Universe
Scientific Publishing

Intelligent Data Simulation And Predictive Analysis In Soybean Breeding

Donghui Zhang, Qingchun Yang, Zhen Geng, Wentao Shu, Jinhua Li

Zhoukou Academy of Agricultural Sciences, Zhoukou 466001, China.

Abstract: The application of intelligent data simulation and predictive analysis shows great potential in improving the efficiency and effectiveness of soybean breeding programs. Based on the research of intelligent data simulation and predictive analysis, the information of complex genetic and environmental factors affecting soybean traits was analyzed, and the application of advanced data simulation technology and predictive analysis method in soybean breeding was discussed. The findings not only contribute to the scientific understanding of soybean breeding, but also provide practical recommendations for harnessing the power of data-driven approaches in agricultural research and breeding programs.

Keywords: Soybean Breeding; Intelligent Data; Simulation Prediction Analysis

1. Introduction

Soybean breeding is a complex and critical process, the purpose of which is to improve the genetic traits of soybean plants. It involves selecting and breeding soybean varieties with desirable characteristics, such as high yield, disease resistance and adaptability to different environmental conditions. The goal of soybean breeding is to improve the overall productivity and quality of soybean crops, ultimately contributing to global food security and sustainable agricultural practices.

Traditional soybean breeding methods rely heavily on observational and experimental methods, which are time-consuming and limited in scope. However, with the advent of intelligent data simulation and predictive analytics, the soybean breeding landscape has changed. These advanced technologies provide powerful means for analyzing complex genetic and environmental interactions, enabling breeders to make more informed decisions and speed up the breeding process.

2. Intelligent data simulation technology and predictive analysis method

2.1 Overview of intelligent data simulation technology

2.1.1 Genome simulation model

Genome simulation models play a key role in soybean breeding by integrating genomic data to simulate genetic variation and its influence on phenotypic traits [1]. By integrating different genomic data sources, such as single nucleotide polymorphisms (SNPS), copy number variants (CNVs), and gene expression profiles, genome simulation models enable researchers to predict phenotypic outcomes of different genetic combinations [2].

The integration of genomic simulation models in soybean breeding helps to identify key genomic regions associated with desirable traits, such as yield, disease resistance, and nutritional quality [3]. In addition, the predictive power of genomic simulation models enables breeders to prioritize promising genetic combinations for further experimental validation, simplifying the breeding process and accelerating the development of improved soybean varieties. [4]

In practical applications, genome simulation models are used to simulate breeding populations, predict the performance of candidate varieties under different environmental conditions, and optimize breeding strategies to achieve the improvement of specific traits. [5]

2.1.2 Phenotypic simulation methods

Phenotypic modeling methods in soybean breeding include a variety of methods to simulate the expression of phenotypic traits under different environmental conditions. These methods use environmental data, including climatic factors, soil properties, and management practices, to model the phenotypic response of soybean varieties in different growing environments[6].

One of the main advantages of phenotypic modeling methods is the ability to assess the stability and adaptability of soybean varieties in different geographical regions and environmental conditions[7]. In addition, phenotypic simulation methods contribute to the optimization of breeding strategies by providing breeders with a comprehensive assessment of variety performance in multiple environments, thus providing information for the targeted deployment of varieties in specific agro-ecological zones [8].

In practical applications, phenotypic modeling methods are used to simulate the performance of soybean varieties under climate change scenarios, to assess the impact of different management practices on crop performance, and to inform the development of varieties suitable for specific environmental niches[9].

2.1.3 Binding genotype-phenotype interactions

Incorporating genotype-phenotype interactions in intelligent data simulation techniques is a key aspect of advancing soybean breeding practices. These interactions capture the complex relationship between genetic variation and its phenotypic expression, providing a comprehensive understanding of how genetic factors influence soybean traits. By utilizing multiple data sources, including genomic, phenotypic, and environmental data, the integration of genotype-phenotypic interactions enables researchers to model the multifaceted effects of genetic variation on trait expression under dynamic environmental conditions [10].

2.2 Overview of predictive analysis methods

2.2.1 Random Forest algorithm

One of the most widely used machine learning algorithms in predictive analysis of soybean breeding is the random forest algorithm. This ensemble learning approach builds multiple decision trees and combines their predictions to produce a robust and accurate prediction model. In soybean breeding, random forest algorithms perform well in processing high-dimensional genomic data and capturing complex genetic interactions that affect soybean traits. Random forest algorithm utilizes a large number of decision trees and summarizes their prediction results, effectively reducing overfitting and improving the prediction accuracy of various soybean traits [11].

2.2.2 Support Vector Machine (SVM)

Support vector machines are also widely used in predictive analysis of soybean breeding. Support vector machines are particularly suited to modeling relationships between genetic markers and important agronomic traits, enabling the identification of non-linear patterns and associations in high-dimensional genomic data. The ability of SVM to handle nonlinear relationships and adapt to large-scale data sets makes it a valuable tool for predicting complex soybean traits and guiding breeding decisions. Using the predictive ability of support vector machines, soybean breeders can gain a deeper understanding of the genetic basis of trait variation, thereby improving the effectiveness of breeding strategies [12].

