Molecular Breeding

REVIEW

Genomic selection for crop improvement in fruits and vegetables: a systematic scoping review

Adrian Ming Jern Lee[1,2] · Melissa Yuin Mern Foong[3] · Beng Kah Song[3] · Fook Tim Chew[1,2]

Abstract

To ensure the nutritional needs of an expanding global population, it is crucial to optimize the growing capabilities and breeding values of fruit and vegetable crops. While genomic selection, initially implemented in animal breeding, holds tremendous potential, its utilization in fruit and vegetable crops remains underexplored. In this systematic review, we reviewed 63 articles covering genomic selection and its applications across 25 different types of fruit and vegetable crops over the last decade. The traits examined were directly related to the edible parts of the crops and carried significant economic importance. Comparative analysis with WHO/FAO data identified potential economic drivers underlying the study focus of some crops and highlighted crops with potential for further genomic selection research and application. Factors affecting genomic selection accuracy in fruit and vegetable studies are discussed and suggestions made to assist in their implementation into plant breeding schemes. Genetic gain in fruits and vegetables can be improved by utilizing genomic selection to improve selection intensity, accuracy, and integration of genetic variation. However, the reduction of breeding cycle times may not be beneficial in crops with shorter life cycles such as leafy greens as compared to fruit trees. There is an urgent need to integrate genomic selection methods into ongoing breeding programs and assess the actual genomic estimated breeding values of progeny resulting from these breeding programs against the prediction models.

Keywords Systematic scoping review · Fruit and vegetables · Genomic selection · In silico breeding · Genetic gain

✉ Fook Tim Chew
   dbscft@nus.edu.sg

1   Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Republic of Singapore

2   NUS Agritech Centre, National University of Singapore, 85 Science Park Dr, #01-03, Singapore 118258, Republic of Singapore

3   School of Science, Monash University Malaysia, Bandar Sunway, 47500 Subang Jaya, Selangor Darul Ehsan, Malaysia

Springer

Introduction

Definition of fruits and vegetables

Fruits and vegetables, as defined by The Food and Agriculture Organization of the United Nations (WHO/FAO) include "edible parts of plants, either cultivated or harvested wild, in their raw state or in a minimally processed form." This definition excludes "starchy roots and tubers, dry grain legumes, cereals, medicinal plants, stimulants (e.g. tea, coffee and cacao) and ultra-processed foods" (FAO 2021). Furthermore, staple food crops such as rice, wheat, maize, millet, barley, oats, and others are also excluded from the classification of fruits and vegetables for the purposes of this review.

Definition of genomic selection

Genomic selection (GS), also known as genomic prediction or genetic improvement, is the use of genome wide genetic markers to build models that predict plant phenotypic traits and can hence be applied to both animal and plant breeding programs (Desta & Ortiz 2014). All statistical and machine learning models calculating genomic estimated breeding values (GEBV) using genome wide genetic markers were considered as GS tools, and hence included in this review.

Background

Food security is a growing need presented to many communities, as nations work to become self-sufficient in the worrying scenarios of geo-political strife. Climate change is also drastically pressuring global food production and decreasing productivity levels (Campbell et al. 2016). Global wheat production was estimated to decrease by a staggering 6% per °C of global warming (Asseng et al. 2015). Lastly, an ever-growing world population is steadily increasing the world's food demands. An estimated global population of 9.1 billion people is expected to increase global crop demands by up to 110% from 2005 to 2050 (Tilman et al. 2011). A key aspect of food security is nutrition security. WHO/FAO recommends a healthy diet to consist of "at least 400 g of fruit and vegetables per day, excluding potatoes, sweet potatoes, cassava and other starchy roots" (Who & Consultation 2003). Hence, the issue of food security must also be addressed from a nutritional standpoint (Ingram 2020), where sufficient fruits and vegetables must be produced to allow for healthy diet consumption for the world's population. Increasingly, genomics and its applications in agriculture has been regarded as the key to global food security issues (Kumar et al. 2021). Through genomics approaches, the speed of agricultural breeding processes for livestock such as cattle and crops such as rice, has increased tremendously (Hayes et al. 2009) (Xu et al. 2021). Despite its apparent translational successes in certain

areas of agriculture, to our knowledge, little work has been done to apply GS methods to fruit and vegetable crops.

Aim of review

This systematic scoping review aims to evaluate the extent of application of GS in fruits and vegetables, focusing on a wide variety of crops with a large array of desirable traits. To this end, our objectives are: 1) To identify the fruit and vegetable crops that have been studied extensively, thereby highlighting the specific crops that may require more research attention; 2) To examine the traits that have been improved through GS, and those that could potentially be benefited from the same approach; 3) To assess the methods involved in GS applications within fruits and vegetables, and, 4) To present findings on how GS studies have been implemented in actual fruit and vegetable breeding strategies. To date, there is no systematic review available that specifically addresses the use and application of GS methods targeted at this specific category of crops.

Methodology

Search Strategy

This review was conducted in accordance with the Preferred Reporting Item for Systematic Review and Meta-Analyses (PRISMA) guidelines (Page et al. 2021). The PRISMA checklist is found in Supplementary Table 1. Scopus, Web of Science and BIOSIS databases were used to conduct the literature search on 21 December 2023. Full search terms for the databases are found in Table 1. An initial primary search was performed using these search terms followed by a secondary search using the same parameters but removing the "vegetable" and "fruit" search term to identify any other eligible results to comprehensively identify all relevant studies.

Table 1 Full search terms applied to databases; Scopus, Web of Science and BIOSIS for the literature search. All three databases were filtered for only English, research articles

Database	Full search term
Scopus	("Genom* select*" OR "Genom* improve*" OR "Genom* predict *") AND ("vegetable*" OR "crop*" OR "fruit*") AND NOT (" wheat " OR " oil " OR " grain " OR " rice " OR "barley" OR "maize" OR "soy")
Web of Science/BIOSIS	("Genom* select*" OR "Genom* improve*" OR "Genom* predict *") AND ("vegetable*" OR "crop*" OR "fruit*") NOT (" wheat " OR " oil " OR " grain " OR " rice " OR "barley" OR "maize" OR "soy")

Eligibility Criteria

With reference to the pre-stated definitions of "fruits and vegetables", as well as "genomic selection", this review selected articles that focus on the application of GS methods in these crops. All articles in English with full text available were included. Articles were excluded for one or more of the following reasons: the target crop was not included in the pre-defined definition of vegetables, or no target crop was specified; the article was a method paper reporting a protocol of GS with no model accuracy reported; the article did not use GS methods; the article was a review or meta-analysis article; or the full text was unavailable. Using these criteria, the primary search yielded an initial 850 articles that were screened. Their titles and abstracts were initially reviewed to identify suitable articles, before full texts were retrieved and reassessed for suitability based on the same exclusion criteria. Only articles that passed both abstract and title screening, as well as full text screening were included in the data extraction step for the final review. 2608 initial articles were identified in the secondary search, and they were screened using the same eligibility criteria. Duplicate studies identified in the primary search were removed so that only previously unidentified, eligible studies were finally included in the review.

Data Extraction

The following data was extracted from the filtered full-text articles: author(s), year, publishing journal, species studied, GS model used, statistical validation method used, prediction accuracy method used, accuracy of GS models reported, DNA extraction method used, sequencing method used, study sample size and traits of interest studied. The extracted data can be found in Table 2 (further details are included in Supplementary Table 2). Where details were unclear, not stated, or could not accurately be extracted from the article, data was not extracted from the article and regarded as 'NA'. All further data analysis and visualization were performed in RStudio.

Results and Discussion

Literature Search

Abstracts and titles from 850 articles were screened in the primary search for suitable articles. This resulted in 69 articles that qualified for further full-text screening. A total of 59 full-text articles were finally included into this review from the primary search after fulfilling the eligibility criteria (Fig. 1, Supplementary Table 3). In summary, Nagano et al. (2017) was excluded from the review during full-text screening as it was deemed to be a method paper reporting a protocol of GS with no accuracy reported while Yamamoto (2021) was excluded due to it being a review article, despite the two papers studying suitable target crops of strawberry and pear.

Table 2 Extracted data from all 59 full-text articles included in this review (Further details found in Supplementary Table 2)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
A Genome-Wide Association Study of Apple Quality and Scab Resistance (McClure et al. 2018)	Fruit	Apple	fivefold cross validation	Pearson correlation	NA	Genotyping-by-sequencing	172	75130 SNP	Fruit-related	BLUP
Accelerating genetic gains for quantitative resistance to verticillium wilt through predictive breeding in strawberry (Feldmann et al. 2023)	Fruit	Strawberry	Monte Carlo cross validation	Pearson correlation	E-Z 96®Plant DNA Kit	Affymetrix®50 K SNP Axiom®Array	921	25865 SNP	Disease resistance	BLUP / Bayesian
Accuracy and responses of genomic selection on key traits in apple breeding (Muranty et al. 2015)	Fruit	Apple	NA	Pearson correlation	NA	Illumina 20 K SNP Array	977	7829 SNP	Fruit-related	Bayesian

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Accuracy of genomic selection and long-term genetic gain for resistance to Verticillium wilt in strawberry (Pincot et al. 2020)	Fruit	Strawberry	Monte Carlo cross validation	Pearson correlation	E-Z 96 Plant DNA Kit (Omega Bio-Tek)	Affymetrix®50 K SNP Axiom®Array	388	35144 SNP	Disease resistance	BLUP / Bayesian / Others
Across-population genomic prediction in grapevine opens up promising prospects for breeding (Brault et al. 2022)	Fruit	Grape	fivefold cross validation	Pearson correlation	NA	Genotyping-by-sequencing	NA	32894 SNP	Fruit-related	RR
Adoption and Optimization of Genomic Selection To Sustain Breeding for Apricot Fruit Quality (Nsibi et al. 2020)	Fruit	Apricot	Cross validation	Pearson correlation	NA	Genotyping-by-sequencing	153	61030 SNP	Fruit-related	BLUP / Bayesian

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
An experimental validation of genomic selection in octoploid strawberry (Gezan et al. 2017)	Fruit	Strawberry	fivefold cross validation	Pearson correlation	NA	Affymetrix Axiom IStraw90 single-nucleotide polymorphism array	1628	17479 SNP	Fruit-related	BLUP / Bayesian / Others
Bayesian ridge regression shows the best fit for SSR markers in Psidium guajava among Bayesian models (da Silva et al. 2021)	Fruit	Guava	Cross validation	Pearson correlation	CTAB	SSR Genotyping	NA	200 SSR	Fruit-related	Bayesian
Combining genetic resources and elite material populations to improve the accuracy of genomic prediction in apple (Cazenave et al. 2022)	Fruit	Apple	fivefold cross validation	Pearson correlation	NA	Affymetrix Axiom Apple 480 K SNP genotyping array Illumina Infinium 20 K SNP genotyping array	NA	303239 SNP	Fruit-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Comparison of genotypic and phenotypic selection of breeding parents in a carrot (Daucus carota) germplasm collection (Corak et al. 2023)	Vegetable	Carrot	NA	NA	CTAB	Genotyping-by-sequencing	749	19944 SNP	Plant-related	BLUP
Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato (Liabeuf et al. 2018)	Vegetable	Tomato	Cross validation	Pearson correlation	CTAB	"SolCAP" Infinium array	109	397 SNP	Disease resistance	Bayesian

Molecular Breeding   (2024) 44:60   Page 9 of 50   60

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Comparison of Representative and Custom Methods of Generating Core Subsets of a Carrot Germplasm Collection (Corak et al. 2019)	Vegetable	Carrot	NA	Pearson correlation	CTAB	Genotyping-by-sequencing	749	19944	Plant-related	BLUP
Efficiency of genomic selection for breeding population design and phenotype prediction in tomato (Yamamoto et al. 2017)	Vegetable	Tomato	tenfold cross validation	Pearson correlation	Qiagen DNeasy Plant Mini Kit	96 Dynamic SNP Array	96	337 SNP	Fruit-related	BLUP / Bayesian / Others
Efficiency of genomic selection for tomato fruit quality (Duangjit et al. 2016)	Vegetable	Tomato	Cross validation	Pearson correlation	NA	7720 SNP Array Infinium assay	122	5995 SNP	Fruit-related	BLUP

