prakapenka@gmail.com
(773) 299-8727

dprakapenka.github.io
US Citizen

Dzianis Prakapenka

SUMMARY

I am an experienced computational biologist with over 13 years of experience in high performance scientific computing. I have a strong background in developing and optimizing computing pipelines, tools, and methods for analyzing large data efficiently. I have advanced knowledge of Python, C++, Fortran, R, and extensive experience with parallel data processing, scripting, developing and running software on HPC clusters.

EDUCATION

University of Minnesota Minneapolis & Saint Paul, MN
Ph.D. in Bioinformatics and Computational Biology *Aug 2022*

- **Thesis**: Haplotype Computing Tools for Genomic Prediction and Estimation and Detection of Epistasis Effects in Holstein Cattle

University of Minnesota Minneapolis & Saint Paul, MN
M.S. in Bioinformatics and Computational Biology *March 2019*

- **Concentrations and coursework**: Matrix Theory (CS), Mixed Model Methods for Genetic Analysis, Computation and Biology, Genetics, Statistical Genetics and Genomics, Computational Techniques in Genomics

University of Chicago Chicago, IL
B.A. in Economics *Dec 2013*

- **Concentrations and coursework**: Econometrics, Statistical Models and Methods, Analysis in R

COMPUTING SKILLS

Languages and Software: Python, C/C++, Fortran · MKL, ScaLAPACK, OpenMP/MPI, CUDA · R, Matlab, Mathematica, Stata · Linux · Bash, SLURM, PBS · Java, Visual Basic, IDL

EXPERIENCE

University of Minnesota Minneapolis & Saint Paul, MN
Postdoctoral Associate in Quantitative Genetics *Sept 2022 – Present*

- Evaluating and integrating GPU acceleration to improve efficiency of genomic prediction tools
- Processing bovine sequencing data and imputing it to high density
- Integrating complex epistasis effects in GVCHAPmpi (written as part of my thesis) genomic prediction software.
- Conducting genome-wide analysis of pairwise epistasis effects of quantitative traits with samples of over one million individuals
- Conducting comprehensive, large scale haplotype analysis of all traits in USDA Holstein database

University of Minnesota Minneapolis & Saint Paul, MN
Research Assistant in Quantitative Genetics *Sept 2014 – Sept 2022*

- Authored and published GVCHAP - a computing pipeline written in C++, Python, and R to integrate haplotypes with SNP markers for genomic prediction, estimation of variance components, and heritability
- Developed and optimized computing tools and methods for quantitative analysis using SNPs and haplotypes, conducted genomic prediction and variance component estimation analyses, GWAS, and interpreted results
- Published a total of 13 peer-reviewed journal articles and conducted data analysis to secure funding from the NIH, NIFA, and USDA in collaboration with researchers from multiple disciplines on projects involving genomic selection and gene discovery in dairy cattle, swine, wheat, panda and human data

Research Computing Center at the University of Chicago Chicago, IL
Research Assistant in Scientific Computing *Nov 2012 – Aug 2014*

- Optimized and parallelized computational tasks to run efficiently on high-performance computing clusters
- Redesigned applications to process intensive, time-consuming calculations faster with minimal memory usage
- Collaborated with researchers in various disciplines to analyze and compute large data efficiently

Spallation Neutron Source at Oak Ridge National Laboratory Oak Ridge, TN

Physics Programming Intern *June 2008 – Dec 2010*

- Designed application to calculate and graphically display experimental detector data for interpretation/analysis
- Wrote software to filter and catalog experimental data for users of neutron spectrometers
- Developed tools to run spectrometry in collaboration with a team of software engineers and researchers

University of Chicago Chicago, IL

Laboratory Assistant/Programmer *Summer 2005*

- Developed GUI and backend software for on-line synchrotron equipment (precise sample positioning, optical path correction)
- Supported experiments for high-pressure X-ray analysis at the synchrotron

PEER-REVIEWED JOURNAL ARTICLES

- **Prakapenka D**, Liang Z, Zaabza HB, VanRaden PM, Van Tassell CP, Da Y. (2024) A Million-Cow Validation of a Chromosome 14 Region Interacting with All Chromosomes for Fat Percentage in U.S. Holstein Cows. International Journal of Molecular Sciences. 2024; 25(1):674. https://doi.org/10.3390/ijms25010674

