



Prospects for increasing yield in macadamia using component traits and genomics

Katie O'Connor¹ · Ben Hayes¹ · Bruce Topp¹

Received: 10 October 2017 / Revised: 18 December 2017 / Accepted: 20 December 2017 / Published online: 2 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Selection of candidate cultivars in macadamia requires extensive phenotypic measurements over many years and trials. In particular, yield traits such as nut-in-shell yield and kernel yield are economically vital characteristics and therefore guide the selection process for new cultivars. However, these traits can only be measured in mature trees, resulting in long generation intervals and slow rates of genetic gain. In addition, these traits are expensive to measure. Strategies to reduce the generation interval and increase the intensity of selection include using yield component traits, identification of markers associated with component traits, and genomic selection for yield. Yield component traits that contribute to resource availability for fruit formation include floral and nut characteristics. In this review, these traits will be investigated to estimate their relative importance in macadamia breeding and their heritability and correlations with yield. Furthermore, the usefulness of genome-wide association studies regarding yield component traits will be reviewed. Genetic-based breeding techniques could exploit this information to increase yield gains per breeding cycle and estimate the quantitative nature of yield traits. Genomic selection uses genome-wide molecular markers to predict the phenotype of individuals at an early age before maturity, thereby reducing the cycle time and increasing gain per unit time in plant breeding programmes. This review evaluates the potential for measurement of yield component traits, genome-wide association studies, and genomic selection to be employed in the Australian macadamia breeding programme to accelerate gains for nut yield.

Keywords Nut crop · Breeding · Perennial · Genome-wide association study (GWAS) · Genomic selection

Introduction

In the past few decades, substantial increases in yield have resulted from genetic improvement in many crops including maize (Duvick 1984) and apple (Igarashi et al. 2016). However, genetic improvement is still in its infancy in many tree species due to their long generation times and the cost of screening new cultivars (Khan and Korban 2012; Kumar et al. 2013a; Isik 2014; van Nocker and Gardiner 2014; Cros et al. 2015). High yield is often the focus in crop breeding programmes, yet selection gains can be hindered since yield is commonly difficult to select due to its complex nature. The

process of yield genetic gain in fruit tree crops can be accelerated in a number of ways.

One method of improving yield is by mining for yield component traits. Component traits that are correlated with yield, and are more heritable and easier to measure, may be used to indirectly select for high yield (Fraser and Eaton 1983; Sparnaaij and Bos 1993; Piepho 1995). This indirect selection may increase breeding gains by reducing cycle times if the component traits are measured earlier in the process than yield.

Other methods to increase yield in crop breeding programmes include employing DNA-based technologies. This includes combining genome-wide association studies (GWAS) with marker-assisted selection (MAS), and using genomic selection (GS) (Lande and Thompson 1990; Varshney et al. 2005; Endresen 2010; Khan and Korban 2012; van Nocker and Gardiner 2014; Isik et al. 2015). GWAS can help identify genetic markers associated with key yield component traits, which can then be screened for in a population and elite candidates selected using MAS. GS

Communicated by S. C. González-Martínez

✉ Katie O'Connor
k.oconnor@uq.edu.au

¹ Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, QLD 4072, Australia

can be used to select for the more complex trait yield by modelling genetic markers across the genome and their effect on the trait to predict the yield of each candidate.

Luby and Shaw (2001) proposed that fruit crops have more to gain from MAS than annual crops due to their large tree size and long generation times, and the time and cost involved in maintaining the trees. However, they recognised that this may be true only if the trait in question is simply inherited, is economically important, and is conventionally very expensive to measure (Luby and Shaw 2001). Since that time, the technology of molecular markers has dramatically expanded and advanced. Genomics-based methods for improving the efficiency of breeding programmes such as GWAS and GS are now particularly pertinent for fruit trees (Wong and Bernardo 2008; Kumar et al. 2012b; Iwata et al. 2016; Yamamoto and Terakami 2016; Peace 2017). These methods have advanced from the more fundamental marker-assisted breeding and trait mapping, have higher accuracies and wider applications (Iwata et al. 2016), and have potential use in breeding for increased yield in a crop such as macadamia.

This review investigates genomic improvement in crop breeding, with specific reference to fruit and nut tree crops including macadamia. The potential use of yield component traits, GWAS, and GS in improving yield in macadamia will be explored.

Macadamia: a native Australian nut

Macadamia (Proteaceae) is a subtropical rainforest tree, native to the east coast of Australia between Mount Bauple, Queensland, and Lismore, New South Wales (Gross 1995; Hardner et al. 2009). The genus contains four species: *M. integrifolia*, *M. tetraphylla*, *M. ternifolia*, and *M. janseni* (Peace et al. 2008; Hardner et al. 2009). Individual trees are produced predominantly from outcrossing, similar to many other rainforest species, are large in size, and have a long juvenile period (Fig. 1; Sedgley et al. 1990; Trueman and Turnbull 1994). Both *M. integrifolia* and *M. tetraphylla* and their hybrids are cultivated around the world for their edible nuts (Fig. 2; Hardner et al. 2009).

