of rust symptom scoring were evaluated using ANOVA and Tukey's test. Individuals were clustered in groups based on rust severity using several statistics methods; a T-test was performed to check for differences among groups. Using a leave-one-out validation strategy, SNPs were used to predict severity groups via machine learning (ML) with various algorithms (K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forests, Multilayer Perceptron, AdaBoost and Gaussian Naive Bayes) and feature selection strategies. In the final set, 14,540 putative SNPs and 2 phenotypic groups with considerable differences on rust severity were identified. The first models built using ML algorithms achieved a mean accuracy of 52%. However, using feature selection strategies we could reduce our dataset to 136 SNPs achieving a mean accuracy of 72%, and a maximum of 95% in the most promising algorithm. These results show the great potential of ML strategies to evaluate phenotypes in highly complex polyploids. Through the identification of non-linear relationships among genomic regions, these methodologies can auxiliate sugarcane genomic studies and be used for genomic selection in breeding programs to predict potential phenotypes.
Keywords: GBS, Polyploidy, Aneuploidy, Classification.

ORGANIZATION

GVENCK

Genetics and Plant Breeding Group
"Prof. Roland Vencovsky"

SUPPORT