Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2582-2845.8702

ISSN: 2582 – 2845

Ind. J. Pure App. Biosci. (2021) *9*(3), 237-246

Indian Journal of Pure & Applied Biosciences

Peer-Reviewed, Refereed, Open Access Journal

Review Article

Application and Potential Use of Advanced Bioinformatics Techniques in Agriculture and Animal Sciences

Muhammad Nouman Khalid[1], Ahsan Abdullah[2], Zainab Ijaz[3], Nargis Naheed[4], Ahmad Hamad Sheir[5], Muhammad Ahmad[6], Faisal Shabir[7*], Khalida Parveen[4], Muhammad Dilawaiz Khan[7]

[1]Department of Plant Breeding and Genetics, University of Agriculture Faisalabad
[2]Department of Plant Pathology, College of Plant Protection,
China Agricultural university, 100193, Beijing, China
[3]Joint International Research Laboratory of Agriculture and Agri-Product Safety,
College of Agriculture, Yangzhou University, China
[4]Department of Zoology, Wildlife and Fisheries University of Agriculture, Faisalabad-Pakistan
[5]Department Poultry Science, University of Agriculture Faisalabad
[6]Department of Horticulture, College of vegetables Science,
China Agricultural University, 100193, Beijing, China.
[7]Department of agronomy, University of Agriculture Faisalabad
*Corresponding Author E-mail: faisal.shabir3612@gmail.com

Received: 10.04.2021 | Revised: 19.05.2021 | Accepted: 27.05.2021

ABSTRACT

The amount of biological information generated in the last two decades is enormous because of Next generation sequencing (NGS) discovery that has enabled researches to sequence and model almost every organism and also due to rapid advancements in techniques and tools in experimental research. The research which was first carried out at fields, labs and clinics is now started with computational analysis (in-silico) of information, modeling, experiment planning and hypothesis development. Various applications of bioinformatics are algorithms, databases, and other data analysis tools and softwares that enable storage, analysis, retrieval, annotation and visual interpretation of biological information which in turn increases the knowledge of various biological systems that help in making new discoveries regarding production, human health, animal health and plant health keeping in mind the challenges of climate change, water and area shortage. This will help not only in increased plant and animal production but also in management and treatment of various human, animal and plant diseases in addition to the underlying mechanisms and strategies of the rapidly evolving pathogenic microorganism and antibiotic resistance.

Keywords: *Animal health, Plant development, Human health, Crop improvement, Genomics, Computational biology, Vaccines, Animal/human diseases, Plant breeding and genetics, Computer aided tools/softwares.*

Cite this article: Khalid, M.N., Abdullah, A., Ijaz, Z., Naheed, N., Sheir, A.H., Ahmad, M., Shabir, F., Parveen, K., Khan, M.D. (2021). Application and Potential Use of Advanced Bioinformatics Techniques in Agriculture and Animal Sciences, *Ind. J. Pure App. Biosci.* 9(3), 237-246. doi: http://dx.doi.org/10.18782/2582-2845.8702

INTRODUCTION

Recent advances in the fields of molecular biology especially recombinant DNA technology, genomics, transcriptomics, metabolomics and proteomics which are sometimes also referred to as ‘omics’ has led to enormous amount of data generation related to biological information that needs to be processed, managed, analyzed and stored. For this purpose the field of bioinformatics plays an important role that is generally defined as the use of computer technology for processing and management of biological information. Bioinformatics in multidisciplinary approach involving biology, biostatistics, computational biology and mathematics/physics not only store and manage the information but also extracts useful biological significance from it; which has wide range of applications in various fields like human health, animal health and agriculture.

