

RESEARCH ARTICLE

Open Access

Genomic selection and genetic gain for nut yield in an Australian macadamia breeding population



Katie M. O'Connor^{1,2*} , Ben J. Hayes³, Craig M. Hardner³, Mobashwer Alam², Robert J. Henry³ and Bruce L. Topp²

Abstract

Background: Improving yield prediction and selection efficiency is critical for tree breeding. This is vital for macadamia trees with the time from crossing to production of new cultivars being almost a quarter of a century. Genomic selection (GS) is a useful tool in plant breeding, particularly with perennial trees, contributing to an increased rate of genetic gain and reducing the length of the breeding cycle. We investigated the potential of using GS methods to increase genetic gain and accelerate selection efficiency in the Australian macadamia breeding program with comparison to traditional breeding methods. This study evaluated the prediction accuracy of GS in a macadamia breeding population of 295 full-sib progeny from 32 families (29 parents, reciprocals combined), along with a subset of parents. Historical yield data for tree ages 5 to 8 years were used in the study, along with a set of 4113 SNP markers. The traits of focus were average nut yield from tree ages 5 to 8 years and yield stability, measured as the standard deviation of yield over these 4 years. GBLUP GS models were used to obtain genomic estimated breeding values for each genotype, with a five-fold cross-validation method and two techniques: prediction across related populations and prediction across unrelated populations.

Results: Narrow-sense heritability of yield and yield stability was low ($h^2 = 0.30$ and 0.04 , respectively). Prediction accuracy for yield was 0.57 for predictions across related populations and 0.14 when predicted across unrelated populations. Accuracy of prediction of yield stability was high ($r = 0.79$) for predictions across related populations. Predicted genetic gain of yield using GS in related populations was 474 g/year, more than double that of traditional breeding methods (226 g/year), due to the halving of generation length from 8 to 4 years.

Conclusions: The results of this study indicate that the incorporation of GS for yield into the Australian macadamia breeding program may accelerate genetic gain due to reduction in generation length, though the cost of genotyping appears to be a constraint at present.

Keywords: Horticulture, Plant breeding, Genome-based prediction, Phenotype, Fruit tree

* Correspondence: katie.oconnor@daf.qld.gov.au

¹Queensland Department of Agriculture and Fisheries, Maroochy Research Facility, 47 Mayers Road, Nambour, QLD 4560, Australia

²Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Maroochy Research Facility, 47 Mayers Road, Nambour, QLD 4560, Australia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Nut yield is the most economically important selection trait of macadamia [1]. In 2017, the Australian industry—the world's largest—produced a crop of 46,000 t of nut-in-shell [2]. Although nut yield is the main trait of focus when selecting new macadamia varieties, it is expensive and difficult to assess in breeding. Nuts are comprised of an outer pericarp green husk, a hard shell testa, and an internal edible kernel. The husk either abscises from the tree along with the nut-in-shell (NIS), or dehisces (splits along a single suture) and the NIS falls to the ground [1]. After harvest, nuts are dehusked mechanically. Yield measurements are usually expressed as NIS or kernel yield per tree [1]. Yield is a complex trait affected by many processes and environmental influences, and is likely controlled by many genes [3, 4]. Previous estimates of yield heritability in macadamia are low (< 0.20) [5], indicating that yield is highly likely to be controlled by many loci of small effect. As such, selection for high yield is often made difficult by environmental and genotype \times environment interaction ($G \times E$) effects [6]. $G \times E$ has been previously documented in macadamia yield [5, 7], though this appeared to be due to a particular characteristic at a particular location, and no work has yet been conducted to understand the repeatable factors behind $G \times E$ for yield.

In addition to increased yield, precocious cultivars—those that produce nuts at an early age—may be commercially attractive due to cash flow at an early age. However, it is not yet known how precocity might affect the rate at which yield begins to plateau in macadamia varieties. In coffee and apple, early-yielding varieties are desirable, particularly those with stable yields over time [8, 9]. For perennial horticulture crops like macadamia, yield stability may be defined as the consistency of yield of individual trees across consecutive years [10]. Unstable yields, due to alternate bearing, is common in some perennial fruit crops and is undesirable as regular income is vital for growers [9, 10]. Research regarding genetic architecture surrounding consistency of yield over years has been limited outside of biennial bearing in apple (e.g. 11, 12). Yield stability is considered an important trait in macadamia by industry [1]. Some macadamia growers report biennial bearing in certain cultivars, such as 'H2' and '344', which can be problematic.

Selection of new macadamia varieties involves two stages: thousands of seedlings are produced by cross-pollination to create diversity and are assessed in an unreplicated seedling progeny trial (SPT) (sometimes across multiple sites due to space restrictions), then the best performing trees are clonally propagated and evaluated in replicated trials across multiple environments in a candidate cultivar regional variety trial (RVT) [11]. Trees begin to flower and bear fruit around 4 to 5 years

after planting, and yield is evaluated for at least another 4 years [5]. Due to the crop's long juvenile stage and the need to assess yield over several years to increase the accuracy of predicting performance, traditional breeding has a selection cycle of almost a quarter of a century (22 years) [1, 12, 13]. Candidates are then selected for commercial release using a selection index including traits such as yield, kernel recovery (the ratio of kernel to NIS weight; KR), precocity and tree size [12]. Alternative selection strategies are sought to shorten the selection cycle and increase genetic gain.

Genomic selection is a form of marker-assisted selection (MAS) that utilises genome-wide markers to predict genomic estimated breeding values (GEBVs) of individuals, after which the best performers are selected, possibly without phenotyping candidates [14, 15]. As GEBVs can be predicted for individuals at the seedling stage, early selection for elite individuals is possible, thus greatly reducing the selection cycle [15, 16]. GS uses a training or reference population of individuals with known genotypes and phenotypes to construct a model of each marker's effect on the trait. To estimate accuracy of prediction, the model is then applied to predict the GEBV of individuals in a validation population, for which measured phenotypes are available. The accuracy of prediction is determined by the correlation between GEBVs and phenotypic observations as a proxy for the unknown true genetic values. MAS can also be conducted using genetic markers of large effect detected through genome-wide association studies (GWAS). GWAS has been conducted in macadamia for nut and kernel traits, but not for yield [17, 18].

Genomic selection was first used in dairy cattle and is being increasingly used to improve genetic gain in both animal and plant breeding programs. With the potential to shorten breeding cycles, long-lived species with slow maturation times may have the most to gain from MAS and GS [19, 20]. Grattapaglia [21] and Lin, Hayes [22] have extensive reviews on the use of GS in forestry and annual species, respectively. The main attraction of GS for perennial crops may be that it can accelerate breeding cycles, thereby increasing the gain per unit time and reducing field trial costs [3, 23, 24]. Sweet cherry [25], peach [26], oil palm [27–29], citrus [30], apple [31, 32], and pear [33, 34] researchers have evaluated the use of GS to increase genetic gain in their breeding programs. A recent study in Japanese chestnut [35] achieved high prediction accuracies for harvest date ($r = 0.84$) and insect infestation (0.60), though yield was not studied.