🖄 Springer

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Evaluation of Selection Methods for Resistance to a Specialist Insect Pest of Squash (Cucurbita pepo) (Brzozowski & Mazourek 2020)	Vegetable	Squash	fivefold cross validation	NA	Qiagen DNeasy 96 Plant kit	Genotyping-by-sequencing	190, 130	16052 SNP	Disease resistance	BLUP
Feasibility of genomic prediction for brown rot (Monilinia spp.) resistance in peach (Fu et al. 2022)	Fruit	Peach	fivefold cross validation	Pearson correlation	NA	9 + 9 K peach SNP Array	256	8014 SNP	Disease resistance	BLUP / Bayesian / Others
Genetic architecture and genomic predictive ability of apple quantitative traits across environments (Jung et al. 2022)	Fruit	Apple	fivefold cross validation	Pearson correlation	NA	SNP Array	534	303148 SNP	Fruit-related	BLUP / Bayesian / Others
Genetic architecture of fresh-market tomato yield (Bhandari et al. 2023)	Vegetable	Tomato	fivefold cross validation	NA	NA	Whole genome resequencing	425	7266 SNP	Fruit-related	BLUP / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genetic characterization of carrot root shape and size using genome-wide association analysis and genomic-estimated breeding values (Brainard et al. 2022)	Vegetable	Carrot	Cross validation	Pearson correlation	Macherey–Nagel NucleoSpin 96-well kits	Whole genome resequencing	749	12370 SNP	Root-related	BLUP
Genome selection in fruit breeding Application to table grapes (Viana et al. 2016)	Fruit	Grape	Cross validation	Pearson correlation	NA	NA	143	243 SSR	Fruit-related	BLUP / Bayesian
Genome wide selection in citrus breeding (Góis et al. 2016)	Fruit	Citrus	Cross validation	Pearson correlation	NA	5287 DArT_seq	180	5287 SNP	Fruit-related	BLUP
Genome-Based Prediction of Time to Curd Induction in Cauliflower (Rosen et al. 2018)	Vegetable	Cauliflower	NA	Pearson correlation	NA	20 k *B. oleracea* Illumina Infinium iSelect single-nucleotide polymorphism (SNP) array	265	176 SNP	Induction time	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies (Biscarini et al. 2017)	Fruit	Peach	fivefold cross validation	Pearson correlation	NA	IPSC 9 K SNP array	1147	6076 SNP	Fruit-related	BLUP
Genome-wide association study and genomic prediction in citrus Potential of genomics-assisted breeding for fruit quality traits (Minamikawa et al. 2017)	Fruit	Citrus	tenfold cross validation	Pearson correlation	NA	SNP Array	676	1841 SNP	Fruit-related	BLUP / Bayesian / RR / Others
Genome-wide association study and genomic prediction of white rust resistance in USDA GRIN spinach germplasm (Shi et al. 2022)	Vegetable	Spinach	Cross validation	Pearson correlation	NA	Whole genome resequencing	346	13235 SNP	Disease resistance	BLUP / Bayesian / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genome-wide association study of individual sugar content in fruit of Japanese pear (Pyrus spp.) (Nishio et al. 2021)	Fruit	Pear	Cross validation	Pearson correlation	Qiagen DNeasy Plant Mini Kit	ddRAD Sequencing	1218	3483 SNP	Fruit-related	BLUP / Bayesian
Genome-Wide Prediction Methods in Highly Diverse and Heterozygous Species Proof-of-Concept through Simulation in Grapevine (Fodor et al. 2014)	Fruit	Grape	Cross validation	Pearson correlation	NA	NA	1000	NA	NA	Bayesian / RR
Genomewide selection for fruit quality traits in apple breeding insights gained from prediction and postdiction (Kostick et al. 2023)	Fruit	Apple	NA	Pearson correlation	NA	SNP Array	852	977 SNP	Fruit-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits (Piaskowski et al. 2018)	Fruit	Cherry	fivefold cross validation	Pearson correlation	NA	RosBREED cherry 6 K SNP array	644–774	1615 SNP	Fruit-related	BLUP
Genomic prediction and association mapping of curd-related traits in gene bank accessions of cauliflower (Thorwarth et al. 2018)	Vegetable	Cauliflower	fivefold cross validation	Pearson correlation	NA	Genotyping-by-sequencing	192	120693 SNP	Plant-related	BLUP / Bayesian
Genomic prediction and selection for fruit traits in winter squash (Hernandez et al. 2020)	Vegetable	Squash	Cross validation	Pearson correlation	Qiagen DNeasy 96 Plant Kit (Qiagen, Valencia, CA, USA)	Genotyping-by-sequencing	179	535 SNP	Fruit-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genomic prediction in a multiploid crop Genotype by environment interaction and allele dosage effects on predictive ability in Banana (Nyine et al. 2018)	Fruit	Banana	Cross validation	Pearson correlation	NA	Genotyping-by-sequencing	307	10807 SNP	Plant-related	Bayesian / Others
Genomic prediction of autotetraploids influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction (de Bem Oliveira et al. 2019)	Fruit	Blueberry	tenfold cross validation	Pearson correlation	NA	Whole genome resequencing	1847	85973 SNP	Fruit-related	Bayesian
Genomic Prediction of Pumpkin Hybrid Performance (Wu et al. 2019)	Vegetable	Pumpkin	Cross validation	Pearson correlation	Qiagen DNeasy Plant Mini Kit	Genotyping-by-sequencing	320	61179 SNP	Fruit-related	BLUP / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genomic prediction of strawberry resistance to postharvest fruit decay caused by the fungal pathogen Botrytis cinerea (Petrasch et al. 2022)	Fruit	Strawberry	Cross validation	Pearson correlation	E-Z 96 Plant DNA Kit (Omega Bio-Tek)	Affymetrix®50 K SNP Axiom®Array	380	11946 SNP	Disease resistance	BLUP / Others
Genomic prediction of trait segregation in a progeny population a case study of Japanese pear (Pyrus pyrifolia) (Iwata et al. 2013a, b)	Fruit	Pear	Cross validation	Pearson correlation	NA	SSR Genotyping	84	333 SSR	Fruit-related	Bayesian
Genomic predictions in diploid Actinidia chinensis (kiwifruit) (Cheng et al. 2019)	Fruit	Kiwifruit	NA	Pearson correlation	Qiagen Dneasy plant mini kit	Random-tag genotyping-by-sequencing	88	8372 SNP	Fruit-related	BLUP / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genomic Selection for F1 Hybrid Breeding in Strawberry (Fragaria × ananassa) (Yamamoto et al. 2021)	Fruit	Strawberry	twofold cross validation	Pearson correlation	Qiagen DNeasy Plant Mini Kit	Whole-genome shotgun sequencing	105	20811 SNP	Plant-related	BLUP / Bayesian / Others
Genomic Selection for Prediction of Fruit-Related Traits in Pepper (Capsicum spp.) (Hong et al. 2020)	Vegetable	Pepper	Cross validation	Pearson correlation	NA	Genotyping-by-sequencing	302	18,663 SNP	Fruit-related	Others
Genomic selection in algae with biphasic lifecycles A Saccharina latissima (sugar kelp) case study (Huang et al. 2023)	Vegetable	Sugar kelp	tenfold cross validation	Pearson correlation	CTAB	Whole genome resequencing	866	909747 SNP	Plant-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genomic Selection in an Outcrossing Autotetraploid Fruit Crop Lessons From Blueberry Breeding (Ferrão et al. 2021)	Fruit	Blueberry	tenfold cross validation	NA	NA	"Capture-Seq"	280	48829 SNP	Fruit-related	BLUP
Harnessing multivariate, penalized regression methods for genomic prediction and QTL detection of drought-related traits in grapevine (Brault et al. 2021)	Fruit	Grape	Nested cross validation	Pearson correlation	NA	Genotyping by sequencing, RAD sequencing	NA	17298 SNP	Fruit-related	RR / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Harnessing underutilized gene bank diversity and genomic prediction of cross usefulness to enhance resistance to Phytophthora cactorum in strawberry (Jiménez et al. 2023)	Fruit	Strawberry	Cross validation	Pearson correlation	E-Z 96 Plant DNA Kit (Omega Bio-Tek)	AxiomTM Strawberry 50 K SNP array	437	40334 SNP	Disease resistance	BLUP / Bayesian / Others
Impact of dominance effects on autotetraploid genomic prediction (Amadeu et al. 2020)	Fruit	Blueberry	Replicated train-ing–testing evaluation	Pearson correlation	NA	NA	1804	97314 SNP	Fruit-related	BLUP / Bayesian / RR
Implementation of different relationship estimate methodologies in breeding value prediction in kiwiberry (Actinidia arguta) (Mertten et al. 2023)	Fruit	Kiwiberry	Cross validation	Pearson correlation	NA	RAPID sequencing	842	NA	Fruit-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Implementing genomic selection in sour passion fruit population (Viana et al. 2017)	Fruit	Passion fruit	Cross validation	NA	NA	NA	95	183 SSR and ISSR	Fruit-related	Bayesian / Others
Implications of genomic selection for obtaining $F_{2:3}$ families of tomato (Hernández-Bautista et al. 2016)	Vegetable	Tomato	NA	Pearson correlation	CTAB	SSR Genotyping	152	31 SSR	Fruit-related	BLUP / Bayesian / RR
Inheritance of Yield Components and Morphological Traits in Avocado cv. Hass From Criollo Elite Trees via Half-Sib Seedling Rootstocks (Cañas-Gutiérrez et al. 2022)	Fruit	Avocado	NA	Pearson correlation	NA	SSR Genotyping	104	13 SSR	Plant-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model BLUP / Bayesian / RR / Others
Mapping and selection of downy mildew resistance in spinach cv. whale by low coverage whole genome sequencing (Bhattarai et al. 2022a)	Vegetable	Spinach	fivefold cross validation	Pearson correlation	Omega Mag-Bind Plant DNA DS kit	Whole genome resequencing	192	8189 SNP	Disease resistance	BLUP
Marker-trait associations and genomic predictions of interspecific pear (Pyrus) fruit characteristics (Kumar et al. 2019)	Fruit	Pear	Cross validation	Pearson correlation	NA	Genotyping-by-sequencing	550	7509 SNP	Fruit-related	BLUP
Multiple-statistical genome-wide association analysis and genomic prediction of fruit aroma and agronomic traits in peaches (Li et al. 2023)	Fruit	Peach	fivefold cross validation	Pearson correlation	NA	NA	242	145456 SNP	Fruit-related	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Multivariate GBLUP Improves Accuracy of Genomic Selection for Yield and Fruit Weight in Biparental Populations of Vaccinium macrocarpon Ait (Covarrubias-Pazaran et al. 2018)	Fruit	Cranberry	NA	Pearson correlation	NA	Genotyping-by-sequencing	NA	7389 SNP	Fruit-related	BLUP
Potential assessment of genome-wide association study and genomic selection in Japanese pear Pyrus pyrifolia (Iwata et al. 2013a, b)	Fruit	Pear	Cross validation	Pearson correlation	NA	SSR Genotyping	76	162 SSR	Fruit-related	Bayesian

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Prediction accuracy of genomic selection models for earliness in tomato (Hernández-Bautista et al. 2020)	Vegetable	Tomato	tenfold cross validation	Pearson correlation	NA	NA	172	61 SSR	Induction time	Bayesian / Others
Resequencing worldwide spinach germplasm for identification of field resistance QTLs to downy mildew and assessment of genomic selection methods (Bhattarai et al. 2022b)	Vegetable	Spinach	fivefold cross validation	Pearson correlation	Omega Mag-Bind Plant DNA DS kit	Whole genome resequencing	434	88682 SNP	Disease resistance	BLUP / Bayesian / RR / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Role of MdERF3 and MdERF118 natural variations in apple flesh firmness/crispness retainability and development of QTL–based genomics-assisted prediction (Wu et al. 2021)	Fruit	Apple	fivefold cross validation	Pearson correlation	CTAB	KASP Assay	394–473	83 SNP	Fruit-related	Others
Simulation-based optimization of genomic selection scheme for accelerating genetic gain while preventing inbreeding depression in onion breeding (Sekine & Yabe 2020)	Vegetable	Onion	NA	NA	NA	NA	Simulated data	Simulated data	NA	BLUP
The apple REF-POP-a reference population for genomics-assisted breeding in apple (Jung et al. 2020)	Fruit	Apple	fivefold cross validation	Pearson correlation	NA	SNP Array	534	303329 SNP	Induction time	BLUP

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Turbocharging introgression breeding of perennial fruit crops a case study on apple (Kumar et al. 2020)	Fruit	Apple	11-fold cross validation	Pearson correlation	Qiagen DNeasy Plant mini kit	Random-tag genotyping-by-sequencing	274	6400 SNP	Fruit-related	BLUP
Defining strawberry shape uniformity using 3D imaging and genetic mapping (Li et al. 2020)*	Fruit	Strawberry	tenfold cross validation	Pearson correlation	Qiagen DNeasy Plant mini kit	SNP Array	416	34949 SNP	Fruit-related	BLUP
Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (Pyrus pyrifolia Nakai) (Minamikawa et al. 2018)*	Fruit	Pear	tenfold cross validation	Pearson correlation	NA	SNP Array	502	1536 SNP	Fruit-related	BLUP / Bayesian / RR / Others

Table 2 (continued)