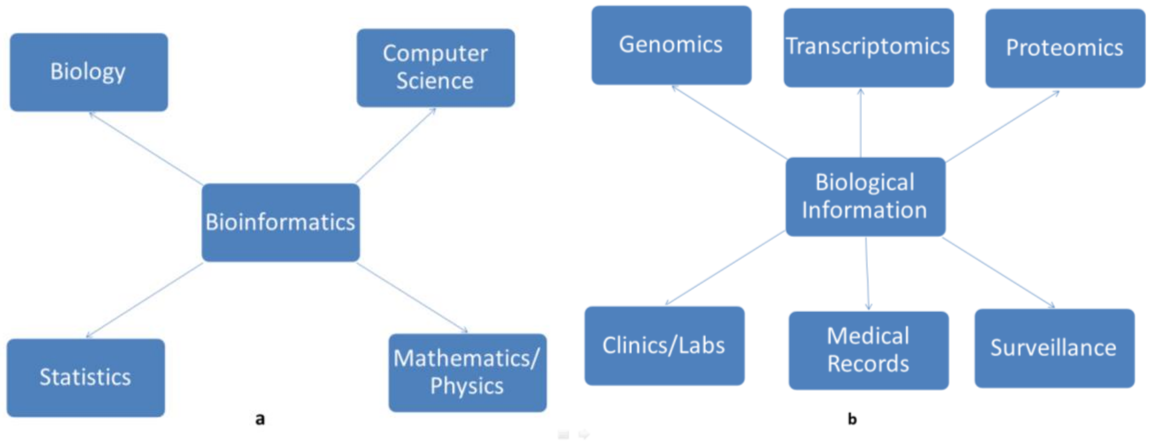

Figure 1: (a) Different disciplines that are used by Bioinformatics (b) Sources of Biological Information

The two main branches of bioinformatics are animal bioinformatics; that is further subdivided into mammals, birds, fishes and reptile bioinformatics, and plant bioinformatics that is further subdivided into agriculture bioinformatics also known as crop bioinformatics, horticulture, medicinal plant and forest plant bioinformatics.

Applications of Bioinformatics in Agriculture:

Due to the rapid increase in human population challenges to breeders have also increased to enhance crop production (Tilman et al., 2011) and yield as human are dependent on 2/3 of their daily calories intake (Smit et al., 1999; & Ulijaszek, 1991) and agriculture also faces challenges like shortage of both water and land, and climate changes (Godfray et al., 2010). To overcome these needs and challenges an interdisciplinary approach (Moose & Mumm, 2008a,b) involving genomics, bioinformatics and other new glasshouse technologies (Watson et al., 2018) is required to be adapted to develop improved cultivars with increased plant development and in turn increase crop production. Besides the large amount of agricultural data also requires processing and analyzing to gain deep functional insights into plant genomes and obtain desired breeding outcomes (Batley & Edwards 2016 a, b; & Santos et al., 2017). Bioinformatics downstream analysis like comparative genome analysis, variant calling and GWAS can facilitate crop improvement (Ong et al., 2016) while adaption of existing databases like Grain Genes (Matthews et al., 2003) and Gramene (Ware et al., 2020), and development of new databases like WheatIS (Wheat Information System) (Scheben et al., 2018) will help to store the large amount of data generated related to agriculture and make it available for breeders globally.

Third generation sequencing:

With the advent of NGS, the discipline of plant biology has been revolutionized with genome sequencing (Goodwin et al., 2016) which has allowed researchers to develop more standard genome assemblies and sequence even the complex regions with high repetitive sequences and different variants using different reads. Genome annotation has also been improved with efficient study of exons and spliced regions by full length transcripts that have been sequenced with third generation sequencing (Li et al., 2018) which in turn has enabled the researchers to identify and characterize various genes not only structurally and functionally but also their relation to different agronomical traits has been efficiently elucidated.