Macadamias are diploid ($2n = 28$), highly heterozygous, with genome size estimates ranging from 652 Mb (Nock et al. 2016) to 780 Mb (Chagné 2015). A draft genome assembly of short-read Illumina sequences from cultivar 'HAES 741' covers 79% of the total estimated genome, at 518 Mb in length (Nock et al. 2016). Nock et al. (2016) discussed ongoing work to improve genome coverage by incorporating deeper, long-read PacBio sequence data and develop a high-density linkage map, which will be advantageous for future genomics studies. The Australian National Macadamia Germplasm Collection contains accessions across all four

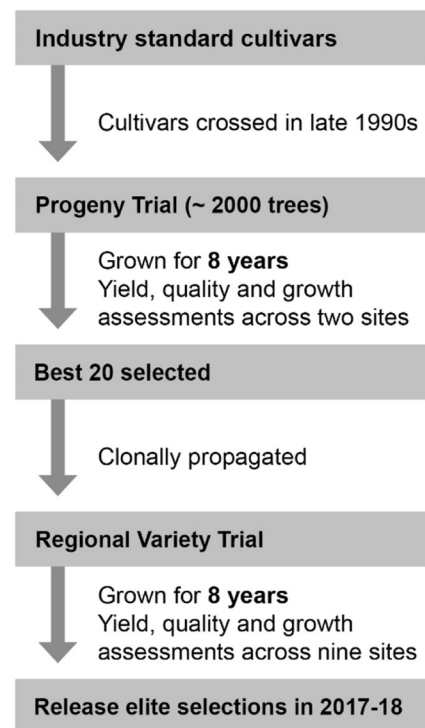


Fig. 1 Timeline of the Australian macadamia breeding programme's first generation, showing evaluation steps indicative of traditional breeding practices

Macadamia species, which is also available as a genomics resource for sequencing (Hardner et al. 2004).

Domestication and cultivation

Domestication of macadamia is only relatively recent, with cultivation beginning in the late 1800s (Peace et al. 2008; Hardner et al. 2009; Hardner 2015). Two early importations of *M. integrifolia* nuts from Australia to Hawaii occurred in the 1880s and 1890s (Hamilton and Fukunaga 1959). It has been suggested that the first exports originated from near Mount Bauple (Hardner 2015). Planting of seedlings began around 1920 by the Hawaii Agricultural Experiment Station, whilst evaluation and selection of new cultivars commenced in the mid-1930s (Hamilton and Fukunaga 1959; Hardner 2015). This programme provided the majority of commercial cultivars currently grown around the world (Hardner 2015). In Australia, the first orchards were established near Lismore, NSW, in the 1880s and in Queensland in 1910 (Hardner et al. 2009). As such, cultivated varieties are only a few generations removed from their wild relatives. Macadamias are mainly produced in Australia, South Africa, the USA (Hawaii), and Kenya (Australian Macadamia Society 2012).

Trees begin to bear nuts after 4 to 5 years, are fully mature after 10 to 15 years, and can be commercially productive for up to 60 years (Hardner et al. 2009). In 1997, the Commonwealth Scientific and Industrial Research



Fig. 2 Macadamia nuts—the edible kernel is enclosed in a hard, woody shell and the outer husk. Left to right: nut in husk, split husk, nut in shell, cracked shell, and kernel. Illustration by Todd Fox

Organisation (CSIRO) launched an Australian macadamia breeding programme from which subsequent selections have been made for parental crossing and regional variety trials (RVTs) to evaluate elite selections (Fig. 1; Hardner et al. 2002). Large areas are required over various environments to evaluate new cultivars. RVTs in macadamias are usually maintained for 8 years from planting date, with many traits measured each year (Hardner et al. 2002). Several years of data are required in order to select high-yielding cultivars (Hardner et al. 2001). Yield and growth data up to year 8 are available for ~2000 trees in Australia from these trials.

High yield is the primary trait used to select new cultivars (Stephenson et al. 1986; Hardner et al. 2009; Howlett et al. 2015). Other important traits selected in the Australian breeding programme are high kernel recovery, small tree size, and high proportion of intact kernels (Topp et al. 2012). Nut-in-shell (NIS) yield refers to the weight of de-husked nuts at 1% moisture content with the shell intact (Hardner et al. 2002). Kernel recovery (KR) is the ratio of kernel to nut mass (Kester and Asay 1975; Hardner et al. 2002); high KR is desired as this indicates that the kernel is relatively large compared with the weight of the shell. However, cultivars with high KR have thin shells, which are susceptible to pests and diseases (Hardner et al. 2009). Depending on the use of the product, whole unbroken kernels may be desirable, so this is also an important trait (O'Hare et al. 2004; Hardner et al. 2009). Industry standards for these traits are as follows: 5 t per ha NIS of > 18 mm diameter, > 36% KR, 2–3 g kernels, and > 50% whole kernels (O'Hare et al. 2004); however, production can fall short of these standards.