3. Suggestions and conclusions

3.1 Suggestions

The practical applications of intelligent data simulation and predictive analytics in soybean breeding are multifaceted and provide valuable insights for breeders and researchers. First, integrating a variety of data sources, including genomic, phenotypic and environmental data, enables breeders to make more accurate and robust analyses with comprehensive datasets. This integration promotes a holistic understanding of the complex interactions that shape soybean traits, laying the foundation for targeted breeding interventions. Second, combining predictive analysis methods with data simulation can produce actionable predictions of trait performance under different environmental conditions. This foresight has proven useful in guiding breeding decisions and prioritizing trait optimization for specific agro-ecological environments.

Based on the research results and practical application of intelligent data simulation and predictive analysis, some suggestions are put forward to promote the development of soybean breeding. First, there needs to be a continued emphasis on data integration and interoperability to simplify the utilization of various data sources. This involves the development of standardized data formats and interoperable platforms to facilitate seamless data sharing and analysis between different breeding programs and research programs. In addition, validation and improvement of predictive models should be prioritized to improve the reliability and accuracy of predictive analysis. Collaborative efforts on

model validation and improvement contribute to the establishment of best practices in predictive modeling for soybean breeding. In addition, developing user-friendly platforms and tools for breeders to access and interpret predictive analytics is critical to translating research insights into actual breeding decisions.

3.2 Conclusion

Taken together, the integration of intelligent data simulation and predictive analytics shows great promise in transforming soybean breeding by providing valuable insights, informing breeding decisions, and optimizing trait performance. By adopting the above recommendations and promoting collaborative research, the soybean breeding community can harness the power of data-driven approaches to propel breeding practices into a new era of efficiency and impact.

References

[1] Zeng Shunan, Jia Shihao, Cao Yongce. Genome-wide association analysis of plant height and main stem number of soybean [J]. Intelligent Agriculture Guide, 2019,3(19) : 34-38.

[2] Yin Zhengong, Wang Qiang, Meng Xianxin, Liu Guangyang, Guo Yifan, Wang Xiujun. Candidate genes mining for plant height traits in soybean based on Overview and physical map [J]. Soybean Science. 2019,38(06) : 914-920.

[3] Ren Jiaojiao, Wu Penghao. The innovative application of biological breeding technology in the transformation of traditional seed industry [J]. Molecular Plant Breeding, 2019,21(19) : 6483-6487.

[4] Zhang W. Accelerate the R&D and application of biological breeding to promote agricultural science and technology self-reliance [J]. Journal of Agricultural Science and Technology of China. 2022,24(12) : 8-14.

[5] Zhang J. The prospect of industrialization application of biological breeding for important crops in China [J]. Journal of Agricultural Science and Technology of China. 2022,24(12) : 15-24.

[6] Hu Jiang, Qian Qian. Current situation and prospect of crop biological breeding technology [J]. China Basic Science, 2019,24(06) : 1-8.

[7] Guo Jingli, Zhang Yuhong, Sheng Caijiao. Countermeasures and suggestions for promoting the industrialization application of biological breeding in China in an orderly manner [J]. Journal of Agricultural Science and Technology of China, 2019,25(12) : 1-5.

[8] Cui Ningbo, Liu Wang. Social welfare prediction of industrialization of transgenic herbicide-resistant soybeans in China: Based on DREAM model [J]. Jiangsu Agricultural Sciences, 2018,46(13) : 304-307.

[9] Wang Youhua, Cai Jingjing, Yang Ming, Zhang Tian, Ren Hongmei, Zou Wannong, Sun Guoqing. Patent information analysis and technology prospect of global transgenic soybean [J]. Chinese Journal of Bioengineering, 2018,38(02) : 116-125.

[10] Shen Ping, Wu Yuhua, Liang Jinguang, Lu Xin, Zhang Qiuyan, Wang Haoqian, Liu Pengcheng. Overview of the development and application of transgenic crops [J]. Chinese Journal of Bioengineering, 2017,37(01) : 119-128.

[11] Fan Shengxu, Yang Chunxi, Yang Qiliang, Han Shichang. Leaf area growth prediction model of Panax Notoginseng based on particle swarm optimization and random forest algorithm and meteorological data [J]. Chinese Herbal Medicine. 202,53(10) : 3103-3110.

[12] Feng Huiyan. Corn starch content estimation based on optimized support vector machine [J]. Science and Technology Innovation, 2022(27) : 21-26.

About the author: Zhang Donghui, born in March 1977, male, Han nationality, native of Zhoukou, Henan Province, associate researcher, master candidate.

Research direction: Soybean breeding, cultivation, demonstration and extension.

Fund project source: Henan Research Platform 2022 Industry Fund; Project name: Creation and utilization of high yield and suitable soybean germplasm; Project number: 0000200045