High prediction accuracy of GS models will improve confidence in selecting elite candidates. Prediction accuracy depends on many factors, including the model, crop, size of the reference population, extent of linkage disequilibrium (LD), marker set, and heritability of the

trait of interest [36]. Genetic markers should be in high LD with the genes controlling the trait, in order to capture the genetic variance [14, 37, 38]. In a simulation using animal data, Calus et al. [39] suggested that models using marker densities of LD $r^2 = 0.2$ (average distance of 0.128 cM between markers) were superior to those at lower densities. Accurate phenotyping of a large training population, preferably over multiple environments and years (allowing for the study of multiple seasons and tree ages), is required for perennial crops to derive accurate predictions due to the interactions between these factors [23, 40–42].

Recently, an updated version of the *M. integrifolia* genome (v2) was published with 4098 scaffolds anchored to 14 pseudo-chromosomes (745 Mb, N50 413 kb) [43]. In a study using 4113 SNP markers, of which 90% mapped to v2 genome scaffolds, O'Connor, Kilian [44] found that LD decayed rapidly over short distances of the genome. Here, using these same 4113 SNP markers, we explore the potential of GS in macadamia breeding, examining the contribution to genetic gains relative to phenotypic- and pedigree-based selection due to a substantial reduction in generation length. This study aimed to: (i) determine the prediction accuracy of GBLUP (genomic best linear unbiased prediction) methods in predicting GEBVs for nut yield and yield stability across years; (ii) estimate genetic gain using GS strategies compared with traditional breeding methods; and (iii) discuss potential strategies in which GS can be employed to increase genetic gain in macadamia breeding programs. This research is the first study to utilise molecular marker technology for GS in macadamia and, to our knowledge, the first to use GS to predict yield stability over consecutive years for a fruit or nut tree crop.

Results

Heritability and accuracy of prediction models

Narrow-sense heritability for yield in the study population was 0.30 ± 0.08 . For yield stability across 4 years, heritability was close to zero (Table 1). Variance components, from which estimates of heritability were based, are included in the [Supplementary Materials](#) (Supplementary Tables 1 and 2).

Moderate prediction accuracy was achieved for yield from cross-validation (CV) using randomly masked individuals (prediction across related populations; 0.57 ± 0.11). In comparison, yield prediction accuracy was not

Table 1 Narrow-sense heritability (h^2) and standard errors (SE) for yield and yield stability

Trait	h^2	SE
Yield	0.30	0.08
Yield stability	0.04	0.04

significantly different from zero for prediction across unrelated populations where families were grouped (0.14 ± 0.14 ; Table 2). Boxplots showing the observed relationship distributions from the GRM used in predictions are included in the [Supplementary Materials](#) (Supplementary Figs. 1 and 2).

For yield stability, high prediction accuracy was achieved for randomly-grouped individuals (0.79 ± 0.23 , $p < 0.01$). However, when families were grouped, prediction accuracy was not significantly different from zero (0.28 ± 0.18 ; Table 3).

Comparison of breeding strategies and genetic gain

Two breeding strategies were compared to demonstrate how implementing GS could decrease the breeding cycle and subsequently increase genetic gain (Table 4). The number of trees involved in each stage and specific costs are excluded (given uncertainties of, and constantly evolving, genotyping costs).

1. Traditional breeding: Progeny are evaluated in a SPT for at least 8 years to select individuals with elite clonal values for yield and other economically important traits (such as KR, precocity, and tree size) using a selection index. SPT is then followed by a RVT for at least 8 years, where selected elites are clonally propagated and evaluated for more economically important traits across multiple environments.
2. Genomic selection: After germination, the first leaves of each progeny seedling are genotyped for a

Table 2 Predictive ability and prediction accuracy for yield for each of the five cross-validation (CV) sets for random and family groupings of individuals, and mean and standard error (SE) for each grouping

Grouping	CV set	Predictive Ability	Prediction Accuracy (r)
Random	1	0.36	0.66
	2	0.43	0.79
	3	0.30	0.56
	4	0.08	0.16
	5	0.39	0.71
	Mean		
	SE		0.11
Family	1	-0.16	-0.29
	2	0.15	0.28
	3	-0.03	-0.05
	4	0.28	0.51
	5	0.15	0.27
	Mean		
	SE		0.14

Results of t-test: ** $p < 0.01$, NS indicates not significantly different from zero

Table 3 Predictive ability and prediction accuracy for yield stability for each of the five cross-validation (CV) sets for random and family groupings of individuals, and mean and standard error (SE) for each grouping

Grouping	CV set	Predictive Ability	Prediction Accuracy (r)
Random	1	0.17	0.86
	2	0.08	0.38
	3	0.29	1.46
	4	0.03	0.15
	5	0.22	1.11
		Mean	
	SE		0.23
Family	1	0.09	0.43
	2	0.12	0.60
	3	0.06	0.30
	4	-0.09	-0.43
	5	0.10	0.50
		Mean	
	SE		0.18

Results of t-test: * $p < 0.05$, NS indicates not significantly different from zero

large number of markers. Genomic prediction is then used to predict GEBVs for yield (and other traits, not shown here). Elite candidates are selected, using a weighted selection index for multiple traits, for establishment and evaluation in the RVT.

Here, we consider the generation length (years; L) as the time taken to select individuals to use as parents to produce the next generation of seedlings. Generation length for traditional breeding was 8 (Table 4), as elite individuals are identified after evaluations from age 5 to 8 and are then used as parents for the next generation [1]. By comparison, for strategies employing GS, L was 4. This difference from 8 to 4 years is because elite individuals may be identified from genetic markers at a very early age, but cannot be used as parents until reproductive maturity around the age of 4 [1, 45]. The strategy using GS has a much shorter selection cycle (14 years) than traditional breeding (21 years; Table 1), because it negates the SPT altogether. Both strategies employ RVTs, as it is vital to test the performance of candidate cultivars across multiple environments before commercial release.

For traditional breeding methods, r was calculated as the square-root of yield heritability ($\sqrt{0.30} = 0.55$; Table 5). The genetic standard deviation of PBLUPs (phenotypic best linear unbiased prediction; using unstandardised yield data) was 1237 g. Genetic gain using traditional breeding methods was estimated as 226 g/year for 1% selection intensity. At 2.5% selection

Table 4 Activities involved in a traditional breeding strategy compared with a simple example of how genomic selection (GS) could be employed in a breeding program. The number of years involved in each activity for the two strategies is shown. Information for traditional breeding is adapted from Topp, Hardner [13]. RVT, regional variety trial; SPT, seedling progeny trial

