VÚŽV
Institute of Animal Science

Workshop

Immunogenetics and Breeding in Livestock Species

Virtual, November 8-9, 2022

Held as part of the project LTV20020 “Support for the participation in the activities of the International Society for Animal Genetics (ISAG)”
of the program INTER/VECTOR – INTER/EXCELLENCE
of the European Union
granted by the Ministry of Education, Youth and Sports of the CR.

Department of Genetics and Breeding of Livestock Species
Institute of Animal Science
Prague
Czech Republic

Workshop schedule

Tuesday, November 8, 2022, CET	Presenter	Affiliation	Title
14.00–14.10	Karel Novak	Institute of Animal Science, Prague	Introductory comments
14.10–14.50	Dr Sabine Hammer	Institute of Immunology, University of Veterinary Medicine, Vienna	The swine leukocyte antigen (SLA) complex: molecular genetics and importance in veterinary vaccine research
14.50–15.30	Prof Petr Hořín	Faculty of Veterinary Medicine, University of Veterinary Sciences, Brno	Genetics and comparative genomics of NK cell receptors
15.30–15.45	Thales Galdino Andrade	Ribeirão Preto School of Medicine, University of São Paulo	Insights on the diversity and distribution of cattle major histocompatibility complex *DRB3* alleles
15.45–16.00	Prof Abbas Rafat	Department of Animal Science, University of Tabriz	Identification of functional candidate genes associated with mastitis and milk production traits based on transcriptome-wide association and multi-tissue colocalization approaches using the available FarmGTEx dataset
16.00–16.10			Break
16.10–16.50	Prof Dirk Werling	Royal Veterinary College, London	Species and breed-specific innate pattern recognition receptors: an immunologist’s plea to take genetics more into account when designing vaccines
16.50–17.30	Dr Terhi Iso-Touru	Natural Resources Institute - LUKE, Finland	Bovine mammary epithelial cells response to two different mastitis pathogens
17.30–17.45	Dr Nidhi Sukhija	ICAR-National Dairy Research Institute, Karnal, India	Genomic clues for selective sweep regions in Indian Gir and Tharparkar cattle on BTA6
17.45–18.00			Discussion, conclusions

Wednesday, November 9, 2021, CET	Presenter	Affiliation	Title
14.00–14.10	Karel Novak	Institute of Animal Science, Prague	Introductory comments
14.10–14.50	Prof Ottmar Distl	University of Hannover	Screening selection signatures associated with disease resistance in sheep using whole-genome sequencing data
14.50–15.30	Dr Steven Fiddaman	Department of Biology, University of Oxford	Investigating the immunogenetics and function of Toll-like receptors in large genomic datasets
15.30–15.45	Prof Lior David	Department of Animal Sciences, The Hebrew University of Jerusalem	Multigenic resistance of common carp to the viral CyHV-3 disease, infection and infectivity
15.45–16.00	Avon Augustin Nalpadan	Center for Integrated Genetics, Norwegian University of Life Sciences, Norway	Developing Cas9 expressing MDBK cell lines for GeCKO screening against bovine diseases
16.00–16.10			Break
16.10–16.25	Karel Novák	Institute of Animal Science, Prague	Effects of *TLR* diversity on reproductive traits
16.25–17.05	Prof Yana Safonova	Department of Computer Science, John Hopkins University, Baltimore	Variations in cattle antibody repertoires correlate with immune responses to vaccines against the bovine respiratory disease
17.05–17.20	Kalifa Samaké	Faculty of Natural Sciences, Charles University, Prague	Balancing selection acting on haplotypes of bovine *TLR*s
17.20–17.35	Kanaka K.K.	ICAR-Indian Veterinary Research Institute, Bareilly, India	Characterization of ovalbumin coding sequence revealed altered antigenic index in predicted ovalbumin protein structure in two strains of WLH chicken
17.35–18.00			Discussion, conclusions

Abstracts

Introductory comments

Karel Novák*

Institute of Animal Science, Prague - Uhříněves, Czech Republic

It is obvious that since the beginning of animal immunogenetics, efforts have been aimed at improving resistance to animal diseases. The most straightforward way to achieve this goal is to apply immunogenetics knowledge in breeding. The importance of this topic can be exemplified by the health-oriented programs for cattle of large breeding companies, e.g., Immunity+ (Semex), and DWP$ and CW$ breeding value indices (Zoetis).

A factor changing this situation is that commercial livestock breeding has become a self-standing discipline based on statistical models. Moreover, during the last decade, tens of thousands of rather formal genomic markers started being used instead of the approximately several tens used before. The question arises of whether the efficiency of formalized breeding systems can be improved by the inclusion of functional knowledge.

Some recent projects were probably in response to the perceived low efficiency of the purely statistical approach for finding breeding values of formal markers. Among others, this situation was defined in the paper by Boichard et al. (2016), who stated that after the initial black box strategy, genomic evaluation might be more accurate and more persistent when integrating biological knowledge.