Integrated Crop Databases:

New and efficient third generation sequencing technologies are generating a lot of data that can be used to investigate different characteristics of agricultural plants from gene level to population level (Pareek et al., 2011). Certain databases store and manage information without integrating phenotypic or variant information like European Molecular Biological Laboratory (EMBL) (Stoesser et al., 2001), Genbank (Benson et al., 2009), Phytozome (Goodstein et al., 2012) and PlantGBD (Duvick et al., 2007) while other tools like KnetMiner which is an intelligent web based mining tool and can detect novel connection between various traits and linked genes (Hassani-Pak & Rawlings 2017a, b) by searching biological information for possible links and approaches. KnetMiner adapts a four step approach; creating a graph from the huge biological information, searching through the literature to improve the graph, linking different evident nodes to genes and finally an intelligent algorithm for ranking genes through application of evidences visualize the integrated information. So far KnetMiner has been utilized for constructing various databases for agronomical important crops like barley and wheat which provides deep understanding of biological process and its association with distant traits.

Varietal information system:

When a genotype is released for commercial cultivation then it is called variety. Varietal information system is important for modeling of agricultural plants using ideal traits which is accomplished through the applications of computer aided programs and tools followed by application of breeding techniques to grow it in field (Gupta et al., 2020).

Pedigree analysis:

Pedigree analysis is important for sorting out parents that are then used in hybridization of various crop plants through breeding techniques. Through application of various bioinformatics tools, pedigree analysis sort out the parents which are common in different cultivars and hybrids that are used to develop hybrid crops through breeding (Gupta et al., 2020).

Biometrical analysis:

For achieving specific targets in plant breeding programs, proper planning and management is necessary that is acquired through biometric analysis and carried out through computer programs. Various kinds of biometric analysis that are carried out in plant breeding and genetics consists of allel like; partial di, di, tri and quadri, path coefficient, correlation, triple test cross, discriminant function, stability parameters, generation means, metrogyl path, line x tester and D2 statistics analysis (Gupta et al., 2020).

Forecasting models:

Remote sensing techniques based on computer program are used to predict various biological outcomes related to crop production, incidence of diseases and insects in crop plants and these predictions are mainly based on different weather parameters (Gupta et al., 2020).

Plant disease management:

The advent of whole genome sequencing with bioinformatics tools have created a better understanding of the plant-pathogen interaction and mechanism of various diseases which has led to creation of various genetically modified crops that are resistant to pests and diseases. Various agriculturally important organisms have been sequenced and studied in different genetic plant models (Koltai & Volpin 2003 a, b).

Insect Resistance

Insect resistant crops have developed due to advances in the field of sequencing and crop breeding. For example the genes of bacteria *Bacillus thuringiensis* have been inserted in many plants like corn, cotton and potatoes to make it insect resistant which was possible only because of sequencing of the bacterial genome and identification of genes that increases soil fertility and also provide protection to plants against insects. The development of insect resistant plants has in turn reduced the dependency on insecticides and also their productivity and nutritional value has been enhanced.

Nutritional Quality

The nutritional quality of different crops has also been improved with due to the mutating its genome which has been made possible through the use of bioinformatics tools like Golden rice which has been genetically modified to have high quantity of vitamin A that is essential for eyes.

Identification of breeding targets by Genome-Wide Association Studies (GWAS):

Bioinformatics tools that are based on standard regression analysis like PLINK for carrying GWAS studies are less sensitive when used for variants (Ma et al., 2013) in associating genotypes with phenotypes (Purcell et al., 2007). Other tools that are based on mixed linear model like TASSEL (Bradbury et al., 2007) are desirable for studying population effects because they incorporates both family and population structure in the analysis. Another more advance tool GAPIT that is also based on mixed linear model uses selection method and model based predictions is ideal for handling large data sets that involves over 10k specimens and over 1 million SNPs (Lipka et al., 2012).