Year	Traditional breeding	Genomic selection
1	Cross parents, grow seedlings	Cross parents, grow seedlings
2	Age 1: Plant SPT	Genotype, select seedlings using GS
3	Age 2: Trial maintenance	Propagate RVT
4	Age 3: Trial maintenance	Age 1: Plant RVT
5	Age 4: Field evaluations	Age 2: Trial maintenance
6	Age 5: Field evaluations	Age 3: Trial maintenance
7	Age 6: Field evaluations	Age 4: Field evaluations
8	Age 7: Field evaluations	Age 5: Field evaluations
9	Age 8: Field evaluations, select seedlings	Age 6: Field evaluations
10	Propagate RVT	Age 7: Field evaluations
11	Age 1: Plant RVT	Age 8: Field evaluations
12	Age 2: Trial maintenance	Age 9: Field evaluations
13	Age 3: Trial maintenance	Age 10: Field evaluations
14	Age 4: Field evaluations	Release
15	Age 5: Field evaluations	
16	Age 6: Field evaluations	
17	Age 7: Field evaluations	
18	Age 8: Field evaluations	
19	Age 9: Field evaluations	
20	Age 10: Field evaluations	
21	Release	

intensity, genetic gain was reduced to 197 g/year. The shorter generation cycle of GS strategies compared with traditional breeding influenced estimates of genetic gain. Genetic gain for GS in related families (randomly-grouped individuals) was more than double that of traditional breeding, at 474 g/year for $s\% = 1$ and 416 g/year for $s\% = 2.5$. However, for unrelated population predictions (individuals grouped by family), traditional breeding achieved higher genetic gain than GS, which was estimated to be 119 g/year for $s\% = 1$ and 105 g/year for $s\% = 2.5$.

Discussion

Comparison of prediction models and cross-validation methods

This study is the first to investigate the use of genomic prediction to improve genetic gain for yield and yield stability in macadamia breeding. Our results suggest that yield-based traits are complex and highly polygenic, as

Table 5 Genetic gain of yield (ΔG , in g/year) for traditional breeding and genomic selection methods as outlined in Table 4. Genetic gain was calculated using Eq. 6, where i was a function of the percentage of the population selected ($s\%$) as given by Falconer and Mackay [46], r is the square-root of yield heritability for traditional breeding or the prediction accuracy of genomic selection model, σ is the standard deviation of PBLUPs (in g), and L is the generation length (in years)

Method	s%	i	r	σ (g)	L (years)	ΔG (g/year)
Traditional breeding	1	2.665	0.55	1237	8	226
	2.5	2.338	0.55	1237	8	197
Genomic selection						
Randomly-grouped (related population)	1	2.665	0.57	1237	4	474
	2.5	2.338	0.57	1237	4	416
Family-grouped (unrelated population)	1	2.665	0.14	1237	4	119
	2.5	2.338	0.14	1237	4	105

indicated by low heritability, and that GS offers a suitable method to select genotypes to improve yield. Prediction accuracy is strongly influenced by the relatedness between training and validation populations [15], and unrelated population predictions are expected to perform poorly compared to related family prediction [47]. This pattern was observed across the models in the current study; model prediction accuracy for randomly-grouped individuals was higher than family-grouped individuals (predictions in unrelated populations). This is because with random groupings for CV, the training set includes full-sibs from the validation set (e.g. progeny from the same cross will be split across the training and validation sets), and so large blocks of chromosomes will be shared between the training and validation sets. The low to moderate prediction accuracies observed by Muranty, Troggio [31] in apple were attributed to predictions across unrelated populations. By comparison, Kumar, Chagne [32] achieved high prediction accuracies (0.70 to 0.90) for apple fruit quality traits, with individuals randomly allocated to CV groups.

The CV method of family-grouped prediction represents an extreme version of the potential real-world application of GS in macadamia where predictions are performed across unrelated populations. It is likely that the training and target populations will actually be more closely related as there is often an overlap of cultivars used as parents between breeding populations, and elite individuals from one population are commonly used as parental germplasm in subsequent generations [13]. It is expected that prediction of GEBVs in a breeding program will, therefore, have accuracies closer to that of the randomly-grouped predictions compared with unrelated population predictions presented in this study. Employing GS in a population closely related to that on which the model is based would provide more accurate predictions of yield. However, more research is needed using large training population sizes with validation sets of whole family groups to improve prediction accuracy before GS can be applied in macadamia breeding.

The implementation of GS in macadamia may include prediction and deployment across environments. The current study population had limited replication of genotypes across environments and did not include $G \times E$ interactions in prediction models as preliminary results found no evidence of $G \times E$ in this experimental material [48]. Previous studies have found some evidence that $G \times E$ may affect macadamia yield [5, 12]. However, research has not yet identified any repeatable factors than can be used for targeted deployment.

Factors affecting accuracy of genomic prediction

The prediction accuracy for yield in the current study was moderate for randomly-grouped individuals ($r = 0.57$), and comparable to the prediction accuracy of yield as measured by phenotypes ($h^2 = 0.30$, $h = r = 0.55$). These similar values for r demonstrate that the genomic prediction accuracy estimated in the current study will provide similar gain as phenotypic analysis, regardless of the time advantage in GS strategies. The prediction accuracy achieved in GS in this study was not as high as reported in some other horticulture crops, which may be attributed to several factors. Estimates of macadamia yield in the current and previous studies [5] involve a large non-genetic component, as indicated by the low heritability and/or high non-additive genetic variation for this trait, and suggest a quantitative nature of inheritance. Yield measurement inaccuracies can occur when overlapping canopies result in a mixture of dropped nuts from neighbouring trees. Additionally, the method used to obtain DNIS (dry nut-in-shell) weight per harvest assumes that the moisture content of the 1 kg sample is consistent through the entire harvest. For these reasons, measuring macadamia yield is very different to measuring yield in other fruit crops, which may inhibit accurate yield prediction.

This study is, to our knowledge, the first to estimate heritability of stability of yield over consecutive years for a nut tree, and use genomic prediction to predict genetic values of yield stability. Biennial bearing in apple has

been researched by multiple authors. Guitton, Kelner [49] found three QTLs associated with biennial bearing that explain 50% of phenotypic variability. Additionally, Durand, Guitton [50] suggested that irregular bearing in apple may be more influenced by factors affecting floral induction rather than those affecting fruit set or drop. Predictions using randomly-grouped individuals were moderately high for yield stability, though this may be due to the low heritability of the trait inflating prediction accuracy. The low heritability of yield stability indicates that yield fluctuations between years is very weakly controlled by genetics, and may be more influenced by non-genetic factors. Thus, it would be up to breeders to determine the value of including yield stability in a selection index when identifying elite candidates for further testing.

The population size of this study was limited compared to other studies in fruit crops, though it did consist of a large number of full-sib families. In the first study of GS in cross-pollinated fruit crop species, Kumar, Chagne [32] obtained high model accuracy for fruit quality traits in apple. They used a much larger population (1120 seedlings) than the current study, albeit from a smaller parent population (seven full-sib families from four female and two male parents), and prediction accuracy ranged from $r = 0.70$ to 0.90 using RR-BLUP and Bayesian LASSO methods. GS in citrus achieved high ($r > 0.7$) prediction accuracy for some fruit quality traits using around 800 individuals, with the GBLUP model consistently out-performing other models [30]. Similarly, using a Japanese pear population of 86 parents and 765 progeny, prediction accuracy varied between models and CV methods, and was commonly greater than 0.5 [33]. However, the correlations found for citrus and Japanese pear may be inflated, since negative correlation coefficients were set to zero when calculating prediction accuracy for these studies. Increasing the size of a phenotyped and genotyped training population would increase the accuracy of yield prediction in macadamia.