A logical reaction to the perceived inherent gaps in genomic selection was a systematic search for causative gene variants. However, although genome-wide association studies have been applied on a large scale, only a limited number of causal variants had been identified using the GWAS technique in cattle with this general method by 2015, i.e., in the middle of the last decade and five years after the introduction of whole-genome genotyping arrays.

In this situation, the integration of functional knowledge from bioinformatic databases and from omics approaches including gene expression data has become a realistic alternative approach. Generally, two strategies are applicable for the identification of targets for health breeding. On the one hand, it is possible to increase the amount of experimental work, i.e., to collect more data from larger populations of animals, to increase the resolution of genotyping arrays, to extend the computing capacity, etc. The second approach is based on bioinformatics, which means that the exploitation of the available molecular data on the formation of complex traits can be made more efficient. This approach is applicable to immune traits, an important subset of relevant phenotypic traits.

Among the first projects applying this approach were the projects initiated in the Animal Genomics and Improvement Laboratory of the USDA since July 2017 (Fang et al. 2019). Additionally, the projects in other leading labs included the incorporation of functional annotation data in cattle GWASs and genomic selection. This switch was paralleled by the development in other labs specialized in genomics and advanced breeding in cattle. In the example of the paper by Cai et al. (2018) from the Aarhus Laboratory, the authors suggested prioritizing the candidate genes that are identified using GWASs by considering additional information about the biological role of the neighbouring genes. In other words, this approach integrates pure genomic and functional data to identify causal genes. By combining the two approaches, the requirements of Bayesian inference can be met. It can be reasonably assumed that the possibilities for the transfer of immunogenetic knowledge to breeding are wider and will be supported by emerging bioinformatic approaches in the future.

The importance of including immunity traits in breeding has always been reflected in the scope and activities of the International Society of Animal Genetics (ISAG). Moreover, the ISAG originated from the International Society for Animal Blood Group Research (I.S.A.B.R.) in Torino in 1988. The I.S.A.B.R. was in turn derived from the European Society for Animal Blood Group Research (E.S.A.B.R.), established in 1964 in Prague. This development was preceded by a series of conferences on animal blood groups and histocompatibility antigens beginning in 1954.

The current state of our capacity to transfer immunogenetic knowledge to breeding was surveyed in the presentations of the workshop, although it is not possible to fully cover such a complex topic in the time period of a limited event. Nevertheless, we believe that the workshop will contribute to the determination of further productive directions of research and will be helpful in the identification of possibilities for cooperation.

References:

Boichard D., Ducrocq V., Croiseau P., Fritz S., 2016. Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies* **339**:274–277.

Cai Z., Guldbrandtsen B., Lund M.S., Sahana G., 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics* **19**:656.

Fang, L. et al. 2019. Comprehensive analyses of 723 transcriptomes enhance biological interpretation and genomic prediction for complex traits in cattle. *Genome Research* **30**:790-801.

The swine leukocyte antigen (SLA) complex: Molecular genetics and importance in veterinary vaccine research

Sabine E. Hammer[1*]

[1] *Institute of Immunology, Department of Pathobiology, University of Veterinary Medicine Vienna, Vienna, Austria*

Livestock species are a major source of animal protein worldwide. To ensure animal health and food safety, it is essential to prevent infectious diseases via biosecurity and the use of well-designed vaccines and therapeutics. Advances in genomics have informed our understanding of the complexity of the immune system and the genes that influence disease and vaccine responses, with the most important being the major histocompatibility complex (MHC). Viral, bacterial and parasitic infections have severe influences on animal welfare and livestock economy. Development of an adaptive immune system to fight off these infections relies on effective activation of T lymphocytes and their recognition of pathogen-derived peptides presented by MHC molecules to T-cell receptors (TCR). The highly polymorphic nature of the MHC allows for the presentation of a wide panel of antigenic peptides and thus influences disease resistance and vaccine responsiveness.

The swine leukocyte antigen (SLA) system is among the most well-characterized MHC systems in nonhuman animal species. A systematic nomenclature for the genes, alleles, and haplotypes of the SLA complex is critical to research on swine genetic diversity, immunology, health, and vaccinology, as well as organ and cell transplantation. Based on our new, detailed annotation of the Sscrofa11.1 genome assembly, the SLA complex encodes approximately 150 loci, with at least 120 genes predicted to be functional. Despite the ongoing domestication process involving selection for favourable traits, pigs have maintained a high degree of SLA diversity, as demonstrated by the presence of 266 and 227 class I and class II alleles, respectively. Pig disease models provide a better understanding of host–pathogen interactions. Pathogen effects on SLA gene expression drive the regulation of swine immune responses. Novel trait association data indicate that SLA alleles or haplotypes may be useful genetic markers for improving pig breeding programs. Swine have become the preferred preclinical large animal model for biomedical studies, transplantation, xenotransplantation, and regenerative medicine research. Allogeneic transplantation research in pigs has improved the understanding of rejection mechanisms of both host-versus-graft and graft-versus-host disease. Improved cross-matched genetically engineered pigs could reduce antibody-mediated rejection of pig xenografts in highly HLA-sensitized patients. Modifying SLA genes could improve pigs as donors for xenotransplantation.