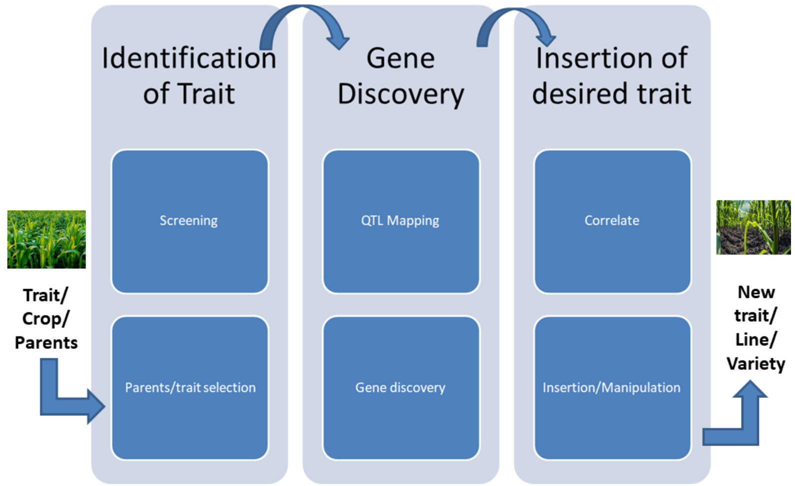

Figure 2: Outline of typical trait discovery/Insertion in Plants

Applying Machine Learning to Crop Breeding:

For analysis of large datasets that contains diverse and raw information machine learning (ML) offers a significant advantage over other analytical approaches because machine learning is based on algorithms that utilizes learning patterns for interpretation of data (Libbrecht & Nobel 2015a, b). The association between various traits and its underlying molecular mechanisms are studied through ML that allows the breeders to efficiently mine large datasets and phenotype plants.

Designing of gRNAs for CRISPR/CAS:

CRISPR/CAS technique for editing genes is widely used method in rDNA technology and many applications in genomics studies. This gRNAs that are used in this technique needs to be properly designed that is carried out through various bioinformatics tools and is necessary for efficient CRISPR/CAS gene editing (Hu et al., 2018).

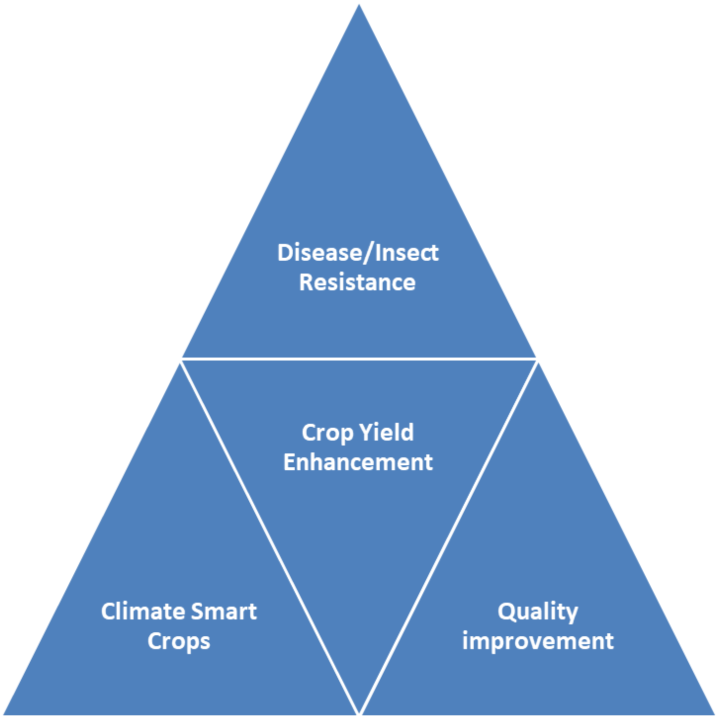

Figure 3: Crispr/Cas applications in Plants

Mining Quantitative Trait Loci Studies (QTL):

The detection of genetic regions that are responsible for quantitative traits is carried out by QTL analysis which in turn has reduced the gap between field and lab. For linking agronomic trait to a gene or genomic region QTL is considered a powerful technique. Various bioinformatics tools like MetaQTL that uses a consensus model and biostatistics can efficiently detect QTL loci (Kumasaka et al., 2016) and its effect by reducing the QTL confidence interval level in mapping of different crop development related traits and biotic and abiotic stress responses in important crop plants like wheat (Hanocq et al., 2007), cotton (Said et al., 2013), maize (Liu et al., 2012) and soya bean (Qi et al., 2011). Rare alleles in whole crop genome can even be identified with limited number of genetic markers with QTL.