LD between markers and genes controlling target traits is essential for GS [15]. Previous studies have suggested increasing the number of markers used in GS may not necessarily achieve better accuracies. Studies investigating the prediction accuracy of GS in citrus, Japanese pear and apple all used fewer SNP markers than the current study (1841, 1502 and 2500, respectively) [30, 32, 33]. Using the same 4113 SNP markers used in the current study, O'Connor, Kilian [44] found that SNPs within 1 kb distance of each other on a scaffold (*M. integrifolia* v2 genome assembly, 4098 scaffolds) had an average LD of $r^2 = 0.124$, with LD decaying rapidly over short distances and more moderately over long distances [44]. These results are important for the current

study to determine that genetic markers capture genetic variance of the target trait [15, 38]. Increasing the density of markers across the genome could lead to increased prediction accuracies, as suggested by Calus, Meuwissen [39], where models with $r^2 = 0.2$ between markers were more accurate than models with fewer markers and lower densities. Future analysis of LD in macadamia could employ the use of an updated macadamia reference genome (45) to determine the distribution of markers across chromosomes, and include corrections for population structure and cryptic relatedness.

Genetic recombination occurs with successive generations of breeding, which may affect the linkage between markers and genes controlling target traits [51]. Furthermore, selection for improved individuals will also alter the frequency of alleles in the population [52]. These changes over generations will have consequences for genomic prediction accuracy. Meuwissen, Hayes [15] estimated that the prediction accuracy of GS models will decrease at around 5% per generation, due to recombination. Thus, it is necessary to recalibrate the model after every few generations as genetic variance explained by the markers will change, along with the allelic frequencies in the population [40, 53]. To aid in model recalibration, Grattapaglia [21] suggested that selection candidates should remain in the field and be grown for 5 to 6 years to provide phenotypes for updating the model. This strategy could be employed in macadamia to ensure accuracy of predictions through subsequent generations of GS.

Genetic gain from genomic selection

The results of this study indicated that genetic gain in macadamia breeding was particularly influenced by the length of the breeding cycle. Genotyping seedlings at a very early age, for example using their first leaf after germination, to identify high-yielding individuals through GS could halve the length of the SPT. Subsequently, elite trees could be cross-pollinated to produce the next generation as soon as they begin to flower, which is usually around the age of four. From there, clonally replicated trees could be phenotyped for other economically important traits, and candidate cultivars identified using a selection index. Similarly in apples, Muranty, Troggio [31] suggested that GS could increase genetic gain per year compared with conventional breeding, by shortening the breeding cycle from 7 to 4 years. In contrast, prediction accuracy was not sufficient for all target traits in oil palm to reduce the generation interval, meaning that breeding would still require the testing of progeny [54]. The authors suggested that if given the resources to increase the size of the training set, and a greater ability to model G x E interactions, GS could be a valid option to increase genetic gain in oil palm.

In their review of GS in apple by Kumar et al. [32, 55], van Nocker and Gardiner [56] proposed using MAS and GS to identify elite apple accessions and then, to decrease time to reproductive maturity, to implement a regime to promote early flowering. Fruit would be phenotyped over two early seasons, and then BVs compared with the predicted GEBVs to analyse genetic gain. Using these methods, candidate cultivars could be clonally propagated 7 years earlier than traditional breeding. However, the predicted beneficial outcomes of using GS in apple may not be as achievable if predictions were to occur across families rather than in randomly-grouped individuals, as has been shown here in macadamia.

Logistics of using genomic selection to increase genetic gain

The opportunity to employ GS in a wider range of crops is increasing with declining genotyping costs and advancements in technology [20, 51, 57]. Implementing genomics-assisted breeding may be expensive due to the cost of genotyping large numbers of candidates at each cycle. However, the cost of genotyping will be a trade-off with a decrease in the costs needed for phenotyping [58] due to the elimination of costs involved in measuring yield during the SPT. An evaluation of costs involved in MAS versus GS has been made for maize and wheat, and GS outperformed MAS even when prediction accuracies were low [58]. Breeders should compare selection strategies to determine which combination of genotyping and phenotyping is most suitable for their crop and program to maximise accuracy of trait prediction in fruit crops [31].

Currently, costs involved in genotyping may restrict the implementation of GS in the Australian macadamia breeding program. To reduce genotyping costs, delaying GS to deploy on a smaller population size may be a viable option, similar to a strategy proposed by Gardiner, Volz [59]; to reduce the size of the seedling population to be genotyped, pre-screen the population for essential traits first. Seedlings could be grown out as per a traditional SPT, but only evaluated to age four, and precocious (early bearing) trees evaluated for KR (high KR attracts a higher commission per kilogram than low KR [60]). Breeders could pre-select precocious seedlings with high KR, genotype this reduced number of potentially elite individuals, and then the highest-yielding trees could be selected through GS for evaluation in RVTs. Longer generation intervals, due to phenotyping for precocity and KR for several years initially, would lead to a lower genetic gain using this strategy than GS of more seedlings at an earlier stage; however, it may be a more cost-effective option. Additionally, whilst implementing GS in macadamia may not decrease the time from seed to reproductive maturity, selecting for precocious

individuals may aid in producing more individuals with a shortened juvenile stage. Reaching reproductive maturity at an earlier stage will further increase genetic gain by reducing the generation length of 4 years in the GS strategy. Extending quantitative modelling of different options for using GS in a breeding program may help to compare possible approaches and identify optimum strategies. Comparing costs of traditional breeding versus strategies using GS is not the focus of this study, though this should be evaluated to determine the prospect of implementing GS in the Australian macadamia breeding program.

Future research using GS in macadamia

Future work employing GS to increase genetic gain in macadamia could investigate other economically important traits, such as tree size. In the same population as the current study, O'Connor et al. [18] found 16 QTLs linked with trunk circumference. The large number of markers associated with this trait, compared with other traits in the study, means that GS may be more appropriate than GWAS and MAS to increase genetic gain, given the seemingly quantitative nature of trunk circumference. GS may also be a good candidate for other traits, such as resistance to diseases and pathogens, including husk spot and phytophthora [61]. Furthermore, the significant associations identified between traits and markers could be incorporated into GS models. Genomic prediction methods including BayesR and BayesB allow the effect of some markers, such as those of significant effect, to be larger than others [15, 62]. Different model types could therefore be tested in the future to determine which are the most accurate in predictions.

Further work could also include multi-trait models to investigate whether the inclusion of additional traits, such as trunk circumference and nut weight, increases the accuracy of yield prediction. Jia and Jannink [63] found that prediction accuracy was increased for a trait with low heritability by including information for a correlated trait with high heritability. Estimates of heritability and genetic correlations between yield and various component traits have been made [48] and, thus, this information could be used to inform multi-trait GS. Distinctions can also be made between linked QTLs (linkage between multiple QTLs affecting different traits) and pleiotropic QTL (one gene affecting multiple traits), using multi-trait methods, like those employed by Bolormaa, Pryce [64].