Future perspective

(1) The impact of SLA genes on swine production and health traits needs to be attributed to individual SLA locus alleles and not just haplotypes.

(2) Renewed typing methods, from PCR SSP to NGS, will enable reliable typing of outbred pigs. To truly explore diversity, data based on large cohorts of pigs are necessary.
(3) Functional studies of MHC effects on cell interactions and on microbiota diversification are needed to understand the impact of SLA genes on the education of the porcine immune system.
(4) In-depth analysis of peptide presentation via major SLA genes will identify a broad range of functionally relevant vaccine targets.
(5) Identification and maintenance of important SLA-defined pig lines (e.g., NIH/MGH, Yucatan, or Babraham pigs) are essential to provide resources for pig biomedical models.
(6) Future tool development is needed for the swine biomedical model; this includes SLA class I and first SLA class II tetramers, T-cell receptor profiling, SLA-informed SNP chips, and panels of monoclonal antibody reagents to swine immune proteins.
(7) The availability of well-characterized, genetically engineered pigs for human disease models will lead to the development and validation of novel therapeutics and improvements in xenotransplantation research.
(8) Human cross-matching with SLA class I and II will facilitate xenotransplantation. Histocompatibility testing of pigs needs to be improved in analogy to human allogeneic transplantation.

Keywords: *Sus scrofa*, swine leukocyte antigen, SLA polymorphism, allogeneic, xenogeneic, vaccine responses

Genetics and comparative genomics of NK cell receptors

Petr Hořín[1,2*], Ján Futas[1,2], Jana Bubeníková[1,2], April Jelinek[1], Martin Plášil[1,2], Karla Stejskalová[1], Leona Vychodilová[1], Eva Jánová[1,2]

[1] *Dept. of Animal Genetics, Faculty of Veterinary Medicine, University of Veterinary Sciences Brno, Czech Republic*

[2] *CEITEC VETUNI, RG Animal Immunogenomics, University of Veterinary Sciences Brno, Czech Republic*

Natural killer (NK) cells are an important part of innate immunity. They are able to directly recognize and kill virus-infected and/or tumour cells. They form a heterogeneous population consisting of subsets differing in their functional capacities and cell surface phenotypes. This heterogeneity is due to the differential expression of natural killer cell receptors (NKRs), representing a complex repertoire of germline-encoded receptors. Two types of NKRs regulating multiple NK cell activities can be distinguished: activating and inhibitory. After they bind an appropriate ligand, they can activate or inhibit, respectively, immune mechanisms mediated by NK cells. In these mechanisms, both qualitative and quantitative variability of NK cells and their functions are important. This variability is based on different types of

polymorphisms of NKR coding genes, including single nucleotide polymorphisms (SNPs) and copy number variations (CNVs). In addition to a number of individual genes belonging to various gene families, such as natural cytotoxicity triggering receptor (*NCR*) genes, two complex genomic regions encode NKRs that play crucial roles in the disease, infection and reproduction of placental mammals. The leukocyte receptor complex (LRC) contains genes encoding killer immunoglobulin-like receptors (KIR), while the natural killer complex (NKC) codes for NKRs with lectin-like structure (KLR, formerly Ly49). Both types of receptors can bind major histocompatibility complex (MHC) class I molecules as ligands, and in evolutionary terms, they can substitute for each other. Similar to MHC class I genes, *NKR* genes are highly polymorphic. Their allelic SNP and CNV variability underlies the overall avidity in the receptor–ligand interaction and as such is crucial for educating NK cells to distinguish between self, nonself, induced-self and missing self.

The repertoire of NKR genes differs between mammalian families, with no single conservative model of NKR genes observed across all placental mammals. Even closely related taxa may differ in LRC/NKC genomic structure and gene contents. In extreme cases, such as in humans and mice, genes for one type of receptor are missing or pseudogenized, while the other type is expanded. In humans, an expansion of KIR genes was observed within LRC, while a single Ly49 (*KLRA*) sequence identified in the NKC is a pseudogene. On the other hand, expansion of Ly49 (*KLRA*) genes was reported in mice, while within LRC, no KIR gene was observed, and two KIR genes were relocated on the X chromosome. Various combinations of LRC and NKC gene contents have been reported in different mammalian species. Therefore, analyses of mammalian families in which NKR genes have not yet been studied, especially domestic animal species, have the potential to provide new information on the evolution of NKRs and may also be of practical importance. In our work, we studied NKR genes and complex genomic regions in three domestic animal families, Equidae, Camelidae and Felidae. All three of them comprise species exposed to various selection pressures exerted by different pathogens, especially in free-ranging species, and/or to strong and long-term pressure from artificial selection, such as in domestic animals. We have identified a new model of the NKC/LRC genomic structure in equids, provided a detailed annotation of these regions in camelids and started analysing the currently unknown repertoire of NKRs in felids. For horses and camelids, we have also developed microsatellite-based tools for analysing the genetic diversity of their NKR and MHC regions. Currently, we study receptor–ligand (NKR-MHC) interactions in their evolutionary context.