Gain from selection

Gain from selection efforts in breeding can be predicted using the following formula:

$$R = \frac{h^2 S}{Y}$$

where R is the response per year, or genetic gain; h^2 is the narrow-sense heritability, and also a function of trait measurement accuracy; S is the selection differential, the amount by which the average parents' performance exceeds the average

breeding population performance; and Y is the selection cycle length in years (Hansche 1983). In macadamia, genetic gain is impeded by a long breeding cycle with the current breeding programme requiring 8 years to evaluate NIS yield (Fig. 1). Reducing the selection cycle is thus an important aim for macadamia breeding. Selection intensity and accuracy are costly to improve due to the large plant size which adds to the cost of increased population size and replication. The following sections address these factors with reference to improvements in macadamia.

Yield and its component traits

Yield is the highest priority in macadamia breeding; yield was consistently ranked by the industry as the most important future cultivar characteristic (O'Hare and . 2010) and was economically weighted the highest of all traits in the selection index used in Australian macadamia breeding (Hardner et al. 2006). Macadamia yield can be quantified as wet nut-in-husk, wet nut-in-shell, nut-in-shell dried to 1% kernel moisture, and expressed on a per tree size or per hectare basis (Hardner et al. 2009). However, selecting for yield in the breeding programme is difficult as it is a complex trait (variation is the result of small effects at many loci). Hardner et al. (2002) found that broad-sense heritability for annual NIS yield ranged from 0.06 to 0.18, whilst cumulative NIS ranged from 0.11 to 0.20 (Table 1).

Yield or other complex traits may be indirectly selected through correlated component traits that are more heritable (Fraser and Eaton 1983; Sparnaaij and Bos 1993; Piepho 1995). It is best to initially explore simple component traits related to yield that may be easier and/or cheaper to measure (Sparnaaij and Bos 1993). The investigation of component traits can reduce cycle times if they can be measured earlier in the tree's life, and selection intensity can be increased if the traits are efficiently measured, allowing evaluation of a larger number of plants. This is particularly true when the trees are young, as less land and fewer resources are required. However, Fraser and Eaton (1983) noted that in broad acre and horticultural crops, it may be ineffective to rely on component traits correlated with the complex target trait as many components are often linked. Other sequential and path

Table 1 Heritability and correlations between various flower and fruit characteristics in macadamia and other nut crops. r_g genetic correlation, r_p phenotypic correlation, H^2 broad-sense heritability, h^2 narrow-sense heritability

Crop	Trait/s	Heritability	Correlation between traits	Source
Macadamia	Annual nut-in-shell yield (10 yrs)	$H^2 = 0.14$		Hardner et al. 2002
<i>Macadamia integrifolia</i> and <i>M. tetraphylla</i>	Cumulative nut-in-shell yield (10 yrs)	$H^2 = 0.17$		
	Kernel recovery and cumulative kernel yield (10 yrs)		$r_g = 0.08, r_p = 0.12$	
	Stem girth and cumulative nut-in-shell yield (10 yrs)		$r_g = 0.22, r_p = 0.59$	
	Nut weight	$H^2 = 0.63$		Hardner et al. 2001
	Kernel weight	$H^2 = 0.63$		
	Kernel recovery	$H^2 = 0.63$		
	Nut weight and kernel weight		$r_g = 0.79, r_p = 0.68$	
	Kernel recovery and kernel mass		$r_g = 0.48, r_p = 0.49$	
	Kernel recovery and shell thickness		$r_p = -0.70$	Leverington 1962
Pecan	Nut weight	$h^2 = 0.35$		Thompson and Baker 1993
<i>Carya illinoensis</i>	Kernel weight	$h^2 = 0.38$		
	Kernel recovery and kernel weight		$r_p = 0.394$	
	Nut yield (kg/tree)	$H^2 = 0.855$		Kumar et al. 2013a
	Kernel recovery	$H^2 = 0.897$		
	Kernel recovery and kernel weight		$r = 0.569$	
Hazelnut	Kernel weight	$h^2 = 0.67$		Yao and Mehlenbacher 2000
<i>Carylus avellana</i>	Kernel recovery	$h^2 = 0.87$		
	Relative husk length	$h^2 = 0.91$		
Cashew	Tree nut yield and whole nut weight		$r_p = 0.108$	Aliyu 2006
<i>Anacardium occidentale</i>	Tree nut yield and number nuts per panicle		$r_p = 0.844$	
	Tree nut yield and number hermaphrodite flowers per panicle		$r_p = 0.863$	
Walnut	Crop yield	$h^2 = 0.07$		Hansche et al. 1972
<i>Juglans regia</i>	Nut weight	$h^2 = 0.86$		
	Crop yield and nut weight		$r_p = -0.20$	

analyses have been proposed to overcome this difficulty (e.g. Li 1975; Thomas and Grafius 1976; Eaton and Kyte 1978; Sparnaaij and Bos 1993; Piepho 1995).

It is important to recognise the relationship between different traits and how they affect yield (Samonte et al. 1998). Understanding genetic parameters such as heritability and correlations between various traits can help select parents in breeding programmes (Falconer 1989; Bodzon 2004). In *Prunus*, traits affecting fruit quality such as flavour, colour, and shape are often related (de Souza et al. 1998; Cantín et al. 2010). Correlations of component traits and yield in macadamia and other nut crops are presented in Table 1.