Finally, future GS analyses could involve more genetic markers and/or more evenly-distributed markers across the genome. This approach may ensure that small-effect loci are captured, since LD in macadamia decays rapidly over short distances [44]. With the aid of the recently published macadamia reference genome (45), future

sequencing of individuals for GS analysis and the calling of SNPs may be more accurate and avoid potential issues associated with allelic dropouts [44].

Conclusions

We found that genomic prediction accuracy of nut yield in macadamia in randomly-grouped individuals was moderate, at $r = 0.57$, and similar to the accuracy of traditional breeding. Prediction accuracy across unrelated populations ($r = 0.14$) was lower than for prediction across randomly-grouped individuals. However, due to the relatedness in parental germplasm between subsequent breeding generations, a realistic prediction accuracy would likely be similar to that of randomly-grouped rather than family-grouped individuals. Additionally, we believe that this study is the first to estimate heritability of yield stability across years for a nut tree, as well as genomic prediction for this trait. Genetic gain for yield using GS (474 g/year) was more than double that of traditional breeding methods (226 g/year), largely due to the generation length being halved in GS. Our results indicate that GS is a viable option to increase genetic gain of yield in macadamia, though validation in a separate population is required. Further research and validation into the use of marker-assisted selection for other key traits in the seedling stage, including kernel recovery and tree size, are needed. Additionally, increasing the number and spread of markers across the genome may capture more causal polymorphisms through LD and lead to higher prediction accuracy.

Methods

Plant material and phenotyping

This study involves a subset of individuals from a progeny population in the Australian industry macadamia breeding program, from which seedlings were produced, established and assessed in field trials by Horticulture Australia Limited, CSIRO and the Queensland Government. The entire progeny population consisted of approximately 2000 seedlings across 141 full-sib families from crosses between 47 publicly available parents, with 1–36 progeny per family (mean 14) [7, 11]. Trees were planted across nine sites in south-eastern Queensland and north-eastern New South Wales, Australia, between 2001 and 2003 in single tree plots in incomplete block design with replication of families [11]. Trees were planted at 4 m distances within rows, and 8 m between rows.

As described by O'Connor, Kilian [44], 295 unrepliated macadamia seedlings from 32 full-sib families (reciprocals combined) from crosses between 29 parents (7–11 progeny per family) across four sites were chosen. All families had at least one parent in common with another family. Progeny within families were selected to

achieve an approximately equal number of low- and high-yielding individuals per family, based on breeding values for cumulative nut-in-shell yield to age 8 years after planting (data not shown), to increase power. Trial sites were within the commercial production zone for macadamia and included Hinkler Park (HP) and Alloway (AL) near Bundaberg, Queensland, and East Gympie (EG) and Amamoor (AM) near Gympie, Queensland. Clones of five parental genotypes were planted at each of the four sites, with a further 13 parental clones planted at AL. Eleven of the 29 parents were not included in the study as they were not present at the study sites. Sites were planted in 2001 (EG, HP), 2002 (AL, part of AM), and 2003 (part of AM).

Historical data were used in this study. Yield was evaluated on an individual tree basis from ages 5 to 8 after planting. Each year, nuts-in-husk were manually harvested from the ground in multiple harvests from February to August. A final strip harvest was undertaken at the end of the season, in which the nuts remaining in the tree were removed with poles and hooks. Nuts were dehusked mechanically and weighed to obtain a wet nut-in-shell weight. For each tree, a 1 kg sample was taken (where available) and dried to approximately 1.5% moisture content at 35 °C for 48 h, 45 °C for 48 h, and 55 °C for 48 h, based on protocol by Prichavudhi and Yamamoto [65]. Samples were then weighed to obtain a dry nut-in-shell (DNIS) weight, with the moisture content used to calculate a total DNIS weight per harvest. DNIS weights were summed across harvests to obtain the total NIS yield per tree each year. For some subsequent analyses, individual tree yield data for each age (age 5 to 8) were standardised by dividing the observation by the standard deviation of each site and age to reduce the bias in genetic values due to heterogeneity of variance among trial sites (following Hardner [7]).

Phenotypic data analysis

An individual-tree linear mixed model was used to predict individual tree effects (phenotypic BLUPs; PBLUPs) using yield data across the four years (ages 5 to 8):

$$\begin{aligned} \text{Yield} = & \text{Site} + \text{Block} + \text{Type} \\ & + \text{Number Neighbours} + \text{Age} + \text{Year} \\ & + \text{Tree} + \text{error} \end{aligned} \quad (1)$$

where **Yield** was the standardised yield of an individual tree in 1 year; **Site** was the fixed effect of the location of the tree (AM, AL, EG and HP in Queensland); **Block** was the fixed effect of planting block within a Site; **Type** was the fixed effect of method used to propagate the tree (seedling progeny or clonally propagated parent); **Number Neighbours** was the fixed effect of the number of trees on either side of that tree within the planting row,

to allow for influence on phenotype of gaps created by the death of neighbouring trees; **Age** was the fixed effect of age after planting of the tree; **Year** was the fixed effect of calendar year that yield was harvested (as trees were planted in different years across the sites); **Tree** was the random effect of individual tree (with no replication of genotypes) within Site, without any pedigree or genetic relationship information, and **error** was a vector of random deviations $e \sim N(0, \sigma_e^2)$ where σ_e^2 is the error variance. As multiple years of yield data were used in the model, a single prediction of mean yield across years for each tree was predicted. PBLUPs were also obtained for trees in each of the four ages (ages 5 to 8) as per Eq. 1, without the effects of Age and Year. Yield stability of each tree was quantified as the standard deviation (SD) of the four annual PBLUPs predicted from the above model.

SNP genotyping and imputation

This study used genetic markers obtained as described by O'Connor, Kilian [44] and briefly outlined here. DNA was extracted from leaves of 295 seedlings and their parents, and sequenced by Diversity Arrays Technology (DARt) Pty Ltd. DARt performed digestion/ligation reactions using a combination of *Pst*I and *Hha*I restriction enzymes and barcoded adaptors. After PCR, samples were pooled, applied to c-Bot (Illumina) bridge PCR, and then sequenced using Illumina HiSeq2500 for 77 cycles. Sequences were processed using proprietary DARt pipelines, with SNP markers detected based on parsing sequence clusters. Missing calls were imputed using the probabilistic principal components analysis (PPCA) method [66] with 97.2% accuracy, which was determined by excluding an additional 10% of missing values and calculating the correlation between the imputed calls and the original dataset. Quality control was performed using pre-imputation parameters, including 50% original call rate, 2.5% minor allele frequency, and a test of Mendelian inconsistencies (parent-offspring trio opposing homozygotes) determined using 16 (50%) of the families. This quality control resulted in 4113 SNPs for genomic analysis.