Table 1: Bioinformatics in Plant Science

Applications	**Examples/Uses**	**Sources/Links/Tools**
Model Plants	Arabidopsis thaliana, Tomato, Wheat, Rice, Maize	https://www.arabidopsis.org/
Management of plant genome data	Primary and integrated Databases for different plants like barley and wheat	https://www.ncbi.nlm.nih.gov/ https://www.embl.de/ https://www.ddbj.nig.ac.jp/index-e.html NCBI, EMBL, DDBJ, SWISS-PROT, TrEMBL, GrainGenes, Gramene, WheatIS, KnetMiner
Comparing plant Genomes	Sequence Alignment, Multiple sequence alignment, Sequence Similarity Searching, Genome Comparison, pedigree analysis	ClustalW, https://blast.ncbi.nlm.nih.gov/Blast.cgi https://www.ebi.ac.uk/Tools/sss/fasta/ MegaBlast, BLAT, BLASTZ
Expressed sequence tags (ESTs)	Discovery of gene, polymorphism analysis, mapping, expression studies	Different EST databases, dbEST
Designing various tools for molecular Plant Breeding	Hybrid and new crop varieties, disease and insect resistance, increased yield, improved nutritional quality	gRNAS for CRISPR/CAS
Sequencing	Sequencing genomes, identification and analysis of genes	Third generation sequencing, Next generation sequencing, Illumina sequencing, SMRT[a], MPSS[b]
Genome-Wide Association Studies	Study variants in a population	PLINK, TASSEL, GAPIT
Quantitative Trait Loci (QTL)	Identification of genes responsible for quantitative traits and abiotic and biotic stresses in wheat, soya bean, cotton and maize	MetaQTL,

a= Single molecule real time, b= massively parallel signature sequencing

Applications of bioinformatics in zoology:

Due to a rapid increase in population the global population will cross 10 billion in 2050 which will increase demand for food and in turn animals that are important source of food and fiber for humans and currently accounts for 18% total calories and 39% protein content (Rexroad et al., 2019). Advance animal breeding techniques for increased animal production requires a multidisciplinary approach that will not only ensure efficient output by combatting diseases but also adapt to minimizing the use of antibiotics in this era of antimicrobial resistance, climate changes and reduce global warming by reducing emission of greenhouse gases (Rexroad et al., 2019).

The information in human healthcare comes at a rapid face from variety of data sources including electronic medical records, lab databases, surveillance systems and other platforms. The top two research areas that were indexed in Google scholar were bioinformatics and health informatics (Andreu-Perez et al., 2015). In similar fashion information regarding animal health care also gathers from numerous data sources like animal species specific databases, electronic veterinary medical records, surveillance for diseases, farm production records, social media platforms, wearables and other sensors for pets, horses (Thompson et al., 2018) and production animals (Haladjian et al., 2018).

Veterinary epidemiology:

The focus in veterinary epidemiology like other sciences has shifted from data acquisition to extracting meaningful information regarding animal health. Veterinary epidemiological data including ‘’omics data’’ has become ‘’big data’’ which can be utilized in a number of ways to analyze health risks and minimize its adverse effect. Various examples of such analysis include monitoring health trends and emerging health threats in animals interpreting high velocity information, increased efficiency in surveillance system identifying populations that are at high risk through analysis of spatial and animal movement information (VanderWaal et al., 2017).

Clinical bioinformatics:

Numerous bioinformatics tools and databases are available that provide in depth analysis of disease outbreaks and pathology, variations in genome, evolutionary analysis of host and pathogen and identification of diagnostic markers by dissecting information related to pathophysiology of pathogens. These tools have received spatial attention in the management of various deadly diseases like tuberculosis, malaria, dengue fever and filariasis. The Cancer Genome Atlas (TCGA) database was mined by (Krasnov et al., 2019) for developing reference genes which have stable mRNA levels and can be utilized for real time PCR analysis of tumor cells. This allowed researchers to study previously untried genes targets that are specific for each cancer type and other pan-cancer universal gene targets like SFRS4, SF3A1 and CIAO1 (Fedorova et al., 2019). The sequencing of cell-free DNA (cfDNA) in human plasma has identified new biomarkers that depend on epigenetic signatures which are tissue specific due to nucleosome patterns and can be used for designing of NGS panels that are based on amplicon which has improved the sensitivity in detection of mutations and has also minimized the requirement of body fluids from patients.