There are many components of yield in macadamia, some of which have been evaluated and others that need further exploration (Hardner et al. 2009). This review focuses on those factors that affect flower and nut development, and hence resource utilisation for the nuts. Further research is needed to understand the different components of yield in

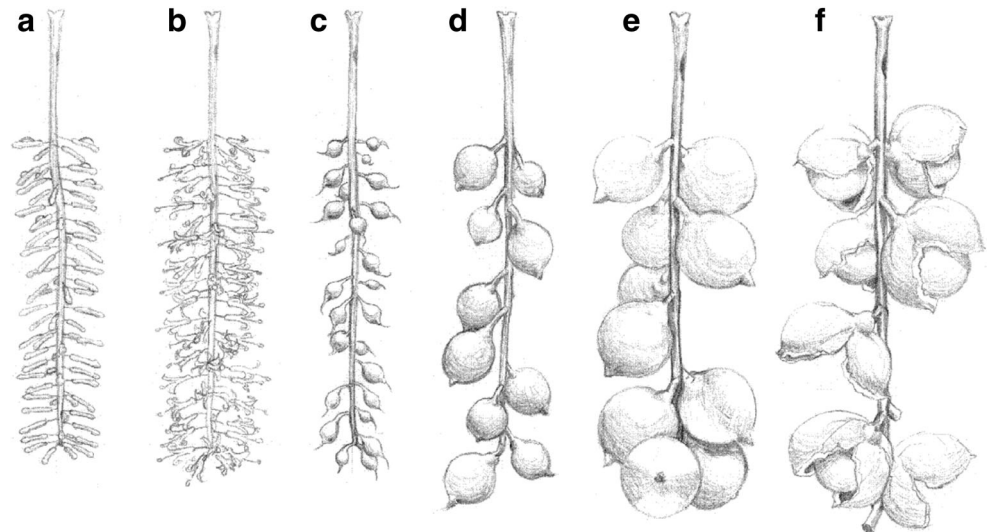
macadamia and to identify the important related traits that can easily be measured.

Flowering and growth traits

Flowering plays a critical role in fruit production, and so it is necessary to understand the factors affecting flower development (Westwood 1993). A review of flowering and fruiting in macadamia was conducted by Trueman (2013). The flowers are initiated on inflorescences called pendant racemes, varying from 6 to 30 cm in length (Huett 2004; Fig. 3). A mature tree can produce about 2500 racemes, with 100–300 flowers (florets) on each raceme. Macadamia flowers are pollinated predominantly by native stingless bees and European honeybees (Trueman 2013; Howlett et al. 2015).

Macadamia is generally self-incompatible through mechanisms including protandry, though there is evidence of self-compatibility in some cultivars (Urata 1954; Sedgley et al. 1985; Sedgley et al. 1990). Self-incompatibility in plants can be controlled by several multi-allelic genes acting at different

Fig. 3 Stages of flower and nut development on a raceme in macadamia. **a** Developing florets, with looping stage shown near base of raceme. **b** Anthesis. **c** Initial nut set, with fewer nutlets than florets. **d** Developing nuts, fewer than previous stage. **e** Nuts in husk at full size. **f** Nuts dehisce from husk and fall to ground. Illustration by Todd Fox



stages of flower development (de Nettancourt 1977; Seavey and Bawa 1986). For example, self-fertility in almond (*Prunus dulcis*) is controlled by a major gene, operating in a quantitative manner (Kester and Asay 1975). Sedgley et al. (1990) found that several macadamia cultivars (predominantly *M. integrifolia*) presented inferior pollen tube growth from self-pollen compared with outcrossed pollen, as well as lower fruit set. *Macadamia tetraphylla* also showed some self-compatibility, though, again, cross-pollen produced higher seed set per raceme (Pisanu et al. 2009). Self-compatibility should be investigated in various genotypes to identify if it is a heritable trait in macadamia, as this may be a target for breeding and selection to increase pollination success. Furthermore, the relationship between self-fertility and nut yield could be a focus of research in macadamia genotypes to determine if inbreeding level affects seed set.

In Australia, flower development occurs from May to October. Bud initiation begins in May, followed by bud dormancy (50–96 days), raceme and floret elongation, style elongation and looping, and anthesis (Moncur et al. 1985). Fertilisation occurs 1 week after anthesis; however, most flowers abscise in the following 2 weeks. Fruits develop and some premature fruit drop occurs; nuts are mature about 28 weeks after anthesis (Nagao and Sakai 1990). Research is required to investigate the heritability of raceme and floret characteristics, and their correlation with yield.

Several studies have investigated the relationship between yield and flowering with variable results. Since the racemes have many florets, there are many opportunities for nuts to be set; floret number does not appear to be a limiting factor. Ito (1980) stated that about 0.3% of the flowers develop into mature, saleable nuts. However, this analysis was based on estimates of the numbers of racemes and flowers and of yield, and not replicated measurements. Further, Urata (1954) argued that it was unreasonable to count the number

of flowers per length of raceme due to the low percentage of flowers setting nuts and the low correlation between the two characters.