- **Prakapenka D**, Liang Z, Da Y (2023) Genome-Wide Association Study of Age at First Calving in U.S. Holstein Cows. Int. J. Mol. Sci. 2023, 24, 7109. doi: 10.3390/ijms24087109

- **Prakapenka D**, Liang Z, Jiang J, Ma L, Da Y. (2021). A Large-Scale Genome-Wide Association Study of Epistasis Effects of Production Traits and Daughter Pregnancy Rate in U.S. Holstein Cattle. Genes. 2021;12(7). doi:10.3390/genes12071089

- **Prakapenka D**, Wang C, Liang Z, Bian C, Tan C and Da Y (2020) GVCHAP: A Computing Pipeline for Genomic Prediction and Variance Component Estimation Using Haplotypes and SNP Markers. Front. Genet. 11:282. doi: 10.3389/fgene.2020.00282

- Liang Z, **Prakapenka D**, Parker Gaddis KL, VandeHaar MJ, Weigel KA, Tempelman RJ, Koltes JE, Santos JEP, Heather White, Francisco Peñagaricano, Ransom L Baldwin, Da, Y. (2022) Impact of epistasis effects on the accuracy of predicting phenotypic values of residual feed intake in U.S. Holstein cows. Front. Genet. doi: 10.3389/fgene.2022.1017490

- Da Y, Liang Z, **Prakapenka D**. (2022). Multifactorial Methods Integrating Haplotype and Epistasis Effects for Genomic Estimation and Prediction of Quantitative Traits. Front. Genet. doi: 10.3389/fgene.2022.922369

- Bian C, **Prakapenka D**, Tan C, Yang R, Zhu D, Guo X, Liu D, Cai G, Li Y, Liang Z, Wu Z, Da Y, Hu X. Haplotype genomic prediction of phenotypic values based on chromosome distance and gene boundaries using low-coverage sequencing in Duroc pigs. *Genet Sel Evol* 53, 78 (2021). doi: 10.1186/s12711-021-00661-y

- Liang, Z., C. Tan, **D. Prakapenka**, L. Ma, and Y. Da. (2020). Haplotype Analysis of Genomic Prediction Using Structural and Functional Genomic Information for Seven Human Phenotypes. Frontiers in Genetics 11(1461).

- Sallam, A. H., E. Conley, **D. Prakapenka**, Y. Da, and J. A. Anderson. (2020). Improving prediction accuracy using multi-allelic haplotype prediction and training population optimization in wheat. G3: Genes, Genomes, Genetics 10(7):2265-2273.

- Ma, L., T. S. Sonstegard, J. B. Cole, C. P. VanTassell, G. R. Wiggans, B. A. Crooker, C. Tan, **D. Prakapenka**, G. E. Liu, and Y. Da. (2019). Genome changes due to artificial selection in U.S. Holstein cattle. BMC genomics 20(1):128.

- Jiang J, Ma L, **Prakapenka D**, VanRaden PM, Cole JB, Da Y. (2019). A large-scale genome-wide association study in U.S. Holstein cattle. Frontiers in Genetics. 2019;10(412).

- Tan,C., Z. Wu, J. Ren, Z. Huang, D. Liu, X. He, **D. Prakapenka**, R. Zhang, N. Li, Y. Da, and X. Hu. (2017). Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing. Genetics Selection Evolution 49:35.

- Garbe, J.R., **D. Prakapenka**, C. Tan, and Y. Da. (2016). Genomic inbreeding and relatedness in wild panda populations. PLoS ONE 11(8): e0160496.

- Wang, C., **D. Prakapenka**, S. Wang, S. Pulugurta, H.B. Runesha, and Y. Da. (2014). GVCBLUP: A computer package for genomic prediction and variance component estimation of additive and dominance effects. BMC Bioinformatics 15:270

ABSTRACTS AND PRESENTATIONS

- **Prakapenka, D.**, Z. Liang, J. Jiang, L. Ma, and Y. Da. 2021a. Genome-wide association study of epistasis effects associated with production and fertility traits in U.S. Holstein cattle. Abstract P171, page 224 of ADSA2021 Abstracts, ADSA 2021 Virtual Annual Meeting.