Vaccine design:

Recent advances in bioinformatics have made it easy to design vaccines and drugs easily which have better immunological response and properties (Paladino et al., 2017). Rational vaccinology is a modern technique that is based on synthetic peptides and has been used to design vaccines against asthma (Bramwell & Perrie 2005a, b). Antigen prediction plays a critical role in vaccine design which can be carried out through an online tool Vaxi Jen that is the first of its kind which utilizes alignment free approach for efficient antigen detection (Gol et al., 2018). Lua et al. (2014) reported the potential applications of rational vaccinology in designing vaccines against different pathogenic viruses like HIV, HCV and influenza virus. Another approach for rapid and effective vaccine design is reverse vaccinology (Dalsass et al., 2019) that is based on accurate epitope detection which in turn generates vigorous response from B and T cells and for specific prediction in silico techniques are utilized (Araújo et al., 2019; & Pyclik et al., 2018). This technique has been used to design multi epitope vaccine against brucellosis. Likewise high IgG antibodies production was reported in mice using multi epitope peptide vaccine for immunological response designed through advanced bioinformatics tools (Ren et al., 2019).

Broadly neutralizing antibodies (bNAbs) have the characteristics to only target specific epitopes that play a crucial rule in pathogenicity (Moore, 2018) however this is a new approach in the bioinformatics pipeline as it is a new term introduced recently to immunoinformatics (He et al., 2017) and can efficiently design vaccines against rapidly mutating viruses like HIV and influenza (Ekiert et al., 2012).

Proper design of vaccines requires proper implementation of bioinformatics tools and techniques which in turn minimizes the challenges in vaccinology. These techniques and tools are equally applicable to veterinary and have wide ranges of application in designing vaccine against deadly diseases in livestock, poultry, fishes, pets and other animals.

CONCLUSION

Agriculture and animals both face significant challenges in meeting the demands of rapid growing world population. Besides food insecurity quality human, animal and plant health is another major challenge in terms of antibiotic resistance, evolving and mutating viral pathogens and other pathogenic microorganisms causing deadly diseases and causes significant loss to human lives, food and animals. To overcome these challenges and meet the food supply demands multidisciplinary approach that involves breeding, genomics, transcriptomics, proteomics, metabolomics and bioinformatics is required to increase crop production and animal production. With the advent of new techniques and approaches now a wide range of data is available which is stored, trimmed, analyzed and integrated through various online softwares and tools. By applying novel technologies and analytical approaches researchers can ensure food security with increased crop yield, animal yield, disease resistance, insect resistance, pesticide resistance, abiotic and biotic stress tolerant and other desirable characteristics in both plants and animals.

REFERENCES

Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T., & Yang, G. Z. (2015). Big data for health. *IEEE journal of biomedical and health informatics*, *19*(4), 1193-1208.

Araújo, C. L., Alves, J., Nogueira, W., Pereira, L. C., Gomide, A. C., Ramos, R., & Folador, A. (2019). Prediction of new vaccine targets in the core genome of Corynebacterium pseudotuberculosis through omics approaches and reverse vaccinology. *Gene, 702*, 36-45.

Batley, J., & Edwards, D. (2016). The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. *Current opinion in plant biology, 30*, 78-81.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009). GenBank. *Nucleic acids research*, *37*(suppl_1), D26-D31.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635.

Bramwell, V. W., & Perrie, Y. (2005). The rational design of vaccines. *Drug discovery today*, *10*(22), 1527-1534.