Title	Crop Type	Species (Common)	Validation	Prediction accuracy method	DNA Extraction	Sequencing method	Sample size	Genetic markers	Trait	Model
Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple (Roth et al. 2020)*	Fruit	Apple	fivefold cross validation	Pearson correlation	Qiagen DNeasy Plant mini kit	SNP Array	259	8294 SNP	Fruit-related	BLUP
Ridge regression and deep learning models for genome-wide selection of complex traits in New Mexican Chile peppers (Lozada et al. 2023)*	Vegetable	Pepper	tenfold cross validation	Pearson correlation	Qiagen DNeasy Plant mini kit	Genotyping-by-sequencing	204	14922 SNP	Plant-related	BLUP / Bayesian / RR / Others

*Article added through secondary search
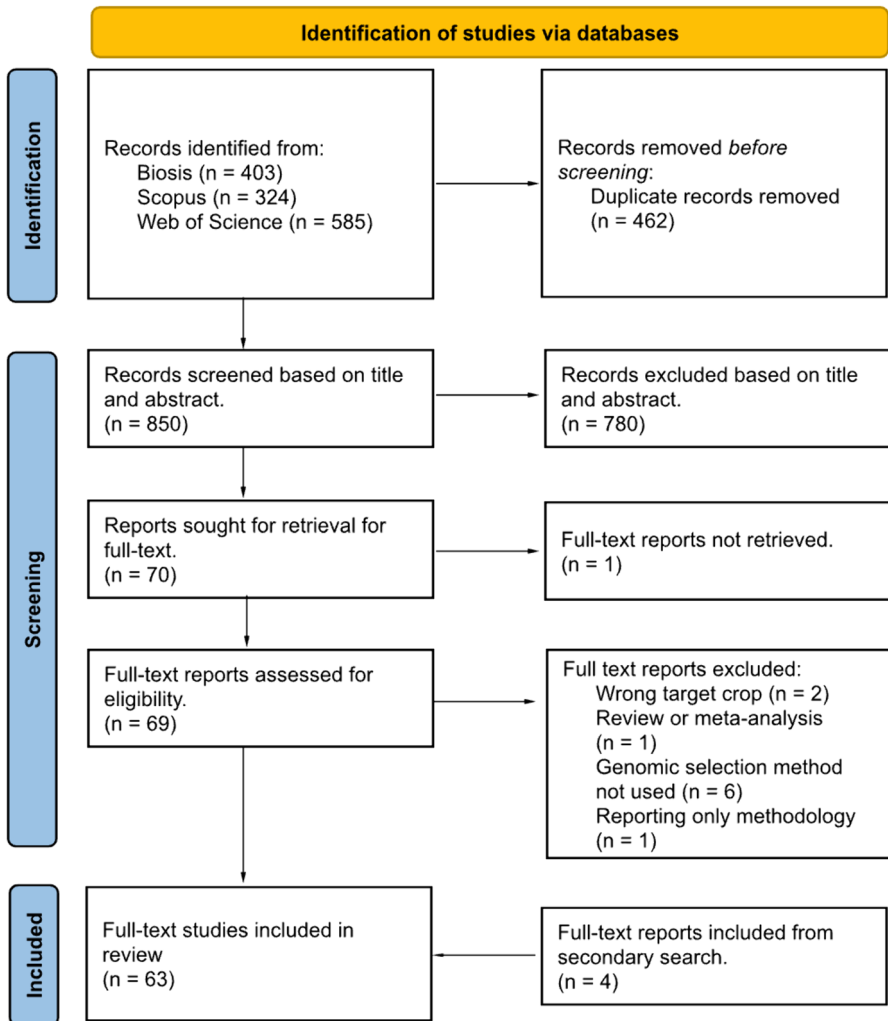
Fig. 1 PRISMA flow diagram of the process used to identify and shortlist articles for systematic review

Similarly, Hardigan et al. (2023); Liabeuf and Francis (2017); Longhi et al. (2013); Nyine et al. (2017); Ohyama et al. (2023); Wilkinson et al. (2022) were all excluded despite their study of fruit and vegetable crops as they did not use GS methods. An additional four articles were identified from the secondary search and included in this review (Li et al. 2020; Lozada et al. 2023; Minamikawa et al. 2018; Roth et al. 2020), totaling the number of articles included in this review to 63.

Based on the definition of fruits and vegetables specified, seventeen crops were excluded from the study at abstract screening stage. They included alfalfa (*Medicago sativa*), black tea (*Camellia sinensis*), cacao (*Theobroma cacao*), cassava (*Manihot esculenta*), chickpea (*Cicer arietinum*), chestnut (*Castanea sativa*), chrysanthemum

(*Chrysanthemum morifolium*), coffee (*Coffea arabica*), common bean/snapbean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*), flax (*Linum usitatissimum*), groundnut/peanut (*Arachis hypogaea*), pea (*Pisum sativum*), potato (*Solanum tuberosum*), sesame (*Sesamum indicum*), sugarcane (*Saccharum officinarum*), and walnut (*Juglans regia*).

Overview of studies

GS in fruits and vegetable crops is relatively new. A database search on related topics revealed publication released only from 2013 onwards. Since then, there has been an overall increase in the number of publications annually (Fig. 2). This increasing trend could be largely attributed to the rapidly decreasing cost of genomic sequencing technology and the rapid availability of genomic data as well as reference genomes (Muir et al. 2016). GS was popularized in 2001 (Meuwissen et al. 2001) and largely applied to livestock as well as staple crops. The successes of GS are well recorded in livestock. Specifically, the cattle industry has greatly benefited from the use of GS breeding (Weller et al. 2017). The aquaculture industry has also largely benefitted from GS breeding particularly in the area of disease resistance (Houston et al. 2020). When discussing key products delivered by genomics-assisted breeding, Wang et al. (2022) shares key developments in staple crops such as rice, wheat and legumes, with no specific mention of fruit or vegetable crops. It has seemingly taken a decade for GS to be applied to fruits and vegetables, starting in 2013. Despite the difference in agricultural organism, there are key lessons that can be taken from such
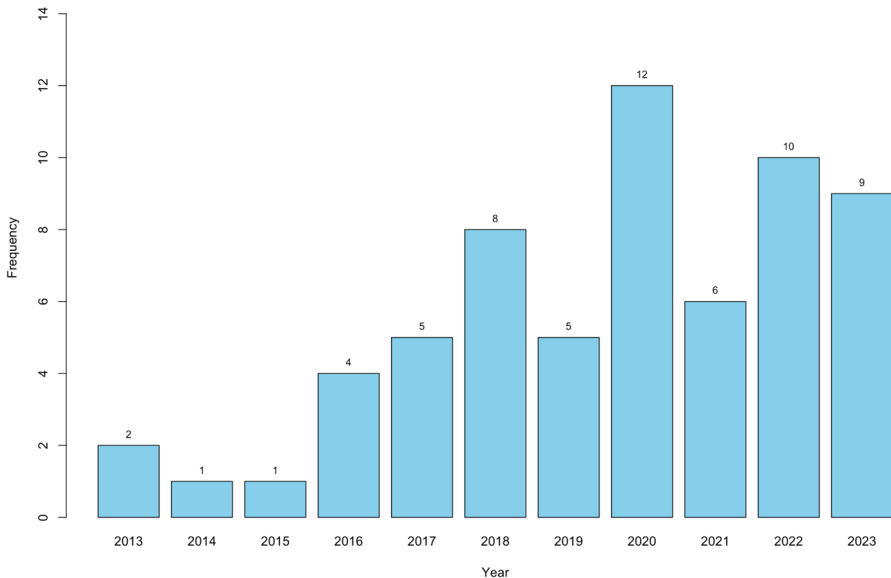
Fig. 2 Number of articles meeting the inclusion criteria of GS studies in fruits and vegetables published each year from 2013 to 2023

successes. In dairy cattle, breeding pedigree for breeding lines have been developed over decades in such a mature industry (Gutierrez-Reinoso et al. 2021). Genotypic and phenotypic data for breeding lines are readily available for purchase together with breeding material, meaning that the initial phenotyping and genotypoing work by the breeders are no longer required. Such data may not be as readily available in some fruits and vegetable species, hence breeding plans must first collect such data and develop breeding lines, which can be costly. In staple crops, Xu et al. (2020)'s genomic selection review recommends the creation of genetic chimeras that contain haplotype combinations of two or more elite parental lines to maximise GS efforts.

Overview of fruit and vegetable crops studied

From the literature search, a total of 25 different fruits and vegetables have been utilized in GS studies in the last 10 years. They comprise 16 fruit and 9 vegetable crops (Table 2). Sugar kelp (*Saccharina latissima*) was classified as a vegetable crop as it is a commonly consumed edible kelp in Asia for its nutritional value and flavor (Stévant et al. 2018). Huang et al. (2023) studied GS for 11 yield and morphological traits from a study population of 866 sugar kelp individuals and reported GS cross validation accuracies ranging from -0.23 to 0.48. GS models were applied to make crosses for experimental validation. There is only one leafy vegetable species included in this review, spinach (*Spinacia oleracea*) (Bhattarai et al. 2022a; Bhattarai et al. 2022b; Shi et al. 2022). All three publications reported GS work in spinach relating to disease resistance. Bhattarai et al. (2022b) and Bhattarai et al. (2022a) both reported genome wide association studies (GWAS) to identify single nucleotide polymorphisms (SNP) associated with resistance to downy mildew disease, caused by the oomycete pathogen *Peronospora effusa* in spinach. They further implemented genomic prediction for this trait by utilizing significant SNPs associated with disease resistance. Another GS study in spinach related to disease resistance was on white rust disease caused by *Albugo occidentalis* (Shi et al. 2022). Cauliflower (*Brassica oleracea* var. *botrytis*) was also studied (Rosen et al. 2018; Thorwarth et al. 2018) on traits such as curd developmental, morphological, and induction time-related traits. Rosen et al. (2018) reported the training of ridge-regression best linear unbiased prediction GS models with prediction accuracies ranging from 0.52–0.61 for predicting curding times, while Thorwarth et al. (2018) reported accuracies ranging from 0.10–0.66 for different curd-related traits using genomic best linear unbiased prediction and Bayesian B GS models. There is a clear distinction between the number of studies conducted on different types of fruit and vegetable crops over the last decade. In total, 39 reports published GS work in fruit type crops, while there were only 20 reports in vegetable crops (Table 2). Of the crops studied, apples (*Malus domestica*) ($n=9$), strawberries (*Fragaria × ananassa*) ($n=7$) and tomatoes (*Solanum lycopersicum*) ($n=6$) were identified to have the greatest number of reports (Fig. 3a).

Data provided by the Food and Agriculture Organization of the United Nations Statistics Division (FAOSTAT) (Food, Agriculture Organization of the United 1997) shows a slight economic importance of fruit over vegetable crops. From the
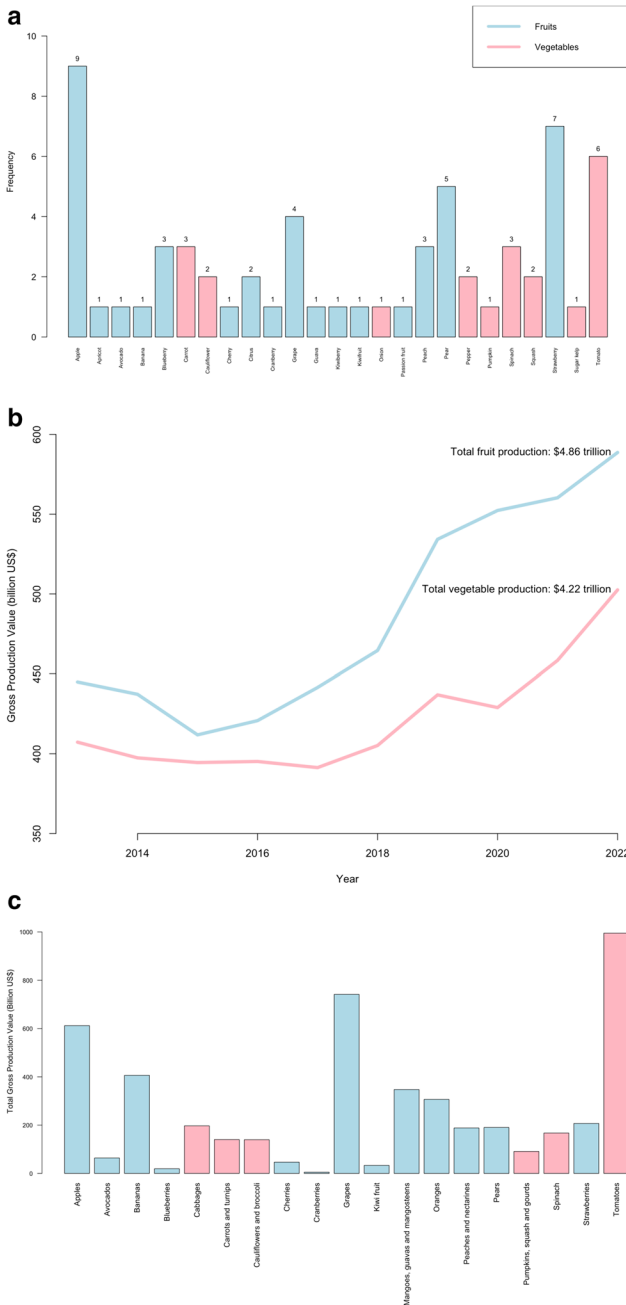
Fig. 3 a: Number of articles published for each fruit and vegetable crop from 2013 to 2023. b: Annual gross production value (billion USD) of fruit and vegetable crops from 2013 to 2022 (Food & Agriculture Organization of the United, 1997). c: Total gross production value of crops from 2013 to 2022 (Food, Agriculture Organization of the United 1997)