This work was supported by the Austrian Science Fund FWF, project P29623-B25, the CEITEC-Central European Institute of Technology, Vetuni Brno, project IGA 107/2020/FVL, and the FWF/GA ČR joint project GA CR 21-28637L.

Insights into the diversity and distribution of cattle major histocompatibility complex DRB3 alleles

Thales Galdino Andrade*, Maurício Scavassini Peña, Jéssica Fiorotti, Renan de Souza Bin, Paulo Henrique do Prado Almeida, Isabel Kinney Ferreira de Miranda Santos
Ribeirão Preto School of Medicine, University of São Paulo, Ribeirão Preto, Brazil

The bovine major histocompatibility complex (MHC), also known as bovine leucocyte antigen (BoLA), is the genomic region that encodes the most important molecules for antigen presentation in immune responses. At first, the evidence for MHC in bovines pointed to a locus of two antigens, one detected by cytotoxic antiserum (MHC I) and another studied by mixed lymphocyte culture tests (MHC II). Currently, it is known that the bovine MHC is located on chromosome 23 with more than 150 genes and multiple alleles. The most studied gene in the BoLA region is the highly polymorphic BoLA-DRB3, which encodes a β chain with a peptide groove domain involved in antigen presentation for helper T cells to develop robust humoral responses. BoLA-DRB3 alleles have also been demonstrated to be associated with a diversity of infectious diseases and production traits, such as mastitis, trypanosomiasis, and tick burden. To catalogue these alleles, two nomenclature methods were proposed, and the current use of both systems makes it difficult to list, comprehend and apply these data in an effective way. Despite this drawback, there are several reports of BoLA-DRB3 allele frequencies in more than 26 countries evaluating many breeds, but studies in Africa and Australia are lacking. We strongly encourage new studies addressing allelic frequencies, which can be incorporated into databases for practical approaches, such as markers for infectious disease resistance, production traits, and ectoparasitic interactions; genetic diversity studies; B- and T-cell epitope mapping; and the conservation of creole breeds.

Identification of functional candidate genes associated with mastitis and milk production traits based on transcriptome-wide association and multi-tissue colocalization approaches using the available FarmGTEx dataset

Sevda Hosseinzadeh[1], Abbas Rafat[1*], Arash Javanmard[1], Lingzhao Fang[2]
[1] *Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran*
[2] *MRC Human Genetics Unit at the Institute of Genetics and Molecular Medicine, the University of Edinburgh, Edinburgh, EH4 2XU, UK*

Research on the simultaneous genetic assessment of milk production and mastitis has a long history. A challenging problem that arises in this area is the existence of a correlation between lactation performance and increased risk of mastitis due to selection. Understanding the complexity and genetic architecture of milk production and mastitis is critical to developing efficient breeding plans. Thus, previous studies have shown that some genes in different tissues