Dalsass, M., Brozzi, A., Medini, D., & Rappuoli, R. (2019). Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery. *Frontiers in immunology, 10*, 113.

Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M. D., Lawrence, C. J., & Brendel, V. (2007). PlantGDB: a resource for comparative plant genomics. *Nucleic acids research, 36*(suppl_1), D959-D965.

Ekiert, D. C., & Wilson, I. A. (2012). Broadly neutralizing antibodies against influenza virus and prospects for universal therapies. *Current opinion in virology, 2*(2), 134-141.

Fedorova, M. S., Krasnov, G. S., Lukyanova, E. N., Zaretsky, A. R., Dmitriev, A. A., Melnikova, N. V., & Kudryavtseva, A. V. (2019). The CIMP-high phenotype is associated with energy metabolism alterations in colon adenocarcinoma. *BMC Medical genetics, 20*(1), 15-27.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *science, 327*(5967), 812-818.

Gol, S., Pena, R. N., Rothschild, M. F., Tor, M., & Estany, J. (2018). A polymorphism in the fatty acid desaturase-2 gene is associated with the arachidonic acid metabolism in pigs. *Scientific reports, 8*(1), 1-9.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research, 40*(D1), D1178-D1186.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*(6), 333.

Gupta, R., Tiwari, S., Tripathi, M. K., & Pooniya, S. K. (2020). Chapter-8 Bioinformatics and Its Applications in Crop Improvement. *Recent Trends in Molecular Biology and Biotechnology, 129*.

Haladjian, J., Haug, J., Nüske, S., & Bruegge, B. (2018). A wearable sensor system for lameness detection in dairy cattle. *Multimodal Technologies and Interaction, 2*(2), 27.

Hanocq, E., Laperche, A., Jaminon, O., Lainé, A. L., & Le Gouis, J. (2007). Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. *Theoretical and Applied Genetics, 114*(3), 569-584.

Hassani-Pak, K., & Rawlings, C. (2017). Knowledge discovery in biological databases for revealing candidate genes linked to complex phenotypes. *Journal of integrative bioinformatics, 14*(1).

He, L., Lin, X., de Val, N., Saye-Francisco, K. L., Mann, C. J., Augst, R., & Zhu, J. (2017). Hidden lineage complexity of glycan-dependent HIV-1 broadly neutralizing antibodies uncovered by digital panning and native-like gp140 trimer. *Frontiers in immunology, 8*, 1025.

Hu, H., Scheben, A., & Edwards, D. (2018). Advances in integrating genomics and bioinformatics in the plant breeding pipeline. *Agriculture, 8*(6), 75.

Koltai, H., & Volpin, H. (2003). Agricultural genomics: An approach to plant protection. *European journal of plant pathology, 109*(2), 101-108.

Krasnov, G. S., Kudryavtseva, A. V., Snezhkina, A. V., Lakunina, V. A., Beniaminov, A. D., Melnikova, N. V., & Dmitriev, A. A. (2019). Pan-cancer analysis of TCGA data revealed promising reference genes for qPCR normalization. *Frontiers in genetics, 10*, 97.

Kumasaka, N., Knights, A. J., & Gaffney, D. J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature genetics, 48*(2), 206-213.

Li, C., Lin, F., An, D., Wang, W., & Huang, R. (2018). Genome sequencing and assembly by long reads in plants. *Genes, 9*(1), 6.

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics, 16*(6), 321-332.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics, 28*(18), 2397-2399.

Liu, R., Zhang, H., Zhao, P., Zhang, Z., Liang, W., Tian, Z., & Zheng, Y. (2012). Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. *Plant Molecular Biology Reporter, 30*(2), 297-308.

Lua, L. H., Connors, N. K., Sainsbury, F., Chuan, Y. P., Wibowo, N., & Middelberg, A. P. (2014). Bioengineering virus-like particles as vaccines. *Biotechnology and bioengineering, 111*(3), 425-440.

Ma, C., Blackwell, T., Boehnke, M., Scott, L. J., & GoT2D Investigators. (2013). Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genetic epidemiology, 37*(6), 539-550.