period of 2013 to 2022, the annual average gross production value (USD) of fruits crops ($4.86 trillion) was 15.14% higher than that of vegetable crops ($4.22 trillion) (Fig. 3b). Further analysis revealed tomatoes ($859.70 billion) to have the highest total gross production value within that time frame, followed by grapes (*Vitis spp*) ($649.36 billion) and apples ($525.41 billion). This may explain the relative research importance represented by increased research publications for apples and tomatoes. However, the similar trend is not observed for grapes, with only four GS publications identified despite its high gross production value (table grape, $n=1$; grapevine, $n=3$) (Fig. 3a, Fig. 3c), potentially identifying grapes as a fruit crop with greater capacity for genomic breeding research using GS. Conversely, despite ranking 12th in gross production value ($172.90 billion), strawberries had the second highest number of publications ($n=7$). The range of fruit and vegetable crop species studied for the application of GS remains limited, particularly in vegetable species, compared to other wheat and grain crops, and to livestock. The higher economic value of each individual, coupled with greater reduction in generation time means that there is a larger genetic gain for GS application in livestock such as cattle (Xu et al. 2020). In turn, the smaller economic importance of vegetable crops and shorter generation time may potentially serve to hinder researchers from studying GS in such crops. However, this research gap needs to be actively addressed as the world strives towards food and further nutritional security, ensuring the availability of healthy diets to the world population. Fruits and vegetables must be readily available, and their growing capacities enhanced. We recommend research of GS methods for enhanced breeding of more fruit and vegetable crops, specifically vegetable crops.

Overview of fruit and vegetable traits improvement

Categorization of traits is difficult to generalize as different studies contain different definitions and methods of measuring various traits. The difficulty in generalizing traits makes their comparison across multiple studies challenging. Not surprisingly, traits reported in this review are mostly directly related to the crop product to be consumed. In fruit crops, these traits were largely related to yield as well as quality of the fruit itself. Of the nine apple GS studies, five studied some form of yield-related trait (such as fruit size or weight) (Jung et al. 2022; Kostick et al. 2023; Kumar et al. 2020; McClure et al. 2018; Muranty et al. 2015) while five studied varying color related traits (Cazenave et al. 2022; Jung et al. 2022; Kostick et al. 2023; McClure et al. 2018; Muranty et al. 2015). Additionally, five looked at fruit firmness (Jung et al. 2022; Kumar et al. 2020; McClure et al. 2018; Roth et al. 2020; Wu et al. 2021), three looked at crispness (Cazenave et al. 2022; Kumar et al. 2020; Roth et al. 2020); all in relation to the quality of the fruit. McClure et al. (2018) studied post-harvest qualities by measuring the firmness, soluble solids content (SSC) and titratable acidity (TA) of apples after three months in refrigerated storage. In six tomato studies, yield and fruit quality traits also played large importance (Bhandari et al. 2023; Duangjit et al. 2016; Hernández-Bautista et al. 2020, 2016; Liabeuf et al. 2018; Yamamoto et al. 2017). Only Liabeuf et al. (2018) studied GS methods for disease resistance in tomatoes, while Hernández-Bautista et al. (2020)

studied fruiting earliness in tomatoes as it "indirectly increases yield and offers more supply opportunities by increasing the production window". Such studies are an example of utilizing GS for traits not directly related to yield or quality, but which may have important indirect impacts that can still result in economic gain for farmers. More of such studies should be performed to understand GS methods in indirect traits that may still affect fruit and vegetable production. Disease resistance was also noted to be an important aspect of breeding in strawberries; four of six strawberry studies reported GS models focused on plant disease resistance, for example, *Phytophthora* crown rot (Jiménez et al. 2023), gray mold (Petrasch et al. 2022), and *Verticillium* wilt resistance (Feldmann et al. 2023; Pincot et al. 2020). Similarly, studies on kiwifruit (*Actinidia chinensis*) (Brzozowski & Mazourek 2020), peach (*Prunus persica*) (Fu et al. 2022), spinach (Bhattarai et al. 2022a; Bhattarai et al. 2022b; Shi et al. 2022), and squash (*Cucurbita*) (Brzozowski & Mazourek 2020) among others, also addressed disease resistance. Meanwhile, Gezan et al. (2017) studied strawberry traits in relation to the marketability of strawberries, Li et al. (2020) explored genomic selection in strawberry shape uniformity and Yamamoto et al. (2021) studied petiole length and leaf area of the leaves of strawberry plants on top of the usual fruit-related traits. The term "phenotyping bottleneck" describes the limitations or bottlenecks in the phenotyping process, which involves measuring and assessing traits or phenotypes of organisms—in this case, fruits, and vegetables. The difficulties in accurately and efficiently measuring traits gives rise to this bottleneck, which can hinder the advancement of GS and breeding programs. Phenotyping is also extremely costly and laborious if one intends to collect high-quality, large datasets, which are crucial for accurate GS model training. Increasingly, there are technological advancements primarily making use of the improvement in image capturing and processing technology to deal with this bottleneck. Li et al. (2020) used a 3D imaging approach to measure strawberry uniformity and show that it is comparative to manual measurements before implementing these data sets into genomic selection models. They present an example of using imaging technology to deal with the phenotyping bottleneck while benefit genomic selection model training. However, there is still work that needs to be done to validate many of these new technological advancements and it will take time before many of these technologies are readily applicable to varying crops. These technologies, challenges and opportunities have been heavily presented in reviews over the years (Bongomin et al. 2024; Sheikh et al. 2024; Yang et al. 2020). Artificial intelligence coupled with advanced imaging technology has potential to truly advance high-throughput phenotyping. Imaging technology, hyperspectral imaging, thermal imaging, 3D imaging and fluorescence imaging, together with artificial intelligence now means that more data than ever can be collected and processed to identify patterns and draw conclusions (Sheikh et al. 2024).Such non-destructive phenotyping allows researchers to actively monitor crop traits over periods of time and collect data without causing potential harm to the plant. Phenotypes collected by the reported studies can largely be categorized into crop quality (including chemical and metabolic analysis of crops), post-harvest related traits, yield-related traits, and growth-related traits (including disease resistance and growth of the plant itself). Within these categories, traits are limited by the availability of methods to measure them effectively and accurately. As

such, traits tend to be biased towards those that are economically important or easily measured through simple phenotyping means. Phenotypic traits with similar underlying genetic architecture can often be correlated with one another to identify associated and secondary traits. In breeding programs, indirect selection for traits that are expensive or difficult to measure using correlated secondary traits is common (Rutkoski et al. 2016). While it is important to gather economically important primary traits, the collection of seemingly unimportant, secondary traits can prove to be vital in reducing the cost of phenotyping in GS breeding schemes. This work will require an initial "deep" phenotyping of desired organisms, followed by phenotypic correlation studies to understand the relationships between phenotypes and identify useful secondary traits. The repertoire of deep phenotypes collected and made available to the public will aid researchers in selecting and identifying cheaper methods of phenotyping desired traits.

Factors affecting genomic selection prediction accuracy in fruits and vegetables

GS model performance is influenced by factors such as study sample size and population structure, marker density, trait genetic architecture and heritability, as well as the GS model selection (Budhlakoti et al. 2022). The correlation coefficient between genomic predicted GEBVs and observed trait values is most used to assess model accuracy. Cross validation methods are regularly used to ensure model stability and robustness as they simulate real world breeding programs. In a k-fold cross validation, available genotypic with coupled phenotypic data is divided into k groups. $k - 1$ subgroups are then used to train the model, and the model tested over the kth group. This process is iteratively repeated to obtain variations of cross validation accuracies to calculate mean and standard deviations of GS model performance. Allowing the user to understand the robustness and reliability of their GS model.

Training population design

For effective application of GS model training, the training population used must be large and representative of the available allele combinations and linkage disequilibrium (LD) patterns of the desired trait and crop target. Without such representation in the model training, model performance often dips considerably when applied to breeding programs (de Los Campos et al. 2013). Models trained on a small diversity of genetically related individuals lack prediction accuracy when tested. Corak et al. (2019) applied five different strategies of sampling from a collection of 433 domesticated carrot (*Daucus carota*) accessions to develop a smaller, core collection of representative accessions. They found that none of the strategies significantly affected GS model accuracy ($r^2 = 0.2$–0.3) and concluded that although it is still possible to use diverse germplasm collections to train accurate GS models, the aim of using core collections of 10–20% may be unreasonable as there is insufficient genetic diversity in the suggested training data. Similarly, breeding programs implementing GS models in populations trained with different genetic backgrounds will find that model performance no longer holds to the prediction accuracies first

recorded. In Cazenave et al. (2022), models trained for apple fruit traits and tested on a genetically different population consistently showed a reduction of prediction accuracy. However, they demonstrated that models trained using a combination of both differing populations exhibited prediction accuracies like that of the within population model tests. Thus, it is important to factor in the genetic background and composition of the training and eventual breeding population of the breeding program. Where training and subsequent breeding populations may be genetically dissimilar, efforts should be made to introduce more diversity from the breeding population into training the GS model.

Minimally effective population size is important to reduce genotyping and logistic costs in a breeding program yet provide maximum model training efficiency. Model accuracy is often well correlated with training population sample size. Nsibi et al. (2020) tested GS model application using 25%, 50% and 75% of the total population size ($n = 152$) as the training population and consistently found that models trained on 75% of the total population exhibited greater model accuracies in apricots (*Prunus armeniaca*). The same was observed in Duangjit et al. (2016) with 75% of the total population size ($n = 163$) of tomatoes trained models with the highest accuracy. However, model accuracy eventually evens out after which the addition of more samples is no longer cost-effective. (Arruda et al. 2015; Fernández-González et al. 2023; Sverrisdóttir et al. 2018) The elbow point serves as a crucial intersection to identify the most cost-effective training population sample size for use in breeding programs. In wheat, Edwards et al. (2019) concluded that a small number of close relatives or a very large number of distant relatives are expected to improve model training accuracy. Given that breeders often rely on a relatively narrow pool of genetically similar crops for breeding, smaller sample sizes should suffice in training accurate GS models. However, breeders that would like to incorporate more trait and genetic diversity will have to factor in the increased costs of phenotyping and genotyping a larger population. While this may be possible for smaller fruit and vegetable crops such as leafy greens, this may prove to be difficult when applied to crops such as fruit trees, which are logistically large and often take many years to mature.

Trait heritability

Phenotypic variance, V_P, is the combined effects of genetic variance, V_G, and environmental variance, V_E.

$$V_P = V_G + V_E$$

Heritability measures the extent of which the phenotypic variation is contributed to by genetic factors. Together with the observed selection differential, S, that is the phenotypic difference between the selected parents and the parental population, heritability, h^2, can be used to calculate the observed selection response, R. This is also known as the breeder's equation.

$$R = h^2 \times S$$

The prediction accuracy of a GS model is greatly influenced by the heritability of the trait of interest. That is, traits that are largely contributed by genetic variance are more accurately predicted using genomic markers. The heritability of the trait can also affect the optimum sample size required for model training, as previously discussed. Self-pollinated crops with huge LD blocks passed down over many generations maintain a relatively high heritability for many traits (Anilkumar et al. 2022). Poor predictive accuracy is often attributed with traits of low heritability as the variation can largely be attributed to the environmental variance. Duangjit et al. (2016) demonstrated in 45 tomato crop traits the positive correlation between trait heritability and predictive accuracy ($r = 0.69$). The predictive accuracy of traits decreased with the heritability of the trait when models were trained on the same training data. In cauliflower, GS prediction worked best for highly heritable traits (number of days to budding and cluster width) compared with traits of low heritability (apical length and length of nearest branch) (Thorwarth et al. 2018). An increase in sample size of training data could demonstrate an increase in accuracy for traits with low heritability. Breeding programs must hence consider trait heritability in designing breeding programs of maximum efficiency and value. Although GS is effective for complex traits controlled by many gene loci, traits that are largely affected by environmental variability can greatly lower GS accuracy. In such situations, multi-environmental trials and models incorporating random and fixed effects can be used as an attempt to rectify this. Nyine et al. (2018) found that models which included averaged environmental data as added effects across multiple test years improved GS accuracy. They evaluated 15 traits in bananas (*Musa spp*) in two environmental crop conditions over two crop cycles. Jung et al. (2020) describes a method to deal with the spatial heterogeneity within environments to reduce the effect of environmental variability that may affect GS accuracy. These methods can be employed to deal specifically with traits of low heritability when designing GS experiments.