Genomic BLUP models

An additive genomic relationship matrix (GRM) was constructed among all individuals using the 4113 SNPs, as per VanRaden [67] and detailed in O'Connor, Kilian [44]. GBLUP models were used to estimate GEBVs for each tree using ASReml-R [68]. Preliminary analyses indicated no significant difference in prediction accuracy between additive genomic effects and sites [48], and thus G x E was not included in analyses.

$$\begin{aligned} \text{Yield} = & \text{Site} + \text{Block} + \text{Type} \\ & + \text{Number Neighbours} + \text{Age} + \text{Year} \\ & + \text{Accession} + \text{SiteTree} + \text{error} \end{aligned} \quad (2)$$

where **Accession** is the tree effect modelled as the additive genetic effect of the individual, assumed random $\sim N(0, \mathbf{G}\sigma_g^2)$, where **G** is the GRM, modelled from SNP effects (where 0, 1, and 2 represents the dosage of the reference allele), and σ_g^2 is the additive genetic variance captured by the SNP markers; and **SiteTree** is the permanent environment random effect of each tree within site across years, uncorrelated among trees, assumed random $\sim N(0, \sigma_{pe}^2)$ where σ_{pe}^2 is the variance attributed to a permanent environment effect. As multiple years of yield data were used in the model, a single prediction for each accession for mean yield GEBV across years was obtained.

To determine the accuracy of genomic prediction for yield stability over multiple years, GEBVs were obtained using the following model:

$$\text{Yield SD} = \text{mean} + \text{Accession} + \text{error} \quad (3)$$

where **Yield SD** is the standard deviation of PBLUPs for ages 5 to 8 across all sites.

An estimate of genomic narrow-sense genomic heritability (h^2) of average yield was made based on the variance components of Eq. 2 (Supplementary Table 2) using the pin function in R [69]:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{pe}^2 + \sigma_e^2 / 4} \quad (4)$$

where σ_g^2 is the additive genetic variance, σ_{pe}^2 is the variance attributed to a permanent environment effect, and σ_e^2 is the residual variance (divided by 4 for the 4 years of data). For yield stability, an estimate of h^2 was calculated as $\sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$ from variance components of Eq. 3.

Model validation

The prediction accuracies of the GEBVs from the models above were determined using five-fold cross-validation (CV). In turn, 20% of phenotypes were masked (set to missing) in a validation set, and data for the remaining 80% of individuals were used as a training set to train the model and predict the masked values. This process was repeated five times until all subsets were used in the validation set, with each individual used only once in the validation set.

Individuals were assigned to one of five groups for the five-fold CV using two grouping techniques for predictions: random and related family groups. For the random CV, individuals were selected for the training and validation group at random ("randomly-grouped"). Here, full-

sibs were randomly allocated across the training and validation groups, and so predictions were performed on individuals that were related to the training population. For the second method, individuals were grouped by family and related families (those with common parents) and grafted parents (“family-grouped”), to give approximately equal-sized groups. Thus, entire full-sib families were either in the reference set or in the validation set, and predictions were performed on families unrelated to the training population (“unrelated population”). This second method represents a more extreme version of the application of GS, where the two populations are not closely related.

Prediction accuracy (r) for each of the five CVs was calculated as:

$$r = \frac{\text{corr}(\text{GEBVs}, \text{PBLUPs})}{\sqrt{h^2}} \quad (5)$$

where the PBLUPs were estimated from Eq. 1, and heritability (h^2) was estimated from Eq. 4. The correlation between GEBVs and PBLUPs is predictive ability. Mean prediction accuracies and standard errors were calculated across the five CVs, and t-tests were performed to determine if prediction accuracies were significantly different from zero.

Comparison of breeding strategies and genetic gain

A simple comparison of breeding strategies was made to demonstrate how GS could be effectively incorporated into the macadamia breeding program to reduce selection time and increase genetic gain. Genetic gain (ΔG , grams/year) was calculated for yield for traditional breeding and GS methods using the following equation derived from Falconer [52]:

$$\Delta G = \frac{i \times r \times \sigma}{L} \quad (6)$$

where i is selection intensity as a function of the proportion of the population selected, r is square-root of yield heritability for traditional breeding or the prediction accuracy of the GS model, and L is generation length in years. Here, σ is the genetic standard deviation, which was calculated as the standard deviation of PBLUPs from Eq. 1, but using unstandardised yield data to give a value in grams. In traditional breeding, approximately 2000 seedlings are evaluated and 1% (20/2000) of the SPT population are further tested in an RVT [11]. Here, the selected percentage of the population (%) has been increased from 1 to 2.5% in an attempt to reduce the probability of not selecting truly elite germplasm under GS. As such, in this equation, $i = 2.665$ and 2.338 , for 1 and 2.5% selected, respectively, as taken from Falconer and Mackay [46]. We assume that genetic gain for RVT selection is the same across selection strategies, and so genetic gain is only calculated here for the SPT.

Abbreviations

AL: Alloway (study location site); AM: Amamoor (study location site); CV: Cross-validation; DArT: Diversity Arrays Technology; DNIS: Dry nut-in-shell; EG: East Gympie (study location site); GBLUP: Genomic best linear unbiased prediction; GEBV: Genomic estimated breeding value; GRM: Genomic relationship matrix; GS: Genomic selection; GWAS: Genome-wide association study; G x E: Genotype by environment interaction; HP: Hinkler Park (study location site); KR: Kernel recovery (ratio of kernel to nut-in-shell); LD: Linkage disequilibrium; MAS: Marker-assisted selection; NIS: Nut-in-shell; PBLUP: Phenotypic best linear unbiased prediction; QTL: Quantitative trait loci; RVT: Regional variety trial; SD: Standard deviation; SE: Standard error; SNP: Single nucleotide polymorphism; SPT: Seedling progeny trial

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07694-z>.

Additional file 1: Supplementary Table 1. Variance components for phenotypic BLUPs derived from Eq. 1. **Supplementary Table 2.** Variance components for genomic BLUPs derived from Eq. 2.

Supplementary Figure 1. Boxplots showing the observed relationship distributions from the GRM for: full siblings, half siblings, parent offspring, parent-parent, and unrelated relationship groups. **Supplementary Figure 2.** Boxplot showing the observed relationship distributions from the GRM for the diagonals (identity).

Acknowledgements

The authors acknowledge the teams at CSIRO and the Queensland Government who established the research trials, and we thank the many people who contributed to the collection and collation of historical yield data used in the study. Thanks also to the anonymous reviewers for their comments and feedback.

Authors' contributions

KO wrote the paper, performed all analyses and made final edits. KO, BH and CH developed the analytical models. BH, CH, MA, BT and RH assisted in interpretation of results and suggested revisions. All authors read and approved the final manuscript.

Funding

This research has been funded by Hort Innovation Australia, using the Macadamia research and development levy and contributions from the Australian Government. Hort Innovation is the grower owned, not-for-profit research and development corporation for Australian horticulture. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. KO acknowledges the Australian Postgraduate Award and Charles Morphett Peglar scholarship for financial support.