have significantly different expression for mastitis and milk production. For example, *DGAT1* (Fang et al., 2017), *MAPK14*, *FRAP1*, *EIF4EBP2*, *GSK3A*, and TSC1 (Bionaz and Loor, 2011) for milk production and *LOC515333*, *SAA3*, *CD14*, *NFKBIA*, *APOC2*, *LOC100335608* (Pereira et al., 2021), *SPEF2*, *FAM151A*, *IMMP2L*, *CCND2*, *NFKB1*, *KRT32*, *DHX58*, *KCNH4*, STAT5A, *STAT3* (Cai et al., 2018), *GLYCAM1*, *B2M*, *CD74*, *BoLA-DRA*, *FCER1G*, *SDS*, and *NFKBIA* (Asselstine et al., 2019) for mastitis were identified as biomarkers. In recent years, researchers have shown increased interest in genome-wide association studies (GWASs) (Li et al., 2011; Benner et al., 2016). GWASs highlighted thousands of loci or variants associated with complex traits such as disease, production, reproduction and conformation in livestock (Freebern et al., 2019; Jiang et al., 2019a; Tian et al., 2020). Despite its long success, the GWAS have a number of problems in application. For example, the majority of loci are located in the non-coding part of the genome (Maurano et al., 2012; Finucane et al., 2015; Visscher et al., 2017). Hence, it is remarkable that non-coding regions of the genome play an important role in the regulation of gene expression (Barbeira et al., 2021). To overcome this problem, some approaches have been developed. Gene expression is the representation of the molecular phenotype and improves the functional explanation of GWAS results (Guo et al., 2015). A number of recent studies have shown that the integration of RNA-Seq data on intramammary infection of Holstein cattle and large-scale GWAS data from three dairy cattle breeds (Jersey, Nordic Red, and Holstein cattle) improves the biological interpretation of GWAS loci and improves genomic prediction for milk production and mastitis resistance (Fang et al., 2017). Recent methods focus on overcoming these problems by proposing the identification of variants involved in regulating gene expression levels through expression quantitative trait loci (eQTLs). The TWAS method attempts to combine GWAS and eQTL approaches to control complex diseases or traits. With this motivation, we focused on previous bovine datasets from the FarmGTEx project proposed for milk yield and mastitis and mainly compared the results of TWAS and colocalization approaches to identify functional candidate genes associated with mastitis and milk production traits on the basis of multi-tissue transcriptome records. We used Gene Ontology (GO) to identify significant genes with DAVID and the STRING and GeneMANIA databases to identify interaction networks. In addition, to calculate the correlation between tissues, available z scores were used in the TWAS results. Taken together, the present results confirm that *LYNX1*, *DGAT1*, *C14H8orf33* and *LY6E* are significantly associated with milk production in 8, 6, 5, 5 and 5 tissues, respectively. *FBLX1* was identified as a significant gene for mastitis. We identified five significant terms for biological processes, three significant terms for cellular components, and four significant terms for molecular function. The following common genes emerged for both the TWAS and colocalization approaches: *CLN3* and *ZNF34* for milk production. In the present study, genes associated with mastitis and milk production traits were identified in different tissues in cattle, and their function and interactions were investigated. On this basis, we conclude that TWAS colocalization can improve our understanding of the potential health status control mechanism in high-yielding dairy cows.

Species- and breed-specific innate pattern recognition receptors: an immunologist's plea to take genetics more into account when designing vaccines

Dirk Werling[1, *]

[1] *Centre for Vaccinology and Regenerative Medicine, Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, Hatfield, UK*

In recent years, the differences in the composition and function of cellular components of the immune system between various mammalian species have been highlighted. Interestingly, whereas much genetic information is available for different laboratory rodent strains and specific human geo-ethnic groups, we do not yet have a great understanding of how similar differences identified in farmed animals affect their immune response. However, these genetic differences may affect not only how animals respond to pathogens but also how they respond to vaccines and address differences in the adaptive and innate immune systems. Here, pattern recognition receptors play an important role. Indeed, many members of the C-type lectin family of glycan-binding receptors have been ascribed roles in the recognition of microorganisms and serve as key receptors in the innate immune response to pathogens. Other mammalian receptors have become targets through which pathogens enter target cells. These receptor roles have often been documented with binding studies involving individual pairs of receptors and microorganisms. Our own studies using C-type lectin receptors in cattle show that the results are consistent with interactions previously ascribed to the receptors, but they also highlight binding to additional sugar targets that have not previously been recognized. Furthermore, recent evidence suggests not only that cattle behave differently to an infection with Mycobacteria but also that several cattle breeds may be more resistant to infection with the zoonotic pathogen *Mycobacterium bovis*. Whole-genome sequencing of the Brown Swiss genome revealed several potential candidate genes, in particular for Toll-like Receptor-2 (TLR2), a PRR that has previously been described to be involved in mycobacterial recognition. Cloning of the *TLR2* gene and subsequent gene-reporter and chemokine assays revealed that specific SNPs present in BS and *Bos indicus* breeds resulted in a significantly higher response to mycobacterial antigens as well as tri-acylated lipopeptide ligands in general. This clearly indicates that we cannot even extrapolate data generated in one cattle breed to others on a 1:1 basis, and we clearly have to take genotype to phenotype variations far more into account. However, this also potentially means that vaccines generated using "European" breeds may not work with the same efficacy in other breeds, specifically *Bos indicus* breeds, and need to be tested far more rigorously before being brought into the target market.