Marker density

Marker density and type plays an important role in enabling GS model accuracy. Genomic markers reviewed in this article primarily consisted of SNPs, with the exception of eight studies that utilized simple sequence repeats (SSRs) (Cañas-Gutiérrez et al. 2022; da Silva et al. 2021; Hernández-Bautista et al. 2020, 2016; Iwata et al. 2013a, b; Iwata et al. 2013a, b; Viana et al. 2017, 2016). Sekine and Yabe (2020) and Fodor et al. (2014) utilized simulated data to perform simulation genomic breeding program studies and understand its applications within an artificial environment. Next-generation sequencing and high-throughput genotyping processes have resulted in the widespread availability of dense SNP markers, making it the preferred genomic marker to use due to this cost per marker efficiency. Many of the reviewed articles all reported the identification of minimally 10,000 SNP markers through next generation sequencing techniques (Table 2), Huang et al. (2023) identified up to 909,747 high quality SNPs in sugar kelp after genotypic filtering. (estimated genome

size: 537-Mb (Ye et al., 2015)). While the importance of genome size cannot be completely ignored when discussing the importance of marker density, we argue that extreme marker densities may not always be effective in training predictive models. As with training population sample sizes, GS model accuracy has also been found to increase with an increased marker density, before plateauing at an elbow point (Wang et al. 2017). Such a pattern is observed when visualizing the proportion of genetic variance explained with an increasing number of markers (Minamikawa et al. 2017). Again, identification of this elbow point can allow plant breeders to identify the minimally required number of markers to reduce genotyping costs for their programs. In dairy cattle, Zhang et al. (2011) argues that 95% of the prediction accuracy obtained through a high density SNP panel can be obtained with low density SNP panels and present multiple ways to reduce this SNP density. Similar low density SNP panels are recommended by Habier et al. (2009) to maximize cost effectiveness.

Marker selection and filtering have been suggested as means to reduce marker density. Including: LD-pruning, trait association filtering and random sampling. LD-pruning involves a process of removing genetic variants above a certain LD threshold, resulting in a subset of genetically independent variants. These independent variants are representative of the entire set of genetic variants but are represented in a much smaller subset of markers. Tong et al. (2022) applied LD-pruning and selected 13,590 tomato (17% of all SNPs) (estimated genome size: 900 Mb (Tomato Genome Consortium, x, 2012)), and 19,685 pepper (*Capsicum spp*) (25.45% of all SNPs) (estimated genome size: 3.5 Gb (Hulse-Kemp et al. 2018)) tag SNPs. When comparing models trained by the two sets of SNPs, markers trained with the LD-pruned SNPs showed similar accuracies as compared to when using all SNPs. This process can prove to be useful for plant breeders in identifying a smaller, minimal subset of genetic variants to be used in training GS models, further decreasing genotyping costs. Ferrão et al. (2021) further compared GS model accuracy in blueberries (*Vaccinium corymbosum*) with two varying sequencing read depth scenarios (6× and 60x) and showed no significant differences in accuracy between the two. Association filtering can be useful in removing genetic markers that have zero or even negative effects on GS model accuracy. Individual marker effects vary depending on the trait of interest and complexity of trait genetic architecture. In scenarios where traits are contributed by a small number of QTLs, the inclusion of all available genetic marker can often provide little to no additional effect on GS model accuracy. In fact, Minamikawa et al. (2017) observed that traits controlled by an estimated lower number of QTL in citrus required much less marker density in training GS models of the highest accuracy. The process of identifying and filtering markers based on association often requires quantitative genetic processes such as QTL mapping and GWAS to first estimate individual marker effects. Models trained on marker sets containing 215, 20, and 3 GWAS-associated SNPs all showed greater prediction accuracy as compared to 8189 non-filtered SNPs (estimated genome size: 989 Mb (Collins et al. 2019)), as well as smaller standard deviation across 100 test iterations in GS models trained for mildew resistance in spinach (Bhattarai et al. 2022a). Random sampling of genetic variants can be used to reduce marker quantity and identity the minimum number of markers required across the genome to enable high predictive

accuracies. In Nsibi et al. (2020), genetic markers were randomly sampled across the genome and compared to study their effects on GS accuracy for fruit quality traits in apricot (*Prunus armeniaca*). In this study, they demonstrated a mere randomly sampled 10% of 61,030 SNPs (estimated genome size: ~220–230 Mb (Groppi et al. 2021)) were required to train rrBLUP models, any further addition of markers used to train the model did not significantly increase the GS model accuracy. While there is a general positive correlation of marker density with model accuracy, increasing the marker density beyond a certain threshold does not significantly improve model accuracy and genetic gain of the breeding program. Hence, the identification of minimally required markers for effective model training should be analyzed when planning each breeding program to maximize genotyping and phenotyping data.

Genomic selection predictive models

GS methods comprise a wide range of statistical and machine learning techniques that are continuously advancing, utilizing genome wide markers to predict the GEBV of individuals. In our review and subsequent analysis, we categorized these methods into: (1) Bayesian models (43.7%), including Bayes LASSO ($n=20$), Bayes RR ($n=15$), Bayes A ($n=13$) and Bayes B ($n=19$), which utilize Bayesian statistical frameworks to model marker effects and calculate GEBVs (Montesinos López et al. 2022); (2) best linear unbiased prediction (BLUP) methods (30.5%); genomic BLUP (GBLUP) ($n=26$) and its extensions, estimate GEBVs by fitting linear and mixed models (Habier et al. 2013); (3) regression-based methods (2.5%), such as ridge regression (RR) ($n=5$), make use of regression models to predict GEBV while penalizing marker effects (Ogutu et al. 2012); (4) machine learning algorithms (13.7%), including least absolute shrinkage and selection operator (LASSO) ($n=5$), random forest (RF) ($n=9$) elastic net ($n=3$) and support vector machine learning
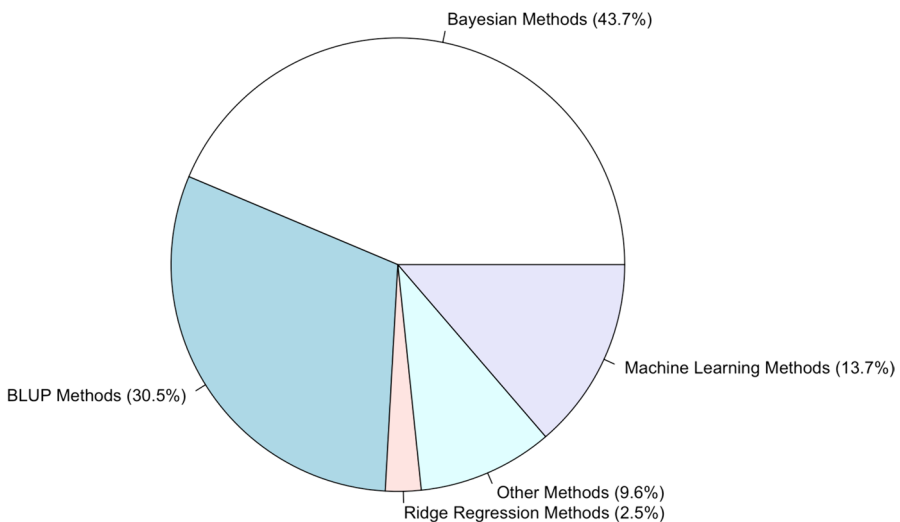
Fig. 4 Pie chart visualizing the frequency of use of GS methods in the articles reviewed

(SVM) ($n=3$), leverage advanced machine learning techniques to model complex relationships between markers and phenotypic traits (Fig. 4). Understanding of statistical and machine learning modelling can be challenging for traditional plant breeders who are not well-versed in the field (Tuggle et al. 2024). The challenge lies in identifying the most suitable model specific to the needs of the breeding program. Understanding the genetic architecture of the trait in question is vital in appropriate selection of GS models. The genetic architecture of a trait, namely the contribution of genes and their interaction to the trait plays an important role in the accuracy of a GS model. GS was quickly suggested to overcome the problems in marker-assisted selection for dealing with complex traits, where many quantitative trait loci (QTL) were found to contribute to the phenotypic trait in question (Bhat et al. 2016). Selection of appropriate GS model and statistical tools can assist in overcoming such problems introduced from trait complexity. For instance, Gezan et al. (2017) compared GS models of 1628 strawberries using GBLUP, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) and Reproducing Kernel Hilbert Spaces (RKHS), finding that Bayes B produced the highest predictive accuracy for yield-related traits. Meuwissen (2009) initially reported the advantages of Bayesian approaches in the prediction of traits controlled by few loci. Compared to regular BLUP models, Bayesian approaches apply differently weighted effects to different loci. In comparison, traditional BLUP methods apply fixed variance effects across all loci and are more suitable when predicting traits affected by many QTLs. Similarly, Minamikawa et al. (2017) showed that GBLUP, RR, and Bayesian regressions were more accurate for most fruit quality traits in citrus, whereas the accuracies of RF, elastic net, and LASSO varied among other traits. Uptake of machine learning models is still low largely due to the lack of reproducibility and interpretability of models when trained with biological data. Machine learning models are often considered 'black boxes' and are notoriously hard to interpret. There are also no evident outperformance of these models as compared to traditional statistical modelling approaches (Chafai et al. 2023). Where conventional machine learning models have struggled to process natural data and required much prior feature engineering, deep learning machine learning models have been able to uncover patterns in data of high dimensionality (LeCun et al. 2015). There have been reported use cases of its effectiveness in human genetics (Leung et al. 2014). In plant genetics, deep learning methods are relatively new. Lozada et al. (2023) reported the use of deep learning models (RF, multilayer perceptron and convolutional neural network models) in comparison to traditional GS models in pepper. However, they noted that there was no pattern identified between model used and accuracy of prediction. They reiterate that deep learning models can be added to the repertoire of available GS models, but large training sample size would be required (Montesinos-López et al. 2021). Initially, most GS models were designed based on additive gene effects, ignoring non-additive effects on the trait. However, non-additive gene effects are increasingly being incorporated into GS models and can provide improved accuracy. Wu et al. (2019) incorporated both additive and dominance gene effects into GS models for pumpkin and regularly showed an improvement in accuracy. They deemed that such models were capable of factoring for the heterozygosity effects of hybrid crossing. Hence, while there has been some attempt at understanding the effect of varying

models within different environmental conditions, traits and population types, it is still largely unexplained. This emphasizes a further research gap on understanding the exact workings and benefits of differing GS models under varying conditions. Tests on eight different pepper traits and twelve models showed minimal differences of performance across models for each phenotype, despite the large differences in models used (Lozada et al. 2023). The same is shown across four different growth periods in strawberries where GS model selected made no apparent difference to prediction accuracy (Gezan et al. 2017). Similar results are observed in aquaculture breeding programs, where systematic review identified no outperforming GS model type (Wang et al. 2022).

Implementation and application of genomic selection into fruit and vegetable breeding programs

Tuggle et al. (2024) performed a survey to understand the pressing and critical needs for advancing agricultural genome-to-phenome research in the USA. Their findings revealed critical needs in funding for "advancing plant and animal breeding", "phenotyping technology development", "advancing genomic (and epigenomic) research" and "predictive analytics development". Despite the abundance of methods and tools available to the breeder, their application remains limited, highlighting the imperative to translate these theoretical methods into tangible gains. GS methods and experiments are often validated using statistical tools such as cross-validation, and adjustment of training and testing population sizes. These tools ensure the reliability and robustness of the trained models. However, the validation using these tools is often theoretical and not realized in practice. Limited research has been performed to apply the GS models into actual breeding schemes and to study the empirical results of these models in their respective crops. Many of the published studies simply focus on GS methodology. In this imperative to study GS methods directly in fruit and vegetable breeding plans to understand their effects against genetic gain. Genetic gain, ΔG, is defined by the increase in mean phenotypic value over time through a selection program.

$$\Delta G = \frac{i \times \sigma A \times rMG}{t}$$

where i is the selection intensity, σA is the genetic variation, rMG is the accuracy of selection and t is the breeding cycle time. In this context, genetic gain is improved where there is an increase in selection intensity, selection accuracy or genetic variation together with the decrease in time taken for each breeding cycle (Xu et al. 2017).