Availability of data and materials

The datasets generated and/or analysed during the current study are available from The University of Queensland's Institutional Data Access/Ethics Committee, but restrictions apply to the availability of these data. Data are however available from the authors upon reasonable request and with permission of The University of Queensland for researchers who meet the criteria for access to confidential data. SNP data can be accessed from <https://doi.org/10.14264/uql.2018.491> and yield data from <https://doi.org/10.14264/5f8abc3>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Queensland Department of Agriculture and Fisheries, Maroochy Research Facility, 47 Mayers Road, Nambour, QLD 4560, Australia. ²Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Maroochy Research Facility, 47 Mayers Road, Nambour, QLD 4560, Australia. ³Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, QLD 4072, Australia.

Received: 23 December 2020 Accepted: 10 May 2021

Published online: 20 May 2021

References

- Hardner CM, Peace C, Lowe AJ, Neal J, Pisanu P, Powell M, et al. Genetic resources and domestication of macadamia. *Hortic Rev.* 2009;35:1–126.
- Australian Macadamia Society. 2017 Australian macadamia crop reaches 46,000 tonnes in-shell. 2017. [Available from: <http://australian-macadamias.org/industry/site/industry/industry-page/industry-news-archive/latest-news-industry/2017-australian-macadamia-crop-reaches-46000-tonnes-in-shell?itemid=133&lang=en>].
- Jannink J-L, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. *Brief Functional Genomics.* 2010;9(2):166–77. <https://doi.org/10.1093/bfpg/elq001>.
- Quarrie S, Rancic D, Radosevic R, Pekic Quarrie S, Kaminska A, Barnes J, et al. Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. *J Exp Bot.* 2006;57(11):2627–37. <https://doi.org/10.1093/jxb/erl026>.
- Hardner CM, Winks CW, Stephenson RA, Gallagher EG, McConchie CA. Genetic parameters for yield in macadamia. *Euphytica.* 2002;125(2):255–64. <https://doi.org/10.1023/A:1015857409317>.
- Allard RW, Bradshaw AD. Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 1964;4(5):503–8. <https://doi.org/10.2135/cropsci1964.0011183X000400050021x>.
- Hardner C. Exploring opportunities for reducing complexity of genotype-by-environment interaction models. *Euphytica.* 2017;213(11):248–64. <https://doi.org/10.1007/s10681-017-2023-0>.
- Kelner J-J, Costes E, Guitton B, Chagné D, Gardiner SE, Velasco R. Genetic control of biennial bearing in apple. *J Exp Bot.* 2011;63(1):131–49.
- Cilas C, Montagnon C, Bar-Hen A. Yield stability in clones of *Coffea canephora* in the short and medium term: longitudinal data analyses and measures of stability over time. *Tree Genet Genomes.* 2011;7(2):421–9. <https://doi.org/10.1007/s11295-010-0344-4>.
- Sharma N, Singh SK, Mahato AK, Ravishanker H, Dubey AK, Singh NK. Physiological and molecular basis of alternate bearing in perennial fruit crops. *Sci Hortic.* 2019;243:214–25. <https://doi.org/10.1016/j.scienta.2018.08.021>.
- Topp B, Hardner CM, Neal J, Kelly A, Russell D, McConchie C, et al. Overview of the Australian macadamia industry breeding program. *Acta Hortic.* 2016; 1127:45–50.
- Hardner C, Costae Silva J, Williams E, Meyers N, McConchie C. Breeding new cultivars for the Australian macadamia industry. *HortScience.* 2019;54(4):621–8. <https://doi.org/10.21273/HORTSCI13286-18>.
- Topp B, Hardner C, Kelly A. Strategies for breeding macadamias in Australia. *Acta Hortic.* 2012;935:47–53.
- Viana AP, Resende MDV, Riaz S, Walker MA. Genome selection in fruit breeding: application to table grapes. *Sci Agric.* 2016;73(2):142–9.
- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157(4):1819–29. <https://doi.org/10.1093/genetics/157.4.1819>.
- O'Connor K, Hayes B, Topp B. Prospects for increasing yield in macadamia using component traits and genomics. *Tree Genet Genomes.* 2018;14(1): Article 7.
- O'Connor K, Hayes B, Hardner C, Alam M, Topp B. Selecting for nut characteristics in macadamia using a genome-wide association study. *HortScience.* 2019;54(4):629–32. <https://doi.org/10.21273/HORTSCI13297-18>.
- O'Connor K, Hayes B, Hardner C, Nock C, Baten A, Alam M, et al. Genome-wide association studies for yield component traits in a macadamia breeding population. *BMC Genomics.* 2020;21(1):1–12.
- Luby JJ, Shaw DV. Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortScience.* 2000;36(5):872–9.
- Iwata H, Minamikawa MF, Kajiya-Kanegae H, Ishimori M, Hayashi T. Genomics-assisted breeding in fruit trees. *Breed Sci.* 2016;66(1):100–15. <https://doi.org/10.1270/jsbbs.66.100>.
- Grattapaglia D. Breeding forest trees by genomic selection: current progress and the way forward. In: Tuberosa R, Graner A, Frison E, editors. *Genomics of plant genetic resources*. Volume 1. Managing, sequencing and mining genetic resources. London: Springer; 2014. p. 651–82.
- Lin Z, Hayes BJ, Daetwyler HD. Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* 2014;65(11):177–91. <https://doi.org/10.1071/CP13363>.
- Destá ZA, Ortiz R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 2014;19(9):592–601. <https://doi.org/10.1016/j.tplants.2014.05.006>.
- Denis M, Bouvet J-M. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. *Tree Genet Genomes.* 2013;9(1):37–51. <https://doi.org/10.1007/s11295-012-0528-1>.
- Piaskowski J, Hardner C, Cai L, Zhao Y, Iezzoni A, Peace C. Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. *BMC Genet.* 2018;19(1):23–38. <https://doi.org/10.1186/s12863-018-0609-8>.
- Biscarini F, Nazzicari N, Bink M, Arús P, Aranzana MJ, Verde I, et al. Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. *BMC Genomics.* 2017;18(1):432–46. <https://doi.org/10.1186/s12864-017-3781-8>.
- Kwong QB, Teh CK, Ong AL, Chew FT, Mayes S, Kulaveerasingam H, et al. Evaluation of methods and marker systems in genomic selection of oil palm (*Elaeis guineensis* Jacq.). *BMC Genet.* 2017;18(1):107–15. <https://doi.org/10.1186/s12863-017-0576-5>.
- Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, Mayes S, et al. Genomic selection in commercial perennial crops: applicability and improvement in oil palm (*Elaeis guineensis* Jacq.). *Sci Rep.* 2017;7(1):2872–81. <https://doi.org/10.1038/s41598-017-02602-6>.
- Wong C, Bernardo R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor Appl Genet.* 2008;116(6):815–24. <https://doi.org/10.1007/s00122-008-0715-5>.
- Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, et al. Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. *Sci Rep.* 2017; 7(1):4721–34. <https://doi.org/10.1038/s41598-017-05100-x>.
- Muranty H, Troggio M, Sadok IB, Al Rifai M, Auwerkerken A, Banchi E, et al. Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Res.* 2015;2:15060.
- Kumar S, Chagne D, Bink MC, Volz RK, Whitworth C, Carlisle C. Genomic selection for fruit quality traits in apple (*Malus x domestica* Borkh.). *PLoS One.* 2012;7(5):e36674.
- Minamikawa MF, Takada N, Terakami S, Saito T, Onogi A, Kajiya-Kanegae H, et al. Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (*Pyrus pyrifolia* Nakai). *Sci Rep.* 2018;8(1):11994. <https://doi.org/10.1038/s41598-018-30154-w>.
- Iwata H, Hayashi T, Terakami S, Takada N, Sawamura Y, Yamamoto T. Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breed Sci.* 2013;63(1):125–40. <https://doi.org/10.1270/jsbbs.63.125>.
- Nishio S, Hayashi T, Yamamoto T, Terakami S, Iwata H, Imai A, et al. Bayesian genome-wide association study of nut traits in Japanese chestnut. *Mol Breed.* 2018;38(8):99–114. <https://doi.org/10.1007/s11032-018-0857-3>.
- Crossa J, de los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics.* 2010;186(2):713–24. <https://doi.org/10.1534/genetics.110.118521>.
- Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity.* 2014;112(1):39–47. <https://doi.org/10.1038/hdy.2013.13>.
- Goddard M. Mapping genes for quantitative traits using linkage disequilibrium. *Genet Sel Evol.* 1991;23(Suppl 1):131s–4s.
- Calus M, Meuwissen T, De Roos A, Veerkamp R. Accuracy of genomic selection using different methods to define haplotypes. *Genetics.* 2008; 178(1):553–61. <https://doi.org/10.1534/genetics.107.080838>.
- Resende MF Jr, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, et al. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol.* 2012; 193(3):617–24. <https://doi.org/10.1111/j.1469-8137.2011.03895.x>.