Genomic clues for selective sweep regions in Indian Gir and Tharparkar cattle on BTA6

Nidhi Sukhija[1], Archana Verma[1], K.K. Kanaka[1], Jayakumar Sivalingam[2], Rekha Sharma[2]

[1] *ICAR-National Dairy Research Institute, Karnal-132001, Haryana, India*

[2] *ICAR-National Bureau of Animal Genetic Resources, Karnal-132001, Haryana, India*

BTA6 harbours QTLs related to milk production and disease susceptibility. Selective sweeps are the regions of co-inheritance of neutral mutations along with beneficial mutations. The search for intersections between selective sweep regions and QTLs was performed in the herd of Livestock Research Centre (LRC) of NDRI in Karnal (n=7,7). A total of 191 selective sweep regions were obtained by employing the composite likelihood ratio (CLR) and FST approaches in the Gir and Tharparkar breeds. The *GRID2*, *SEC31A*, *PTPN13* and *LDB2* genes were linked to selective sweep regions in Gir cattle. The *STIM2* and *UGT2A1* genes were linked to selective sweep regions in Tharparkar cattle using CLR. The *FAM13A*, *RBPJ*, *PDS5A*, *MAPK10*, *HS3ST1* and *SORCS2* genes were identified by the FST approach. The FAM13A gene is related to the production traits and adaptation, while the *MAPK10* gene is related to the reproductive traits and adaptation. A total of 12, 8 and 4 selective sweep regions overlapped with QTLs related to milk protein percent, milk unglycosylated kappa casein percentage and milk kappa casein amount present on BTA6 by the CLR and FST approaches in Gir and Tharparkar cattle. Selective sweep studies are important because they serve as a validation platform for genomic selection strategies, help narrow the dataset by 10 times from the SNP level to the sweep level and serve as a pilot study for GWASs to establish correlations with traits of interest.

Investigating the immunogenetics and function of Toll-like receptors in large genomic datasets

Steven Fiddaman[1*], Christophe Klopp[2], Mathieu Charles[3], Philippe Bardou[4], Ophélie Lebrasseur[5,6], Martijn Derks[7], Jens Schauer[8], Christian Reimer[8], Johannes Geibel[8,9], Almas Gheyas[10], Adrian Smith[1], Robert Schnabel[11], Maria Luisa Martin Cerezo[12], Greger Larson[13], Ismael Ng'ang'a[14], William Muir[15], Margaret Lange[16], Dominic Wright[12], Hans Cheng[17], Henner Simianer[9], Steffen Weigend[8,9], Wesley Warren[11], Richard Crooijmans[7], Olivier Hanotte[18,19,20], Jacqueline Smith[10], Michele Tixier-Boichard[21], Laurent Frantz[22]

[1] *Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK*

[2] *INRAE, MIAT UR875, Sigenae, F-31326, Castanet Tolosan, France*

[3] *University Paris-Saclay, INRAE, AgroParisTech, GABI, Sigenae 78350, Jouy-en-Josas, France*

[4] *Université de Toulouse, INRAE, ENVT, GenPhySE, Sigenae, F-31326, Castanet Tolosan, France*

[5] *Centre d'Anthropobiologie et de Génomique de Toulouse (CAGT), CNRS UMR 5288, Université Toulouse III Paul Sabatier, Toulouse 31000, France*

[6] *Instituto Nacional de Antropología y Pensamiento Latinoamericano, Ciudad Autónoma de Buenos Aires, Argentina*
[7] *Animal Breeding and Genomics, Wageningen University and Research, Wageningen, The Netherlands*
[8] *Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, 31535 Neustadt, Germany*
[9] *Center for Integrated Breeding Research, University of Goettingen, 37075 Göttingen, Germany*
[10] *The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK*
[11] *Department of Animal Sciences, University of Missouri, Columbia, MO, USA*
[12] *AVIAN Behavioural Genomics and Physiology, IFM Biology, Linköping University, Linköping 58183, Sweden*
[13] *The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK*
[14] *Queen Mary University of London, Bethnal Green, London, E1 4NS, UK*
[15] *Purdue University, Department of Animal Sciences, West Lafayette, IN, USA*
[16] *Department of Molecular Microbiology & Immunology, University of Missouri, Columbia, MO, USA*
[17] *USDA, ARS, USNPRC, Avian Disease and Oncology Laboratory, East Lansing, MI, USA*
[18] *Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, NG7 2RD UK*
[19] *Centre for Tropical Livestock Genetics and Health, The Roslin Institute, Edinburgh, EH25 9RG UK*
[20] *LiveGene, International Livestock Research Institute (ILRI), P.O. 5689, Addis Ababa, Ethiopia*
[21] *University Paris-Saclay, INRAE, AgroParisTech, GABI , 78350 Jouy-en-Josas, France*
[22] *Palaeogenomics Group, Department of Veterinary Sciences, LMU Munich, Germany*