Increasing selection intensity and accuracy using genomic selection in breeding plans

The most obvious implementation of genomic selection into plant breeding programs for fruits and vegetables is the use of GS to select genetically superior parents

more efficiently and accurately as compared to the traditional phenotypic truncation selection method. In silico breeding makes use of available genomic data and crossing simulation, incorporated with GS models to select ideal breeding parents. The GEBV of progenies from the simulated crossings can be calculated within these models to rank parental crossings. These methods increase both selection intensity and accuracy to improve genetic gain. However, they are still largely unknown to traditional breeders but have immense economic benefits for both growers and stakeholders alike. Gezan et al. (2017) selected top strawberry parents by genotypic information without any phenotypic information and showed an efficiency of above 50% when selecting parents this way, as compared to selecting parents with the full repertoire of phenotypic and genotypic data, demonstrating the successes of GS models in selecting ideal individuals as parents based on GEBV. Further, Yamamoto et al. (2021) showed high predictability of F_1 strawberry hybrids when using GS models with observed phenotypes. GS models can be used to simulate and select crossing of parents, and progeny testing can be performed to experimentally validate the results of these simulated crossings. Yamamoto et al. (2017) simulated progenies for four crosses of tomatoes to assess the potential of GS for increasing soluble solids content and total fruit weight. They tested the progeny of these crosses and found that the correlation between the predicted GEBVs and the observed phenotypic values were comparable to the estimated predictability calculated in the cross-validation approach. A similar study performed was by Iwata et al. (2013a, b) based on 84 Japanese pear (*Pyrus pyrifolia*) cultivars, predicting progeny traits based on simulated, in silico crossings. Ferrão et al. (2021) have already begun implementing GS models for GEBV-based parental selection in their blueberry breeding program with the aims of increasing genetic gain as well as maintaining genetic diversity. GS methods were tested in pepper by Hong et al. (2020), who evaluated GS models in a separate population of recombinant inbred lines derived from two parental lines. They reported prediction accuracies ranging from 0.32 to 0.50 for fruit related traits and attributed the moderate accuracy to the large difference in genetic diversity between the training population and the test population. The simulation of progeny trait distributions allows for breeders to select crossings based on desired selection intensity on top of increasing selection accuracy, as demonstrated in some studies. Selection intensity can be selected based on the range of simulated crossings to thus direct the actual crosses. Improvement of both selection intensity and accuracy will assist in increasing genetic gains for fruit and vegetable crops.

Introduction of genetic diversity into breeding plans

To increase genetic gain, improved genetic variation and diversity can also support breeding efforts. Leveraging of germplasm diversity as well as other crop wild type cultivar to improve genetic diversity is well discussed, however the cost of accurate and precise phenotyping still serves as a bottleneck (Voss-Fels et al. 2019). If these "phenotyping bottlenecks" can be overcome, GS methods have great potential to aid in the introgression of exotic alleles into current breeding plans to improve genetic gain. Jiménez et al. (2023) underscored the lack of available genomic diversity towards the resistance of strawberries to Phytophthora crown rot. They hence

implemented diverse genetic cultivar of underutilized germplasm resources to train and implement GS models. The implementation of GS models allows for varying complex selection indexes to be introduced to simulate and hence select the best, long-term breeding model. Serial, truncated selection for genetic gain has often resulted in the reduction of genetic diversity and has been a cause for concern for environmentalists and plant breeders alike (Vanavermaete et al. 2020). In these traditional breeding models, the best genotypes are selected from every cycle and used to breed the next cycle, creating a funneling effect on genetic diversity based on desired traits. There have been several suggested means to deal with the concern of a loss of genetic diversity over breeding generations. Indexes find a compromise between short term genetic gain and genetic diversity loss, but studies have shown that there are long term genetic gains to be made. Cowling et al. (2023) has developed means to overcome the loss of genetic diversity by incorporating such genetic diversity indexes into breeding schemes, known as 'Optimal Contribution Selection'. This selection based on an index of key economic traits increased the rate of genetic gain while minimizing population inbreeding in oilseed rape (*Brassica napus*) over three cycles of serial selection. Sonesson et al. (2012) compared truncation selection against optimum contribution selection and found that pedigree-based inbreeding control was a sufficient measure to deal with the potential inbreeding in a breeding plan. A weighted genomic selection method can also be implemented to upscale rare alleles in the model while retaining long-term genetic gain and a lower inbreeding rate (Beukelaer et al. 2017). However, such concerns may be less valid for hybrid breeding schemes using exotic germplasm resources as breeding material. Given a vegetable breeding scheme where breeders return to a diverse pool of parental genotypes to continually develop new hybrids, the genetic diversity of such a breeding plan should continue to improve with new allele combinations. Provided a plant breeder has access to a diverse, exotic gene pool of plant genotypes within the desired crop, and capabilities to precisely phenotype and genotype it, we recommend the use of such a breeding program to develop new and exciting hybrid cultivars to overcome the decrease in genetic diversity resulting from traditional breeding schemes and uncover new and useful phenotypic traits made possible by use of GS methods.

Increasing genetic gain through reduction of breeding cycle time

Another commonly discussed benefit of the implementation of GS models into breeding plans is the reduction of breeding cycle time, which can greatly improve genetic gain in crops. This is achieved through use of GS models to enables breeders to select ideal genotypes early, without having to wait for plant maturity for phenotypic selection. Such methods provide valuable gains in shortening the breeding cycle, especially for long cycle fruit and vegetable crops such as fruit trees. Kumar et al. (2019) recommended the use of GS models to accelerate the pear breeding cycle by making selections without waiting for extensive fruit-quality phenotyping. However, this same improvement is limited in crops with a short life cycle such as leafy greens. The relatively short and quick breeding cycle of leafy greens means that implementation of GS methods may not immediately result in genetic gain by

reducing breeding cycle time. The time taken to genotype such vegetable crops may even be longer than using traditional phenotypic selection due to the short growth cycle. This may provide an explanation as to the lack of GS implementation and research in such crops. GS performed on spinach measured white rust disease severity a mere 65–70 days after planting (Shi et al. 2022). While GS may not necessarily decrease the breeding cycle time in such an instance, GS methods are extremely useful in reducing the need to further measure such difficult traits that involve long, complicated protocols.

While GS implementation methods vary, it is evident that few studies report the actual application of breeding and the actual realized gain. This may be attributed to the sheer duration taken to cross and phenotype the progeny of the simulated cross before further comparing it to the calculated GEBVs to determine experimentally validated prediction accuracies. Breeders may find it hard to apply such technologies as a result. We believe that it is important to empirically validate GS models on the diversity of fruits and vegetables being studied, and report both their successes and failures for both the scientific community and farmers to learn from. While much focus has been placed into genome-based studies of fruits and vegetables, the translation of GS into real world application has lagged and must be furthered to meet nutritional security needs and overcome climate change problems in agriculture.

Conclusion

To comprehensively identify all relevant studies for this review, the primary and secondary search were performed with appropriate eligibility criteria to identify suitable studies. Studies with insignificant genomic prediction studies and breeding validations are unlikely to be published, resulting in a lack of evidence and bias against understanding the failures of genomic predictions in fruits and vegetables.

This scoping review serves to present an understanding of GS and its applications specifically in fruit and vegetable crops. While such methods are well understood in livestock and other agricultural crops, the same understanding and application is not presented in fruit and vegetable crops, although their breeding is vitally important. We have presented an understanding of the specific fruit and vegetable crops that have been widely studied in relation to GS. There is some pattern between the global value of crop production and their corresponding research publication attention. However, other important vegetable crops such as leafy vegetables have barely been studied with reference to GS methods and can yield important advancements for food and nutritional security. Specific traits studied within crops are largely directed towards fruit and vegetable quality, while less research is performed on plant resource use and other traits that are not directly linked to crop yield or quality. Studies in these traits as well as other seemingly unrelated secondary traits can prove to be vital in providing phenotypic trait understanding and selection of secondary traits to reduce phenotyping costs. GS methods are diverse and have varying effects and accuracies depending on trait heritability and genetic architecture, training population sample size and structure as well as marker density. The selection of appropriate methods and study size decided based on traits of interest as well

as population structure is of utmost importance in maximizing GS accuracy with limited cost. While much research has been reported in advancing and improving GS methods, validation and application of such methods are largely theoretical and seldom proven empirically. Even though the models, methods, and availability of genomic data for GS is rapidly growing, the same is not seen in realized genetic gain and implementation resulting in real world results. We highly recommend an increased effort of research translation into tangible, appliable results, specifically towards crops that hold important positions in both food and nutritional security yet are seldom studied.

Data Availability All data generated or analyzed for this study are included in this article and its supplementary material files. Further inquiries can be directed to the corresponding author (F.T.C.).

Declarations

Conflict of interests F. T. Chew reports grants from Singapore Ministry of Education Academic Research Fund, Singapore Immunology Network, National Medical Research Council (Singapore), Biomedical Research Council (Singapore), National Research Foundation (NRF) (Singapore), Singapore Food Agency (SFA), and the Agency for Science Technology and Research (Singapore), during the conduct of the study; and has received consultancy fees from Sime Darby Technology Centre, First Resources Ltd, Genting Plantation, Olam International, and Syngenta Crop Protection, outside the submitted work. The other authors declare no other competing interests.

References

Amadeu RR, Ferrão LF, Oliveira ID, Benevenuto J, Endelman JB, Munoz PR (2020) Impact of dominance effects on autotetraploid genomic prediction. Crop Sci 60(2):656–665. https://doi.org/10.1002/csc2.20075

Anilkumar C, Sah RP, Muhammed AT, Sunitha NC, Beher S, Marndi BC, Singh AK (2022) Genomic selection in rice: current status and future prospects. Genomic Selection in Plants 68–82. https://doi.org/10.1201/9781003214991-4

Arruda MP, Brown PJ, Lipka AE, Krill AM, Thurber C, Kolb FL (2015) Genomic selection for predicting Fusarium head blight resistance in a wheat breeding program. Plant Genome 8(3):plantgenome2015.2001.0003.

Asseng S, Ewert F, Martre P, Rötter RP, Lobell DB, Cammarano D, Kimball BA, Ottman MJ, Wall GW, White JW, Reynolds MP, Alderman PD, Prasad PVV, Aggarwal PK, Anothai J, Basso B, Biernath C, Challinor AJ, De Sanctis G, . . . Zhu Y (2015) Rising temperatures reduce global wheat production. Nat Clim Change 5(2):143–147. https://doi.org/10.1038/nclimate2470

Bhandari P, Kim J, Lee TG (2023) Genetic architecture of fresh-market tomato yield. BMC Plant Biol 23(1):18

Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK, Singh GP, Prabhu KV (2016) Genomic Selection in the era of next generation sequencing for complex traits in plant breeding. Front Genet 7:221. https://doi.org/10.3389/fgene.2016.00221

Bhattarai G, Olaoye D, Mou B, Correll JC, Shi A (2022a) Mapping and selection of downy mildew resistance in spinach cv. Whale by low coverage whole genome sequencing. Front Plant Sci 13:1012923

Bhattarai G, Shi A, Mou B, Correll JC (2022b) Resequencing worldwide spinach germplasm for identification of field resistance QTLs to downy mildew and assessment of genomic selection methods. Hort Res 9:uhac205

Biscarini F, Nazzicari N, Bink M, Arús P, Aranzana MJ, Verde I, Micali S, Pascal T, Quilot-Turion B, Lambert P, Linge CdS, Pacheco I, Bassi D, Stella A, Rossini L (2017) Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. BMC Genom 18. https://doi.org/10.1186/s12864-017-3781-8

Bongomin O, Lamo J, Guina JM, Okello C, Ocen GG, Obura M, Alibu S, Owino CA, Akwero A, Ojok S (2024) UAV image acquisition and processing for high-throughput phenotyping in agricultural research and breeding programs. Plant Phenome J 7(1):e20096

Brainard SH, Ellison SL, Simon PW, Dawson JC, Goldman IL (2022) Genetic characterization of carrot root shape and size using genome-wide association analysis and genomic-estimated breeding values. Theor Appl Genet 135(2):605–622. https://doi.org/10.1007/s00122-021-03988-8

Brault C, Doligez A, Cunff L, Coupel-Ledru A, Simonneau T, Chiquet J, This P, Flutre T (2021). Harnessing multivariate, penalized regression methods for genomic prediction and QTL detection of drought-related traits in grapevine. G3 11(9). https://doi.org/10.1093/g3journal/jkab248

Brault C, Segura V, This P, Le Cunff L, Flutre T, François P, Pons T, Péros JP, Doligez A (2022) Across-population genomic prediction in grapevine opens up promising prospects for breeding. Hort Res 9. https://doi.org/10.1093/hr/uhac041

Brzozowski LJ, Mazourek M (2020) Evaluation of selection methods for resistance to a specialist insect pest of squash (Cucurbita pepo). Agronomy 10(6):847

Budhlakoti N, Kushwaha AK, Rai A, Chaturvedi K, Kumar A, Pradhan AK, Kumar U, Kumar RR, Juliana P, Mishra D (2022) Genomic selection: A tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. Front Genet 13:832153

Campbell BM, Vermeulen SJ, Aggarwal PK, Corner-Dolloff C, Girvetz E, Loboguerrero AM, Ramirez-Villegas J, Rosenstock T, Sebastian L, Thornton PK (2016) Reducing risks to food security from climate change. Glob Food Sec 11:34–43