Trueman and Turnbull (1994) found that number of flowers per raceme affected initial fruit set in different pairs of cross-pollinated cultivars. Both cv. ‘H2’ and cv. ‘333’ racemes bearing 200 flowers when cross-pollinated with cv. ‘246’ had higher initial fruit set (22.3 and 40.3%, respectively), than control racemes (9.1 and 24.0%). For cv. ‘660’, racemes with 50 flowers set more fruits (21.6%) than those with 200 flowers (15.9%) when cross-pollinated with cv. ‘344’. In comparison, the number of fruits per raceme at the final nut set increased with number of flowers per raceme for cv. ‘660’. Trueman and Turnbull (1994) also found that for cv. ‘660’, nut and kernel fresh weights as well as KR were higher in cross-pollinated (with cv. ‘333’ and ‘246’) fruits than control racemes. These results demonstrate significant variation in pollination success between macadamia cultivars. Further research is required across many genotypes to determine if raceme and flower production and fruits per raceme limit yield in macadamia.

Xylem and phloem transport water, nutrients and photosynthates throughout plants (Campbell and Reece 2002), and thus the size of these vessels may influence the growth of limbs and fruit. Hardner et al. (2002) found that cumulative NIS yield up to year 10 was positively phenotypically correlated with girth of the trunk stem (0.59) in 40 cultivars (Table 1). The rachis (raceme stem) in macadamia enlarges after anthesis and is wider in inflorescences with high numbers of nuts than in inflorescences with low numbers of nuts (Urata 1954). Fruiting wood with larger diameters produced larger fruits in two out of six peach (*Prunus persica*) cultivars (Porter et al. 2002). The inheritance of the diameter of tree trunk, raceme stems, and fruit pedicels should be investigated, along with the relationship between yield and these characteristics.

Fruit characteristics

Nut development and abscission affect yield and profitability in macadamia (Boyton and Hardner 2002). After pollination, the pollen tube grows down to the ovary and fertilises one of the two ovules. Occasionally both ovules are fertilised, resulting in twin nuts (Sedgley 1981). The immature nut expands and oil accumulates from 80 to 165 days after anthesis (McConchie et al. 1996; Trueman et al. 2000). Nuts can drop from just after fertilisation to when they are mature (Boyton and Hardner 2002). Nuts are mature when they reach their maximum oil content, about 135 to 165 days after anthesis, or from February to March in Queensland and New South Wales, Australia (McConchie et al. 1996); commercial nut drop can continue through to September, depending on the cultivar (McConchie et al. 1997; Boyton and Hardner 2002). The husk can dehisce (split open) on the tree and the nut-in-shell falls to the ground, or the husk may fall with the shell (Hardner et al. 2009). It is economically advantageous to have cultivars in which the nut-in-husk abscises from the tree. This results in higher yield recovery due to the improved mechanical harvesting efficiency and reduced carry-over due to disease-harboured stick-tight husks.

Saleable kernel yield in macadamia is related to several component traits. Parents and their progeny are usually selected for NIS yield and KR. The nuts consist of an edible kernel enclosed by the shell, a woody testa, and husk (an outer pericarp) (Hardner et al. 2009; Fig. 2). Nut size is also an important yield component trait in almond: larger nuts correspond with higher yields per acre (Kester and Asay 1975). Topp et al. (2012) suggested selecting for high KR after 4 or 5 years as an indirect indication of future precocity. Precocity is a desirable trait in macadamia (Hardner et al. 2009), though as found in the cultivar 'Ikaika' precociousness may mean lower yields at later ages compared with other cultivars (Hamilton and Ito 1984).

It may be useful to investigate the partitioning of the tree's resources in husk, shell, and kernel. Harvest index was initially described by Donald (1962) regarding the ratio of economic yield to total biomass in grain crops. Cannell (1985) proposed that harvest index in perennial fruit trees should relate to the ratio of harvested fruit to total above-ground dry biomass. However, as only a portion of the macadamia nut is edible, then perhaps, an index based on the kernel rather than the nut-in-shell and husk should be used. As much as 30% of the moisture in a macadamia nut may be in the husk (Rosengarten 2004). Husk hardness, which influences the level of pest damage (Hardner et al. 2009), differs between cultivars (Campbell et al. 2005). It is not known whether the size of the husk affects yield. It is of interest to investigate if energy used by the tree in producing husk occurs at the expense of kernel production.

Bazzaz et al. (1987) suggested that perennial plants may not invest as much energy into reproduction as annuals as they have more opportunities to reproduce and can allocate resources to other activities such as defence. Flowering intensity has been inconsistently correlated with reserves of carbohydrates in macadamia trees (McFadyen et al. 2012). Carbohydrate resources may be depleted during flowering and fruit development, meaning that fruit set is negatively affected (Stephenson et al. 1989; Wilkie 2009).