Matthews, D. E., Carollo, V. L., Lazo, G. R., & Anderson, O. D. (2003). GrainGenes, the genome database for small-grain crops. *Nucleic acids research, 31*(1), 183-186.

Moore, P. L. (2018). The neutralizing antibody response to the HIV-1 Env protein. *Current HIV research, 16*(1), 21-28.

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant physiology, 147*(3), 969-977.

Ong, Q., Nguyen, P., Phuong Thao, N., & Le, L. (2016). Bioinformatics approach in plant genomic research. *Current genomics, 17*(4), 368-378.

Paladino, A., Marchetti, F., Rinaldi, S., & Colombo, G. (2017). Protein design: from computer models to artificial intelligence. Wiley Interdisciplinary Reviews: *Computational Molecular Science, 7*(5), e1318.

Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of applied genetics, 52*(4), 413-435.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics, 81*(3), 559-575.

Pyclik, M., Górska, S., Brzozowska, E., Dobrut, A., Ciekot, J., Gamian, A., & Brzychczy-Włoch, M. (2018). Epitope mapping of Streptococcus Agalactiae elongation factor tu protein recognized by human sera. *Frontiers in microbiology, 9*, 125.

Qi, Z. M., Wu, Q., Han, X., Sun, Y. N., Du, X. Y., Liu, C. Y., & Chen, Q. S. (2011). Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica, 179*(3), 499-514.

Ren, S., Guan, L., Dong, Y., Wang, C., Feng, L., & Xie, Y. (2019). Design and evaluation of a multi-epitope assembly peptide vaccine against Acinetobacter baumannii infection in mice. *Swiss medical weekly, 149*(2324).

Rexroad, C., Vallet, J., Matukumalli, L. K., Reecy, J., Bickhart, D., Blackburn, H., & Wells, K. (2019). Genome to phenome: improving animal health, production, and well-being–a new USDA blueprint for animal genome research 2018–2027. *Frontiers in genetics, 10*, 327.

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC genomics*, *14*(1), 1-22.

Santos, R., Algar, A., Field, R., & Mayes, S. (2017). Integrating GIScience and Crop Science datasets: a study involving genetic, geographic and environmental data (No. e2248v4). PeerJ Preprints.

Scheben, A., Batley, J., & Edwards, D. (2018). Revolution in genotyping platforms for crop improvement. In Plant Genetics and Molecular Biology (pp. 37-52). Springer, Cham.

Smit, E., NIETO, F. J., Crespo, C. J., & Mitchell, P. (1999). Estimates of animal and plant protein intake in US adults: results from the Third National Health and Nutrition Examination Survey, 1988–1991. *Journal of the American Dietetic Association*, *99*(7), 813-820.

Stoesser, G., Baker, W., van den Broek, A., Camon, E., Garcia-Pastor, M., Kanz, C., & Tuli, M. A. (2001). The EMBL nucleotide sequence database. *Nucleic acids research, 29*(1), 17-21.

Thompson, C. J., Luck, L. M., Keshwani, J., Pitla, S. K., & Karr, L. K. (2018). Location on the body of a wearable accelerometer affects accuracy of data for identifying equine gaits. *Journal of equine veterinary science, 63*, 1-7.

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the national academy of sciences, 108*(50), 20260-20264.

Ulijaszek, S. J. (1991). Human dietary change. Philos. *Trans. R. Soc. Lond. B Biol. Sci. 334*, 271–279.

VanderWaal, K., Morrison, R. B., Neuhauser, C., Vilalta, C., & Perez, A. M. (2017). Translating big data into smart data for veterinary epidemiology. *Frontiers in veterinary science, 4*, 110.

Ware, D., Naithani, S., & Tello-Ruiz, M. K. (2020, January). Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities. In Plant and Animal Genome XXVIII Conference (January 11-15, 2020). PAG.

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., & Hickey, L. T. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature plants, 4*(1), 23-29.