Cañas-Gutiérrez GP, Sepulveda-Ortega S, López-Hernández F, Navas-Arboleda AA, Cortés AJ (2022). Inheritance of yield components and morphological traits in Avocado cv. Hass from "Criollo"

"Elite Trees" via Half-Sib Seedling Rootstocks. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.843099

Cazenave X, Petit B, Lateur M, Nybom H, Sedlak J, Tartarini S, Laurens F, Durel C-E, Muranty H (2022) Combining genetic resources and elite material populations to improve the accuracy of genomic prediction in apple. G3 12(3):420

Chafai N, Hayah I, Houaga I, Badaoui B (2023) A review of machine learning models applied to genomic prediction in animal breeding. Front Genet 14:1150596

Cheng CH, Datson PM, Hilario E, Deng CH, Manako KI, McNeilage M, Bomert M, Hoeata K (2019) Genomic predictions in diploid Actinidia chinensis (kiwifruit). Eur J Hort Sci 84(4):213–217. https://doi.org/10.17660/eJHS.2019/84.4.3

Collins K, Zhao K, Jiao C, Xu C, Cai X, Wang X, Ge C, Dai S, Wang Q, Wang Q, Fei Z, Zheng Y (2019) SpinachBase: a central portal for spinach genomics. Database 2019. https://doi.org/10.1093/database/baz072

Corak K, Ellison S, Simon P, Spooner D, Dawson J (2019) Comparison of representative and custom methods of generating core subsets of a carrot germplasm collection. Crop Sci 59(3):1107–1121

Corak KE, Genger RK, Simon PW, Dawson JC (2023) Comparison of genotypic and phenotypic selection of breeding parents in a carrot (Daucus carota) germplasm collection. Crop Sci 63(4):1998–2011

Covarrubias-Pazaran G, Schlautman B, Diaz-Garcia L, Grygleski E, Polashock J, Johnson-Cicalese J, Vorsa N, Iorizzo M, Zalapa J (2018) Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of Vaccinium macrocarpon Ait. Front Plant Sci 9:1310

Cowling WA, Castro-Urrea FA, Stefanova KT, Li L, Banks RG, Saradadevi R, Sass O, Kinghorn BP, Siddique KHM (2023) Optimal contribution selection improves the rate of genetic gain in grain yield and yield stability in Spring Canola in Australia and Canada. Plants 12(2):383. https://www.mdpi.com/2223-7747/12/2/383

da Silva FA, Viana AP, Correa CCG, Santos EA, de Oliveira JAVS, Andrade JDG, Ribeiro RM, Glória LS (2021) Bayesian ridge regression shows the best fit for SSR markers in Psidium guajava among Bayesian models. Sci Rep 11(1):13639

de Bem Oliveira I, Resende MF Jr, Ferrão LFV, Amadeu RR, Endelman JB, Kirst M, Coelho AS, Munoz PR (2019) Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. G3 9(4):1189–1198. https://doi.org/10.1534/g3.119.400059

De Beukelaer H, Badke Y, Fack V, De Meyer G (2017) Moving Beyond Managing Realized Genomic Relationship in Long-Term Genomic Selection. Genetics 206(2):1127–1138. https://doi.org/10.1534/genetics.116.194449

de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193(2):327–345. https://doi.org/10.1534/genetics.112.143313

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19(9):592–601

Duangjit J, Causse M, Sauvage C (2016) Efficiency of genomic selection for tomato fruit quality. Mol Breeding 36:1–16

Edwards SM, Buntjer JB, Jackson R, Bentley AR, Lage J, Byrne E, Burt C, Jack P, Berry S, Flatman E, Poupard B, Smith S, Hayes C, Gaynor RC, Gorjanc G, Howell P, Ober E, Mackay IJ, Hickey JM (2019) The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet 132(7):1943–1952. https://doi.org/10.1007/s00122-019-03327-y

FAO and Ministry of Social Development and Family of Chile (2021) Promoting safe and adequate fruit and vegetable consumption to improve health. https://doi.org/10.4060/cb7946en

Feldmann MJ, Pincot DD, Vachev MV, Famula RA, Cole GS, Knapp SJ (2023) Accelerating genetic gains for quantitative resistance to Verticillium wilt through predictive breeding in strawberry. Plant Genome 17:e20405

Fernández-González J, Akdemir D, Isidroy Sánchez J (2023) A comparison of methods for training population optimization in genomic selection. Theoret Appl Genet 136(3):30. https://doi.org/10.1007/s00122-023-04265-6

Ferrão LF, Amadeu RR, Benevenuto J, de Bem Oliveira I, Munoz PR (2021) Genomic selection in an outcrossing autotetraploid fruit crop: lessons from blueberry breeding. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.676326

Fodor A, Segura V, Denis M, Neuenschwander S, Fournier-Level A, Chatelet P, Homa FAA, Lacombe T, This P, Le Cunff L (2014) Genome-wide prediction methods in highly diverse and heterozygous species: proof-of-concept through simulation in grapevine. PLoS ONE 9(11):e110436

Food, & Agriculture Organization of the United, N (1997) FAOSTAT statistical database. In [Rome]: FAO, c1997. https://search.library.wisc.edu/catalog/999890171702121

Fu W, da Silva Linge C, Lawton JM, Gasic K (2022) Feasibility of genomic prediction for brown rot (Monilinia spp.) resistance in peach. Fruit Res 2(1):1–12

Gezan SA, Osorio LF, Verma S, Whitaker VM (2017) An experimental validation of genomic selection in octoploid strawberry. Hort Res 4(1):16070. https://doi.org/10.1038/hortres.2016.70

Gois IB, Borém A, Cristofani-Yaly M, de Resende MDV, Azevedo CF, Bastianel M, Novelli VM, Machado MA (2016) Genome wide selection in citrus breeding. Genet Mol Res 15(4). https://doi.org/10.4238/gmr15048863

Groppi A, Liu S, Cornille A, Decroocq S, Bui QT, Tricon D, Cruaud C, Arribat S, Belser C, Marande W, Salse J, Huneau C, Rodde N, Rhalloussi W, Cauet S, Istace B, Denis E, Carrère S, Audergon J-M, . . . Decroocq V (2021) Population genomics of apricots unravels domestication history and adaptive events. Nat Commun 12(1):3956. https://doi.org/10.1038/s41467-021-24283-6

Gutierrez-Reinoso MA, Aponte PM, Garcia-Herreros M (2021) Genomic analysis, progress and future perspectives in dairy cattle selection: a review. Animals 11(3):599. https://www.mdpi.com/2076-2615/11/3/599

Habier D, Fernando RL, Dekkers JC (2009) Genomic selection using low-density marker panels. Genetics 182(1):343–353. https://doi.org/10.1534/genetics.108.100289

Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics 194(3):597–607. https://doi.org/10.1534/genetics.113.152207

Hardigan MA, Feldmann MJ, Carling J, Zhu A, Kilian A, Famula RA, Cole GS, Knapp SJ (2023) A medium-density genotyping platform for cultivated strawberry using DArTag technology. Plant Genome 16(4):e20399

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92(2):433–443

Hernandez CO, Wyatt LE, Mazourek MR (2020) Genomic prediction and selection for fruit traits in winter squash. G3 10(10):3601–3610. https://doi.org/10.1534/g3.120.401215

Hernández-Bautista A, Lobato-Ortiz R, García-Zavala JJ, Cruz-Izquierdo S, Chávez-Servia JL, Rocandio-Rodríguez M, Moreno-Ramírez YDR, Hernandez-Leal E, Hernández-Rodríguez M, Reyes-Lopez D (2020) Prediction accuracy of genomic selection models for earliness in tomato. Chil J Agric Res 80(4):505–514

Hernández-Bautista A, Lobato-Ortiz R, García-Zavala JJ, Parra-Gómez MA, Cadeza-Espinosa M, Canela-Doñan D, Cruz-Izquierdo S, Chávez-Servia JL (2016) Implications of genomic selection for obtaining F2: 3 families of tomato. Sci Hortic 207:7–13

Hong J-P, Ro N, Lee H-Y, Kim GW, Kwon J-K, Yamamoto E, Kang B-C (2020) Genomic Selection for Prediction of Fruit-Related Traits in Pepper (Capsicum spp.). Front Plant Sci 11. https://doi.org/10.3389/fpls.2020.570871

Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, Selly SLC, Martin SAM, Stevens JR, Santos EM, Davie A, Robledo D (2020) Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet 21(7):389–409. https://doi.org/10.1038/s41576-020-0227-y

Huang M, Robbins KR, Li Y, Umanzor S, Marty-Rivera M, Bailey D, Aydlett M, Schmutz J, Grimwood J, Yarish C (2023) Genomic selection in algae with biphasic lifecycles: a Saccharina latissima (sugar kelp) case study. Front Mar Sci 10:1040979

Hulse-Kemp AM, Maheshwari S, Stoffel K, Hill TA, Jaffe D, Williams SR, Weisenfeld N, Ramakrishnan S, Kumar V, Shah P, Schatz MC, Church DM, Van Deynze A (2018) Reference quality assembly of the 3.5-Gb genome of Capsicum annuum from a single linked-read library. Hortic Res 5:4. https://doi.org/10.1038/s41438-017-0011-0

Ingram J (2020) Nutrition security is more than food security. Nature Food 1(1):2–2. https://doi.org/10.1038/s43016-019-0002-4

Iwata H, Hayashi T, Terakami S, Takada N, Saito T, Yamamoto T (2013a) Genomic prediction of trait segregation in a progeny population: a case study of Japanese pear (Pyrus pyrifolia). BMC Genet 14(1):81. https://doi.org/10.1186/1471-2156-14-81

Iwata H, Hayashi T, Terakami S, Takada N, Sawamura Y, Yamamoto T (2013b) Potential assessment of genome-wide association study and genomic selection in Japanese pear Pyrus pyrifolia. Breed Sci 63(1):125–140

Jiménez NP, Feldmann MJ, Famula RA, Pincot DD, Bjornson M, Cole GS, Knapp SJ (2023) Harnessing underutilized gene bank diversity and genomic prediction of cross usefulness to enhance resistance to Phytophthora cactorum in strawberry. Plant Genome 16(1):e20275

Jung M, Keller B, Roth M, Aranzana MJ, Auwerkerken A, Guerra W, Al-Rifaï M, Lewandowski M, Sanin N, Rymenants M (2022) Genetic architecture and genomic predictive ability of apple quantitative traits across environments. Hort Res 9:uhac028

Jung M, Roth M, Aranzana MJ, Auwerkerken A, Bink M, Denancé C, Dujak C, Durel CE, Font IFC, Cantin CM, Guerra W, Howard NP, Keller B, Lewandowski M, Ordidge M, Rymenants M, Sanin N, Studer B, Zurawicz E, . . . Muranty H (2020) The apple REFPOP-a reference population for genomics-assisted breeding in apple. Hortic Res 7(1):189. https://doi.org/10.1038/s41438-020-00408-8

Kostick SA, Bernardo R, Luby JJ (2023) Genomewide selection for fruit quality traits in apple: breeding insights gained from prediction and postdiction. Hort Res 10(6):uhad088

Kumar A, Mir RR, Sehgal D, Agarwal P, Carter A (2021) Genetics and genomics to enhance crop production, towards food security. Front Genet 12:798308

Kumar S, Hilario E, Deng CH, Molloy C (2020) Turbocharging introgression breeding of perennial fruit crops: a case study on apple. Hort Res 7:47

Kumar S, Kirk C, Deng CH, Shirtliff A, Wiedow C, Qin M, Wu J, Brewer L (2019) Marker-trait associations and genomic predictions of interspecific pear (Pyrus) fruit characteristics. Sci Rep 9(1):9072. https://doi.org/10.1038/s41598-019-45618-w

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539

Leung MK, Xiong HY, Lee LJ, Frey BJ (2014) Deep learning of the tissue-regulated splicing code. Bioinformatics 30(12):i121-129. https://doi.org/10.1093/bioinformatics/btu277

Li B, Cockerton HM, Johnson AW, Karlström A, Stavridou E, Deakin G, Harrison RJ (2020) Defining strawberry shape uniformity using 3D imaging and genetic mapping. Hortic Res 7:115. https://doi.org/10.1038/s41438-020-0337-x

Li X, Wang J, Su M, Zhang M, Hu Y, Du J, Zhou H, Yang X, Zhang X, Jia H, Gao Z (2023) Multiple-statistical genome-wide association analysis and genomic prediction of fruit aroma and agronomic traits in peaches. Hort Res 10(7). https://doi.org/10.1093/hr/uhad117

Liabeuf D, Francis DM (2017) The use of historical datasets to develop multi-trait selection models in processing tomato. Euphytica 213(5):100

Liabeuf D, Sim S-C, Francis DM (2018) Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. Phytopathology 108(3):392–401

Longhi S, Cappellin L, Guerra W, Costa F (2013) Validation of a functional molecular marker suitable for marker-assisted breeding for fruit texture in apple (Malus× domestica Borkh.). Mol Breeding 32(4):841–852