Previous studies have reported correlations between different components of yield. Nut and kernel weight were strongly correlated in macadamia ($r_g = 0.79$, $r_p = 0.68$; Table 1) (Hardner et al. 2001; Peace 2005), and KR decreased significantly with increased shell thickness ($r_p = -0.70$) (Leverington 1962). Kernel recovery and kernel mass were moderately correlated ($r_g = 0.48$, $r_p = 0.49$, $p < 0.005$) in different cultivars (Hardner et al. 2001). Hansche et al. (1972) found that walnut crop decreased with increased nut weight ($r_p = -0.20$), after adjusting for year effect. In other species like cashew nut (*Anacardium occidentale*), the yield per tree was highly correlated with both number of nuts per panicle ($r_p = 0.844$, $p < 0.01$) and number of hermaphrodite flowers per panicle ($r_p = 0.863$, $p < 0.01$) (Aliyu 2006). In pecan (*Carya illinoensis*), Thompson and Baker (1993) found a moderately low phenotypic correlation ($r_p = 0.394$, $p < 0.002$) between KR and kernel weight, whilst Kumar et al. (2013a) found a high correlation ($r = 0.569$). The differences between these pecan studies may be due to alternate fruit bearing in the crop, differences in the study populations, or year of data collection (Thompson and Baker 1993; Kumar et al. 2013a).

No information is available on the link between macadamia yield and raceme length or number of nuts per cluster. However, fruit set per raceme varied in different cultivars (McConchie et al. 1997; Boyton and Hardner 2002). As previously mentioned, the heritability of NIS yield per tree in macadamia is low. Broad-sense heritability based on individual trees ranged from 0.06 to 0.18 for annual NIS, 0.11 to 0.20 for cumulative NIS, and 0.11 to 0.21 for cumulative kernel yield, between 4 and 10 years after planting, respectively (Table 1) (Hardner et al. 2002). Quantification of these traits in a wider group of genotypes will be beneficial for the breeding programme.

Nut size and ratio of edible nut to shell are important breeding factors in nut tree crops. Nut weight, kernel weight, and KR in macadamia were all found to have the same broad-sense heritability of 0.63 by Hardner et al. (2001); Table 1). These characteristics were also measured in 152 pecan families (Thompson and Baker 1993). Pecans are selected for thin shells and high KR, similar to macadamia. Estimates of narrow-sense heritability for nut and kernel weight in pecan were 0.35 and 0.38, respectively. In contrast, in hazelnut (*Corylus avellana*), kernel weight, KR, and relative husk length (husk:nut length) were

highly heritable (Yao and Mehlenbacher 2000; Table 1). Walnut nut weight was also very highly heritable, though crop heritability was extremely low (Hansche et al. 1972; Table 1). Kumar et al. (2013a) selected nuts from 34 pecan selections and three standard cultivars and found that broad-sense heritability for nut yield and KR (> 0.85 ; Table 1) was extremely high compared with macadamias (0.14). However, this may have been due to favourable environmental conditions during the study, rather than genetic influence in pecan (Kumar et al. 2013a).

Genetic gain for yield may be hastened by selecting for yield component traits instead of selection for yield per se. However, indirectly selecting for high yield through component traits depends on a number of factors. Firstly, success depends on the genetic variance of the component trait and its heritability, which also encompasses the accuracy of measuring the trait. Traits with high heritability will be more easily bred and selected for than traits with low heritability. Component traits should be highly correlated with yield and more easily and cheaply measured than yield (Sparnaaij and Bos 1993).

The relative efficiency of indirect selection on a trait X via direct selection for trait Y depends on the ratio of correlated response (CR_X) to direct response (R_X) (Falconer 1989):

$$\frac{CR_X}{R_X} = \frac{i_Y h_Y r_A \sigma_{AX}}{i_X h_X \sigma_{AX}}$$

Or if the selection intensities are the same, more simply, $h_Y r_A / h_X$.

Hardner et al. (2001, 2002) estimated heritability for kernel mass ($H^2 = 0.66$; therefore, $h_Y = 0.81$) and cumulative kernel yield to 10 years ($H^2 = 0.14$; $h_X = 0.37$) with a genetic correlation (r_A) between these traits of 0.30. Thus, using the above equation, for this population, the ratio of the two responses was 0.65, indicating that indirect selection using kernel mass was only 65% as efficient as direct selection for yield. A genetic correlation of > 0.46 would be needed for indirect selection of kernel mass to be more efficient than direct selection for yield. These estimates were from a population of clonally propagated elite selections and cultivars. Genetic estimates are required from segregating progeny populations to allow conclusions of the use of indirect selection in stage one breeding. More combinations of yield and component traits should be investigated to determine if correlated response to selection is promising in macadamia populations.

Genetic gain will also be affected by the stage at which the component trait can be measured and assessed in the tree: if it can be measured when the trees are juvenile, then costs may be reduced by elimination of inferior individuals prior to expensive field evaluations. Trees must flower before crosses

can be made to produce the next generation of seedlings. Therefore, selection for component traits should be coupled with selecting for early flowering (Huett 2004).

Using genomic information to accelerate genetic gains in tree crops

Identifying target genes through genome-wide association studies followed by marker-assisted selection

Economically important traits in fruit trees such as yield and quality are likely to be controlled by several multi-allelic genes or a very large number of genes (Khan and Korban 2012; Iwata et al. 2016). Genome regions identified through linkage (family based) mapping as being associated, or in linkage disequilibrium (LD), with the target or component traits are called quantitative trait loci (QTLs) (Iwata et al. 2016). These QTLs can be used to predict individuals with high breeding values which can be used in marker-assisted selection (MAS) (Lynch and Walsh 1998; Myles et al. 2009; Hayes and Goddard 2010; Muranty et al. 2014; Iwata et al. 2016). Breeding values (BVs) are the sum of the mean additive effects of all alleles in an individual (Heffner et al. 2009).