41. Xu Y, Crouch JH. Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 2008;48(2):391–407. <https://doi.org/10.2135/cropsci2007.04.0191>.
42. Rikkerink EH, Oraguzie NC, Gardiner SE. Prospects of association mapping in perennial horticultural crops. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN, editors. *Association mapping in plants*. New York: Springer; 2007. p. 249–69. https://doi.org/10.1007/978-0-387-36011-9_11.
43. Nock CJ, Baten A, Mauleon R, Langdon KS, Topp B, Hardner CM, et al. Chromosome-scale assembly and annotation of the macadamia genome (*Macadamia integrifolia* HAES 741). *G3: Genes, Genomes, Genetics.* 2020; 10(10):3497–504.
44. O'Connor K, Killian A, Hayes B, Hardner C, Nock C, Baten A, et al. Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silicoDART markers. *Tree Genet Genomes.* 2019;15(2):Article 24.
45. Topp B, Nock C, Hardner C, Alam M, O'Connor K. Macadamia (*Macadamia* spp.) breeding. In: Al-Khayri JM, Jain SM, Johnson DV, editors. *Advances in plant breeding strategies: nut and beverage crops*. 4. Switzerland: Springer International Publishing; 2019.
46. Falconer DS, Mackay TF. *Introduction to quantitative genetics*. 4th ed. Sussex: UK Longman Group; 1996.
47. Pszczola M, Strabel T, Mulder H, Calus M. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci.* 2012;95(1):389–400. <https://doi.org/10.3168/jds.2011-4338>.
48. O'Connor K. *Selection strategies to improve yield in macadamia using component traits and genomics*. Brisbane: University of Queensland; 2019.
49. Guitton B, Kelner J-J, Velasco R, Gardiner SE, Chagne D, Costes E. Genetic control of biennial bearing in apple. *J Exp Bot.* 2012;63(1):131–49. <https://doi.org/10.1093/jxb/err261>.
50. Durand J-B, Guitton B, Peyhardi J, Holtz Y, Guédon Y, Trottier C, et al. New insights for estimating the genetic value of segregating apple progenies for irregular bearing during the first years of tree production. *J Exp Bot.* 2013; 64(16):5099–113. <https://doi.org/10.1093/jxb/ert297>.
51. Khan MA, Korban SS. Association mapping in forest trees and fruit crops. *J Exp Bot.* 2012;63(11):4045–60. <https://doi.org/10.1093/jxb/ers105>.
52. Falconer DS. *Introduction to quantitative genetics*. Essex: Longman Scientific & Technical; 1989.
53. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica.* 2009;136(2):245–57. <https://doi.org/10.1007/s10709-008-9308-0>.
54. Cros D, Bocs S, Riou V, Ortega-Abboud E, Tisné S, Argout X, et al. Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. *BMC Genomics.* 2017;18(1):839–55. <https://doi.org/10.1186/s12864-017-4179-3>.
55. Kumar S, Bink MC, Volz RK, Bus VG, Chagné D. Towards genomic selection in apple (*Malus × domestica* Borkh.) breeding programmes: prospects, challenges and strategies. *Tree Genet Genomes.* 2012;8(1):1–14. <https://doi.org/10.1007/s11295-011-0425-z>.
56. van Nocker S, Gardiner SE. Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. *Horticulture Res.* 2014;1:Article 14022.
57. Heffner EL, Sorrells ME, Jannink J-L. Genomic selection for crop improvement. *Crop Sci.* 2009;49(1):1–12. <https://doi.org/10.2135/cropsci2008.08.0512>.
58. Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME. Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 2010;50(5):1681–90. <https://doi.org/10.2135/cropsci2009.11.0662>.
59. Gardiner S, Volz R, Chagné D. Tools to breed better cultivars faster at Plant & Food Research. In: *Proceedings of the 1st international rapid cycle crop breeding conference*; 2014.
60. Macadamia Processing Co. Ltd. 2018 Notional Price Table for NIS at 10% Moisture Content. Access date 5/12/2018. 2018.
61. Drenth A, Akinsanmi OA, Miles A. Macadamia diseases in Australia. *Southern African Macadamia Growers' Assoc Yearbook.* 2009;17:48–52.
62. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci.* 2012;95(7):4114–29. <https://doi.org/10.3168/jds.2011-5019>.
63. Jia Y, Jannink J-L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics.* 2012;192(4):1513–22. <https://doi.org/10.1534/genetics.112.144246>.
64. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.* 2014;10(3): e1004198. <https://doi.org/10.1371/journal.pgen.1004198>.
65. Prichavudhi K, Yamamoto HY. Effect of drying temperature on chemical composition and quality of macadamia nuts. *Food Technol.* 1965;19(7): 1153–6.
66. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. *pcaMethods*—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics.* 2007;23(9):1164–7. <https://doi.org/10.1093/bioinformatics/btm069>.
67. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91(11):4414–23. <https://doi.org/10.3168/jds.2007-0980>.
68. Butler D, Cullis B, Gilmour A, Gogel B. *Asreml: asreml () fits the linear mixed model*. R package, version 3. 2009.
69. White I. Pin function for asreml-R. <http://www.homepages.ed.ac.uk/iwhite/aa/sreml/>. 2013.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