Lozada DN, Sandhu KS, Bhatta M (2023) Ridge regression and deep learning models for genome-wide selection of complex traits in New Mexican Chile peppers. BMC Genom Data 24(1):80

McClure KA, Gardner KM, Douglas GM, Song J, Forney CF, DeLong J, Fan L, Du L, Toivonen PM, Somers DJ (2018) A genome-wide association study of apple quality and scab resistance. Plant Genome 11(1):170075

Mertten D, Baldwin S, Cheng CH, McCallum J, Thomson S, Ashton DT, McKenzie CM, Lenhard M, Datson PM (2023) Implementation of different relationship estimate methodologies in breeding value prediction in kiwiberry (Actinidia arguta). Mol Breeding 43(10):75. https://doi.org/10.1007/s11032-023-01419-8

Meuwissen TH, Hayes BJ, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Meuwissen THE (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet Sel Evol 41(1):35. https://doi.org/10.1186/1297-9686-41-35

Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T, Imai A, Hamada H, Hayashi T, Matsumoto S, Katayose Y, Toyoda A, Fujiyama A, Nakamura Y, Shimizu T, Iwata H (2017) Genome-wide association study and genomic prediction in citrus: Potential

of genomics-assisted breeding for fruit quality traits. Sci Rep 7(1):4721. https://doi.org/10.1038/s41598-017-05100-x

Minamikawa MF, Takada N, Terakami S, Saito T, Onogi A, Kajiya-Kanegae H, Hayashi T, Yamamoto T, Iwata H (2018) Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (Pyrus pyrifolia Nakai). Sci Rep 8(1):11994

Montesinos López OA, Montesinos López A, Crossa J (2022) Bayesian genomic linear regression. In: Multivariate statistical machine learning methods for genomic prediction. Springer, Cham. https://doi.org/10.1007/978-3-030-89010-0_6

Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, Barrón-López JA, Martini JWR, Fajardo-Flores SB, Gaytan-Lugo LS, Santana-Mancilla PC, Crossa J (2021) A review of deep learning applications for genomic selection. BMC Genom 22(1):19. https://doi.org/10.1186/s12864-020-07319-x

Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J (2016) The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol 17:1–9

Muranty H, Troggio M, Sadok IB, Rifaï MA, Auwerkerken A, Banchi E, Velasco R, Stevanato P, van de Weg WE, Di Guardo M, Kumar S, Laurens F, Bink MC (2015) Accuracy and responses of genomic selection on key traits in apple breeding. Horticul Res 2:15060. https://doi.org/10.1038/hortres.2015.60

Nagano S, Shirasawa K, Maeda F, Watanabe M, Noguchi Y, Kataoka S, Wada T, Oku K, Mori M, Tasaki K, Iimura K, Nakaya A, Yanagi T, Hirakawa H, Isobe S (2018) Challenge to genomic selection in strawberry at four breeding stations in Japan. Acta Hortic 1203:1–8. https://doi.org/10.17660/ActaHortic.2018.1203.1

Nishio S, Hayashi T, Shirasawa K, Saito T, Terakami S, Takada N, Takeuchi Y, Moriya S, Itai A (2021) Genome-wide association study of individual sugar content in fruit of Japanese pear (Pyrus spp.). BMC Plant Biol 21(1):378. https://doi.org/10.1186/s12870-021-03130-2

Nsibi M, Gouble B, Bureau S, Flutre T, Sauvage C, Audergon J-M, Regnard J-L (2020) Adoption and optimization of genomic selection to sustain breeding for apricot fruit quality. G3 10(12):4513–4529. https://doi.org/10.1534/g3.120.401452

Nyine M, Uwimana B, Blavet N, Hřibová E, Vanrespaille H, Batte M, Akech V, Brown A, Lorenzen J, Swennen R, Doležel J (2018) Genomic Prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. Plant Genome 11(2). https://doi.org/10.3835/plantgenome2017.10.0090

Nyine M, Uwimana B, Swennen R, Batte M, Brown A, Christelová P, Hřibová E, Lorenzen J, Doležel J (2017) Trait variation and genetic diversity in a banana genomic selection training population. PLoS ONE 12(6):e0178734

Ogutu JO, Schulz-Streeck T, Piepho H-P (2012) Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC Proc 6(2):S10. https://doi.org/10.1186/1753-6561-6-S2-S10

Ohyama A, Matsunaga H, Kawasaki Y, Shirasawa K, Negoro S, Miyatake K, Yamaguchi H, Nunome T, Iwata H, Fukuoka H (2023) Bayesian estimation of multi-allele QTLs for agricultural traits in tomato using recombinant inbred lines derived from two F1 hybrid cultivars. Euphytica 219(1):21

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, . . . Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj, 372:n71. https://doi.org/10.1136/bmj.n71

Petrasch S, Mesquida-Pesci SD, Pincot DD, Feldmann MJ, López CM, Famula R, Hardigan MA, Cole GS, Knapp SJ, Blanco-Ulate B (2022) Genomic prediction of strawberry resistance to postharvest fruit decay caused by the fungal pathogen Botrytis cinerea. G3 12(1):378

Piaskowski JH, Hardner C, Cai L, Zhao YY, Iezzoni A, Peace C (2018) Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. BMC Genet 19. https://doi.org/10.1186/s12863-018-0609-8

Pincot DD, Hardigan MA, Cole GS, Famula RA, Henry PM, Gordon TR, Knapp SJ (2020) Accuracy of genomic selection and long-term genetic gain for resistance to Verticillium wilt in strawberry. Plant Genome 13(3):e20054

Rosen A, Hasan Y, Briggs W, Uptmoor R (2018) Genome-based prediction of time to curd induction in cauliflower. Front Plant Sci 9:256209

Roth M, Muranty H, Di Guardo M, Guerra W, Patocchi A, Costa F (2020) Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. Horticul Res 7:148. https://doi.org/10.1038/s41438-020-00370-5

Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, Reynolds M, Singh R (2016) Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3 6(9):2799–2808

Sekine D, Yabe S (2020) Simulation-based optimization of genomic selection scheme for accelerating genetic gain while preventing inbreeding depression in onion breeding. Breed Sci 70(5):594–604

Sheikh M, Iqra F, Ambreen H, Pravin KA, Ikra M, Chung YS (2024) Integrating artificial intelligence and high-throughput phenotyping for crop improvement. J Integr Agric 23(6):1787–1802. https://doi.org/10.1016/j.jia.2023.10.019

Shi A, Bhattarai G, Xiong H, Avila CA, Feng C, Liu B, Joshi V, Stein L, Mou B, du Toit LJ (2022) Genome-wide association study and genomic prediction of white rust resistance in USDA GRIN spinach germplasm. Hort Res 9:uhac069

Sonesson AK, Woolliams JA, Meuwissen THE (2012) Genomic selection requires genomic control of inbreeding. Genet Sel Evol 44(1):27. https://doi.org/10.1186/1297-9686-44-27

Stévant P, Marfaing H, Duinker A, Fleurence J, Rustad T, Sandbakken I, Chapman A (2018) Biomass soaking treatments to reduce potentially undesirable compounds in the edible seaweeds sugar kelp (Saccharina latissima) and winged kelp (Alaria esculenta) and health risk estimation for human consumption. J Appl Phycol 30(3):2047–2060. https://doi.org/10.1007/s10811-017-1343-8

Sverrisdóttir E, Sundmark EHR, Johnsen H, Kirk HG, Asp T, Janss L, Bryan G, Nielsen KL (2018) The Value of Expanding the Training Population to Improve Genomic Selection Models in Tetraploid Potato. Front Plant Sci 9:1118. https://doi.org/10.3389/fpls.2018.01118

Thorwarth P, Yousef EA, Schmid KJ (2018) Genomic prediction and association mapping of curd-related traits in gene bank accessions of cauliflower. G3 8(2):707–718

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci 108(50):20260–20264

Tomato Genome Consortium, x (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485(7400):635

Tong H, Nankar AN, Liu J, Todorova V, Ganeva D, Grozeva S, Tringovska I, Pasev G, Radeva-Ivanova V, Gechev T, Kostova D, Nikoloski Z (2022) Genomic prediction of morphometric and colorimetric traits in Solanaceous fruits. Hortic Res 9:uhac072. https://doi.org/10.1093/hr/uhac072

Tuggle CK, Clarke JL, Murdoch BM, Lyons E, Scott NM, Beneš B, Campbell JD, Chung H, Daigle CL, Das Choudhury S (2024) Current challenges and future of agricultural genomes to phenomes in the USA. Genome Biol 25(1):8

Vanavermaete D, Fostier J, Maenhout S, Debaets B (2020) Preservation of Genetic Variation in a Breeding Population for Long-Term Genetic Gain. G3 (Bethesda) 10:2753–2762. https://doi.org/10.1534/g3.120.401354

Viana AP, de Lima e Silva FH, Gloria LS, Ribeiro RM, Krause W, Boechat MSB (2017) Implementing genomic selection in sour passion fruit population. Euphytica 213:1–13

Viana AP, Resende MDVd, Riaz S, Walker MA (2016) Genome selection in fruit breeding: application to table grapes. Sci Agric 73:142–149

Voss-Fels KP, Cooper M, Hayes BJ (2019) Accelerating crop genetic gains with genomic selection. Theor Appl Genet 132(3):669–686. https://doi.org/10.1007/s00122-018-3270-8

Wang K, Yang B, Li Q, Liu S (2022) Systematic evaluation of genomic prediction algorithms for genomic prediction and breeding of aquatic animals. Genes (Basel) 13(12). https://doi.org/10.3390/genes13122247

Wang Q, Yu Y, Yuan J, Zhang X, Huang H, Li F, Xiang J (2017) Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp Litopenaeus vannamei. BMC Genet 18(1):45. https://doi.org/10.1186/s12863-017-0507-5

Weller JI, Ezra E, Ron M (2017) Invited review: a perspective on the future of genomic selection in dairy cattle. J Dairy Sci 100(11):8633–8644. https://doi.org/10.3168/jds.2017-12879

Who J, Consultation FE (2003) Diet, nutrition and the prevention of chronic diseases. World Health Organ Tech Rep Ser 916(i-viii), 1–149.

Wilkinson MJ, Yamashita R, James ME, Bally IS, Dillon NL, Ali A, Hardner CM, Ortiz-Barrientos D (2022) The influence of genetic structure on phenotypic diversity in the Australian mango (Mangifera indica) gene pool. Sci Rep 12(1):20614

Wu B, Shen F, Wang X, Zheng WY, Xiao C, Deng Y, Wang T, Yu Huang Z, Zhou Q, Wang Y (2021) Role of MdERF3 and MdERF118 natural variations in apple flesh firmness/crispness retainability and development of QTL-based genomics-assisted prediction. Plant Biotechnol J 19(5):1022–1037

Wu PY, Tung CW, Lee CY, Liao CT (2019) Genomic prediction of pumpkin hybrid performance. Plant Genome 12(2):180082

Xu Y, Li P, Zou C, Lu Y, Xie C, Zhang X, Prasanna BM, Olsen MS (2017) Enhancing genetic gain in the era of molecular breeding. J Exp Bot 68(11):2641–2666

Xu Y, Liu X, Fu J, Wang H, Wang J, Huang C, Prasanna BM, Olsen MS, Wang G, Zhang A (2020) Enhancing Genetic Gain through Genomic Selection: From Livestock to Plants. Plant Commun 1(1):100005. https://doi.org/10.1016/j.xplc.2019.100005

Xu Y, Ma K, Zhao Y, Wang X, Zhou K, Yu G, Li C, Li P, Yang Z, Xu C (2021) Genomic selection: a breakthrough technology in rice breeding. Crop J 9(3):669–677

Yamamoto E, Kataoka S, Shirasawa K, Noguchi Y, Isobe S (2021) Genomic Selection for F1 Hybrid Breeding in Strawberry (Fragaria × ananassa) [Original Research]. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.645111

Yamamoto E, Matsunaga H, Onogi A, Ohyama A, Miyatake K, Yamaguchi H, Nunome T, Iwata H, Fukuoka H (2017) Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity 118(2):202–209. https://doi.org/10.1038/hdy.2016.84

Yamamoto T (2021) Dna markers and molecular breeding in pear and other rosaceae fruit trees. Hort J 90(1):1–13

Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J (2020) Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. Mol Plant 13(2):187–214

Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y, Wang Y, Shi W, Ji P, Li D, Guan Z, Shao C, Zhuang Z, Gao Z, Qi J, Zhao F (2015) Saccharina genomes provide novel insight into kelp biology. Nat Commun 6:6986. https://doi.org/10.1038/ncomms7986

Zhang Z, Ding X, Liu J, Zhang Q, de Koning DJ (2011) Accuracy of genomic prediction using low-density marker panels. J Dairy Sci 94(7):3642–3650. https://doi.org/10.3168/jds.2010-3917