- **Prakapenka, D.**, Z. Liang, P. M. VanRaden, J. Jiang, L. Ma, J. R. Garbe, C. Melticca, P. J. Hansen, and Y. Da. 2021b. Genetic mechanisms of reproductive heterosis in crossbred dairy cows involve genome-wide additive and nonadditive effects. Abstract P169, page 224 of ADSA2021 Abstracts, ADSA 2021 Virtual Annual Meeting.

- Liang, Z., **D. Prakapenka**, and Y. Da. 2021a. EPIHAP: a computing tool for genomic estimation and prediction using global epistasis effects and haplotype effects. Abstract P167, page 223 of ADSA2021 Abstracts, ADSA 2021 Virtual Annual Meeting.

- Liang, Z., **D. Prakapenka**, and Y. Da. 2021b. Genomic heritability and prediction accuracy of epistasis effects for production and fertility traits in U.S. Holstein cattle. Abstract P170, page 224 of ADSA2021 Abstracts, ADSA 2021 Virtual Annual Meeting.

- Liang, Z., **D. Prakapenka**, P. M. VanRaden, and Y. Da. 2021c. Genomic heritability and prediction accuracy of additive and nonadditive effects for daughter pregnancy rate in crossbred dairy cows. Abstract P168, page 224 of ADSA2021 Abstracts, ADSA 2021 Virtual Annual Meeting.

- Ma, L., J. Jiang, **D. Prakapenka**, J. B. Cole, and Y. Da. 2019. Approximate generalized least squares method for large-scale genome-wide association study. Abstract #M64, American Dairy Science Association Annual Meeting. June 23-26, 2019. Cincinnati.

- **Prakapenka, D.** and Y. Da. 2019. Computing pipeline for genomic prediction and estimation using haplotypes and SNP markers. Abstract #M65, American Dairy Science Association Annual Meeting. June 23-26, 2019. Cincinnati.

- Ma, L., J. Jiang, **D. Prakapenka**, M.E. Tooker, P.M. VanRaden, J.B. Cole, and Y. Da. 2017. Large-scale GWAS reveals reason for the DGAT1 significance and identifies new SNP effects in Holstein cattle. Livestock High-Throughput Phenotyping and Big Data Analytics, Beltsville, MD. Nov. 13-14.

- **Prakapenka, D.**, L. Ma, and Y. Da. 2017. Chromosome-specific genomic relationships using haplotypes for genomic prediction and variance component estimation. Plant and Animal Genome XXV, San Diego, CA. Jan. 13-18.

- Tan, C., **D. Prakapenka**, L. Ma, Z. Wu, X. Hu, and Y. Da. 2017. JBLUP: The joint best linear unbiased prediction using BLUP and GBLUP solutions. Plant and Animal Genome XXV, San Diego, CA. Jan. 13-18.

- Da, Y., C. Tan, and **D. Prakapenka**. 2016. Integrated SNP-haplotype genomic selection based on the invariance property of GBLUP and GREML to duplicate SNPs. 2016 Joint Annual Mtg. ASAS-ADSA-CSAS-WSASAS, Salt Lake City, UT. July 19-23.

- Garbe, J.R., **D. Prakapenka**, J. Yang, C. Tan, C. Wang, and Y. Da. 2016. Genomic inbreeding and relationships in wild panda populations. Plant and Animal Genome XXIV Conf., San Diego, CA. Jan. 9-13. Abstr. P0694.

- Tan, C., **D. Prakapenka**, C. Wang, L. Ma, J.R. Garbe, and Y. Da. 2015. Integration of haplotype analysis of functional genomic information with single SNP analysis improved accuracy of genomic prediction. ADSA/ASAS 2015, Orlando, FL, July 12-16. Abstr. M84.

- Da, Y., C. Wang, C. Tan, **D. Prakapenka**, M. Shigematsu, J.R. Garbe, and L. Ma. 2015. Multi-allelic haplotype model to integrate functional genomic information with genomic prediction and estimation. Plant and Anim. Genome XXIII, San Diego, CA. Jan. 10-14.