Since QTLs can be thousands of kilobases in length, multiple genes may be closely linked with the target gene (Khan and Korban 2012). Linkage drag may occur with adverse results for the breeding programme as a result of undesirable traits positioned in proximity to desired ones (Khan and Korban 2012). As such, a more directed approach is desirable for capturing significant genes using genomic methods (Savolainen and Pyhäjärvi 2007).

Genome-wide association studies (GWAS) can utilise the allelic state of unrelated individuals to detect markers linked with target traits including the broader germplasm pool rather than using family-based methods such as biparental controlled crosses, which can be impractical, laborious, and expensive (Myles et al. 2009; Iwata et al. 2016). Association mapping in the form of GWAS offers a more fine-scale approach than QTLs to identify smaller, individual markers in LD with target traits. This can overcome the detrimental effects of genetic drag as the marker intervals are shorter, as well as accounting for population structure (Rikkerink et al. 2007; Myles et al. 2009; Brachi et al. 2011; Khan and Korban 2012; Isik 2014). The incorporation of population structure and kinship information can reduce a major problem of false associations between markers and phenotypes in a GWAS (Brachi et al. 2011; Khan and Korban 2012; Iwata et al. 2016).

In GWAS, each marker is tested individually for an association with the trait (Hayes and Goddard 2010; Khan and Korban 2012; Huang and Han 2014). This process relies on markers being in LD with the genes controlling the trait

(Balding 2006), and very few markers are located within that causative locus itself (Hayes and Goddard 2010). Single nucleotide polymorphisms (SNPs) are now a commonly used genetic marker for these studies, where there is a variation in the base at a location in the genome which can be compared among individuals (Hayes and Goddard 2010; Huang and Han 2014).

Recent advancements have led to high-throughput and high-density genotyping at lowered costs per marker point for use in genomic analysis (Iwata et al. 2016). Next-generation sequencing technologies such as genotyping by sequencing can detect molecular markers for use in genomics studies like GWAS (He et al. 2014). Marker order along the genome is not strictly required, so association studies and some other genomics studies can be performed without a reference genome. This is particularly useful in novel species (Iwata et al. 2016).

GWAS for quantitative traits such as yield have shown that often there are many markers that influence the corresponding phenotype, and these can have a minor or major effect (Lee et al. 2008; Khan and Korban 2012). The gain from MAS is proportional to the variance of the trait captured by the markers and the significance of the association (Collard et al. 2005). As such, MAS has little value for traits that are complex (that is, affected by a large number of mutations all of small effect); MAS is more effective for monogenic or oligogenic traits (Luby and Shaw 2001; Hayes and Goddard 2010; Huang and Han 2014). Screening for target markers using GWAS and MAS can occur before the plants flower as only DNA is needed for the selection, and thus can substantially reduce the selection cycle and increase genetic gain by eliminating undesirable genotypes (van Nocker and Gardiner 2014).

‘DNA-informed breeding’, a term coined by Peace (2017), is becoming the convention driving breeding direction in Rosaceae crops in the USA. Previously, however, Ru et al. (2015) reviewed the opportunities and constraints of using MAS in Rosaceae breeding. They found that MAS was not yet widely applied in fruit trees, but that affordable and programme-specific testing of DNA for major trait loci at the seedling stage could be effective for the adoption of the technique. There are relatively few published studies employing GWAS in fruit trees. Recently, a study by Minamikawa et al. (2017) investigating fruit quality traits in 676 citrus individuals using 1841 SNPs found that correlated traits were controlled by several common SNPs. In apple, Kumar et al. (2013b) found significant associations in six fruit quality traits using 2500 SNPs across 1200 seedlings. SNP markers with the largest effect across linkage groups individually explained only 2% of the phenotypic variation for fruit firmness and 17% for red flesh, which was reasonably low, yet substantially more than that explained by pedigree-based analysis in many other traits (Kumar et al. 2013b). Kumar et al.

(2013b) also found two genomic regions that were linked with two pairs of fruit quality traits, suggesting a pleiotropic effect.

Further studies have employed GWAS utilising genetic markers other than SNPs. For example, Iwata et al. (2013) and Cao et al. (2012) investigated fruit quality in Japanese pear (*Pyrus pyrifolia*) and peach, respectively, using simple sequence repeat (SSR) markers. Iwata et al. (2013) detected significant associations between markers and resistance to black spot disease, spur number, and harvest time, which indicated links to major QTLs, despite the small scale of the study in terms of number of markers ($n = 162$) and cultivars ($n = 76$). Using 53 SSR markers distributed across linkage groups, Cao et al. (2012) found that the significantly associated markers detected for peach fruit quality were located nearby previously known QTLs. Between 8.1 and 14.5% of the variation in red flesh pigment was explained by four SSRs. Similar to the findings of Kumar et al. (2013b) in apple, two of the pigment markers were associated with two other sets of traits: ripening time and fruit development period, and fruit weight and flowering time (Cao et al. 2012).