Genomic data of domesticated animals are becoming increasingly ubiquitous and easy to generate. Coupled with the relative ease of computationally processing whole genomes, it is now possible to gather and jointly analyse many hundreds or even thousands of genomes. Leveraging datasets of this size brings with it analytical power that has scarcely been seen outside of human data. Motivated by a desire to better understand the domestic chicken, the Chicken Genomic Diversity Consortium was established in 2021, bringing together >25 researchers from >10 institutions. Currently, ~4,800 chicken and junglefowl genomes are being processed through a state-of-the art bioinformatic pipeline at the Leibniz Supercomputing Facility, and it is expected that genomes will be continually added. The dataset comprises chickens from a diverse range of sources, including feral birds and local breeds from a variety of locations worldwide, European fancy breeds, commercial birds, experimental lines, and even

ancient birds from archaeological sites. It is hoped that a sizeable fraction of extant chicken genomic diversity will be represented by this dataset. Alongside several other diverse goals, one of the aims of the consortium is to analyse immune diversity at several key loci. Among the immune loci to be analysed are the Toll-like receptors (TLRs), which are pattern-recognition receptors at the front line of defence against pathogens. Genetic polymorphisms in TLRs are known to influence host disease susceptibility, and thus, an understanding of TLR diversity can be an important adjunct to studying infectious diseases of chickens. In a preliminary analysis of >1600 chickens and junglefowl, we found that chicken TLRs are highly diverse. Moreover, several TLRs deviated significantly from neutrality, and we found evidence of selective sweeps affecting *TLR2A*, *TLR2B*, *TLR5*, *TLR7* and *TLR21*. Further analysing sites under positive selection, we found that a significant fraction of these have known ligand-binding or dimerization functions, implying important functional changes. We also identified a number of high-impact variants, such as internal stop codons, frameshift mutations, lost start codons, and splice site variants, some of which are found at high prevalence in certain breeds of chicken. These data therefore provide the foundation to catalogue chicken TLR diversity – benign, advantageous and disadvantageous – to inform future breeding efforts.

Multigenic resistance of common carp to viral CyHV-3 disease, infection and infectivity

Lior David[1*], Batya Dorfman[1], Roni Tadmor-Levi[1], Mor Amir[1], Evgenia Marcos-Hadad[1]

[1] *Department of Animal Sciences, R.H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel*

Infectious diseases stress animals and damage production in farmed animals. Challenges are even larger under aquaculture conditions, where disease control and prevention measures are limited. Breeding genetically disease-resistant strains stands out as a sustainable solution to this problem. Common carp is a food fish cultured worldwide, and it suffers from outbreaks of a major lethal disease caused by the cyprinid herpes virus 3. Our group have been developing genetically resistant and susceptible common carp strains. Although phenotypic measurements might be similar, several key aspects differentiate monogenic and multigenic resistances. Here, we will touch upon infection and infectivity differences between susceptible and resistant fish. Following infection, the viral load in the spleen of live susceptible fish was 100 times higher than that in resistant fish, leading to four times higher cumulative mortality of the former. CyHV-3 resistance relies on improved immunity reflected in different signalling pathways and sets of cytokines. Interestingly, viral load levels in the spleen and mortalities depended not only on the type of infected fish but also on the type of infecting fish. Infection by susceptible fish led to higher viral loads and mortalities than infection by resistant fish. Therefore, we provide empirical evidence showing that susceptible individuals are not only more prone to viral

infection but also support more replication of the pathogen in their tissues, shedding more viral particles to the environment and infecting more than resistant individuals.

Developing Cas9-expressing MDBK cell lines for GeCKO screening against bovine diseases

Avon Augustin Nalpadan[1*]

[1] *CIGENE - Center for Integrated Genetics, Norwegian University of Life Sciences (NMBU), Norway.*

Recent developments in CRISPR/Cas9 technology have created several possibilities for gene editing in various model organisms. However, little research has focused on applying this research to develop suitable tools for gene editing in large production animals against viral diseases. The objective of this project was to develop Cas9-integrated Madin–Darby bovine kidney (MDBK) cell lines for genome-wide CRISPR/Cas9 knockout screening against bovine viral diarrhoea virus (BVDV) and bovine coronavirus (BCoV). A two-step vector was used for screening. LentiCas9-Blast was transduced into MDBK cells to generate *Cas9*-integrated cell lines. Stable cell lines were then selected using blasticidin selection, followed by expanding homogenous cell lines from single-cell clones obtained through serial dilutions. The lines were then grouped based on the *Cas9* gene expression (using RT–qPCR) and protein levels (using Western blotting). The *ADAM10* gene was then cloned into lentiGuide-Puro, the construct was transduced into the grouped cell lines and selected using puromycin. The lines were then sequenced to confirm genome editing. The major result involved the successful development of multiple *Cas9*-integrated cell lines with varied *Cas9* expression that were ready for subsequent genome-wide screening. Additionally, CRISPR editing efficiencies of cell lines were correlated with their *Cas9* gene expression to obtain a few optimal cell lines ready to perform GeCKO screening.

Keywords: BVDV, BCoV, CRISPR/Cas9, MDBK cell lines, transduction, GeCKO screening

Effects of *TLR* gene diversity on reproductive traits

Karel Novák[1*]

[1] *Institute of Animal Science, Prague, Czech Republic*

To date, a series of observations of the effects of TLR gene variants on different health traits of cattle have been published. They comprise mostly the effects on mastitis incidence, somatic cell count, and paratuberculosis resistance. Consequently, diversity in this group of genes was documented in the population of Czech Red Pied cattle (Czech Simmental), and an association

study with a set of production, health, and reproductive traits was performed. The genotyping workflow included design of the targeted amplification procedure, hybrid resequencing in the population with two independent technologies, discovery of polymorphism and validation with primer extension reactions, genotyping of a set of bulls representing the population and calculation of the association indicators. Eight polymorphisms in *TLR1*, *TLR2* and *TLR6* were significantly associated with reproductive traits, such as calving ease. The calf vitality index seemed to be correlated as well. On the other hand, unaffected traits included early traits such as cyst formation in ovaries and the index of early reproduction disorders. The shared pattern of associated traits in this group of *TLR* genes might be due to the interaction of their products in the heterodimers formed by TLR2 on one side with TLR1 and TLR6 on the other. A similar effect on reproductive traits was observed for *TLR4* and *TLR5*. No effects on udder health traits were observed, in contrast to expectations. The association with milk production traits was present only for the *TLR5* polymorphisms. This finding is consistent with the *TLR* variation effects known in model species, mice and humans. Non-immune functions of *TLR*s, e.g., participation in Rho/ROCK signalling, may be responsible for the observed effect on reproductive traits. However, the absence of correlation between the predicted effects and detected associations for individual alleles points to the localization of causative variants in the non-coding regulatory regions. Consequently, knowledge of the long-range variability in the regulatory regions of *TLR*s would reduce the possibility of misinterpretation of association studies, which would otherwise be difficult to avoid.

Balancing selection acting on haplotypes of bovine *TLR*s

Kalifa Samaké[1*]

[1] *Department of Genetics and Microbiology, Faculty of Natural Sciences, Charles University, Prague*

The *TLR* genes coding for Toll-like receptors of anti-bacterial series, namely, *TLR1*, *-2*, *-4*, *-5* and *-6*, were resequenced in Czech Simmental (Czech Red). PacBio sequencing allowed us to determine haplotypes within the range of the designed amplicons. A more general statistical reconstruction of haplotypes from individual reads was carried out in parallel. The PacBio read results demonstrated non-randomly distributed frequencies of haplotypes. 15 haplotypes revealed in amplicon 1 in the proximal part of the *TLR2* transcript formed two distinct groups. Similar results were found for the haplotypes obtained by statistical reconstruction in *TLR2* and *TLR5*. Similar to direct reconstruction in *TLR2*, the trend for a bimodal distribution was stronger in proximal regions of the transcripts. The bimodal clustering of *TLR* haplotypes was reported earlier in cattle and other animal models; however, a final interpretation of this disequilibrium is still missing. Alternating infectious agents might be a factor causing balancing selection. Two different functions performed by the *TLR2* gene or its product are another possible mechanism.

An example of a dual function might be the formation of two kinds of heterodimers, TLR2/TLR1 and TLR2/TLR6. Nevertheless, the association of the groups of haplotypes with the transcript proximal region suggests that the selection target is located in the 5' regulatory regions of the *TLR* genes, although functional interactions in the proximal part of the transcript cannot be excluded.

Characterization of the ovalbumin coding sequence revealed an altered antigenic index in the predicted ovalbumin protein structure in two strains of WLH chickens

K.K. Kanaka[1,2*], Nidhi Sukhija[2], N.G. Sagar[1], Rajendra Prasad[1]

[1] *ICAR-Indian Veterinary Research Institute, Bareilly, India.*

[2] *ICAR- National Dairy Research Institute, Karnal, India*

Ovalbumin is the most abundant protein in egg white, contributing up to 54%. It is a storage protein and major source of amino acids for the developing embryo. Sequencing of the ovalbumin CDS from the IWI and IWK lines of white leghorn chickens revealed that both products were 1161 bp in length. Comparison of the obtained nucleotide sequence and deduced amino acid sequence with that of available sequences of the reference genome confirmed that the sequence was of the ovalbumin gene encoding a peptide of 386 amino acids in both cases. T>C transition at the 15th position and C>T transition at the 237th position were observed in both lines. T>C transition at the 291st position and G>A transition at the 562nd position were observed in the IWI line. G>A transition at the 562nd transition changed the amino acid alanine to threonine at the 188th position in the amino acid sequence, while other mutations did not change any of the amino acids. Based on the predicted structure of the protein using the coding sequence, the molecular weights of the ovalbumin protein from the IWI and IWK lines were 42.91 kDa and 42.55 kDa, respectively. The antigenic index was determined for the predicted protein structure for both IWI and IWK lines. It was observed that the index was similar in the two lines until amino acid number 300 from the N-terminus and from 300 to 386, and the index graph showed multiple changes. Based on our results, we hypothesize that there may be potential antigenic determinant candidates present in this region that may be associated with consumer egg protein allergy.