A review of breeding progress in tree nut crops by Mehlenbacher (2002), including efforts of trait mapping and MAS, concluded that genetic improvement is limited by small breeding programme size. Hardner et al. (2005) evaluated the potential for MAS to improve macadamia specifically, stating that SSR markers and pedigree will be useful in detecting marker-trait associations. However, the technology in the genomics field has vastly increased and improved; many more markers can now be screened at a lower cost. GWAS and MAS regarding important component traits of yield such as nut and kernel weights and KR using SNP markers appear to be feasible to improve macadamia if the traits are controlled by few genes of moderate to large effect. Determining the number of markers and their effects for these traits should be the focus of future genomics studies, as this is currently unreported.

Yield prediction using genomic selection

Genomic selection (GS) uses genome-wide markers to capture the effects of loci that affect the target trait (Meuwissen et al. 2001). GS is best when markers such as SNPs are in high LD with genes of large effect, hence capturing a large proportion of genetic variance (Goddard 1991; Druet et al. 2014; Viana et al. 2016). A two-step process is involved. In a reference (or training) population, where individuals have both genome-wide marker genotypes and target trait phenotypes available, the effects of all markers on the trait are estimated simultaneously. The effect of each marker is used to establish a prediction equation. The equation can then be used to predict genomic estimated breeding values (GEBVs) for genotyped selection candidates, likely to be seedlings or young trees. The accuracy of GS is assessed with cross validation of the

predicted GEBV against the known and accurate phenotypes in a validation (or testing) population (Meuwissen et al. 2001).

Many simulations suggest that GS is superior to MAS and traditional phenotypic selection for complex traits (Bernardo and Yu 2007; Heffner et al. 2009; Grattapaglia and Resende 2011; Iwata et al. 2011). This is because MAS only uses markers significantly associated with a target trait, yet many yield and quality traits are often controlled by numerous minor-effect genes (Jannink et al. 2010; Kumar et al. 2012a; Iwata et al. 2016). In comparison, GS utilises all available genetic markers with no significance threshold and can therefore explain more of the genetic variability than MAS (Meuwissen et al. 2001; Viana et al. 2016). Thus, GS avoids marker effect biases and produces more highly correlated measured and predicted BVs (Meuwissen et al. 2001; Heffner et al. 2009).

GS can increase genetic gain in horticulture crops by accelerating breeding cycles (Meuwissen et al. 2001; Heffner et al. 2010; Jannink et al. 2010; Desta and Ortiz 2014). By selecting potential elite juvenile individuals, filtering candidates and only proceeding to the field with potentially high-performing trees, time, cost, and labour can be reduced. However, yield and other traits still need to be assessed over several locations before cultivars are recommended (Acquaah 2012). Selection of potential superior cultivars in the juvenile stage can also drastically reduce capital and maintenance costs (Luby and Shaw 2001; Rikkerink et al. 2007). This would be useful in macadamia where the trees do not reach full nut production until they are 8 years old (Hardner et al. 2009). Iwata et al. (2016) and Namkoong et al. (2005) recognised that a combination of traditional and marker selection strategies should be employed.

Denis and Bouvet (2013) concluded that perennial crops may have more to gain from GS than annual crops since genetic gain per unit time in perennial crops is critical for improved cultivars. There is a paucity of published studies for tree nut crops, though there has been some work conducted in citrus (Minamikawa et al. 2017), apple (Kumar et al. 2012b), oil palm (Wong and Bernardo 2008; Kwong et al. 2017), and pear (Iwata et al. 2013).

In their recent study of fruit quality traits in citrus, Minamikawa et al. (2017) obtained high ($r > 0.7$) prediction accuracies for six of the 17 traits including fruit weight. They also found that some model accuracies were trait dependent, but the genomic best linear unbiased prediction (GBLUP) model was the highest for most traits and was more accurate in predictions than MAS based on significant SNPs. Kumar et al. (2012b) investigated the use of GS in improving fruit quality traits in apple. They used 2500 high quality SNPs for 1120 seedlings, and model predictions for fruit quality were high, at 0.70 to 0.90 (Kumar et al. 2012b).

Wong and Bernardo (2008) demonstrated that gain per unit cost and time can be increased in oil palm (*Elaeis guineensis*)

through GS. Costs for GS ranged from USD\$75,000 to \$194,000 per unit gain, depending on cost per marker data point, population size, QTL number, and heritability, compared with USD\$116,000 to \$333,000 per unit gain for 19 years of phenotypic selection per cycle. Also in oil palm, Kwong et al. (2017) found that for 1218 individuals genotyped using a 200K array, GS model accuracy increased with trait heritability, ranging from 0.40 to 0.70. The results of these studies may be applicable to other tree species with long generation intervals and large planting areas.

Pear has a long juvenile period and needs to be evaluated over many years and thus Iwata et al. (2013) found could benefit from GS. They investigated nine disease resistance and fruit set traits in 76 Japanese pear cultivars using 162 markers, mostly SSRs. Prediction of GEBVs was moderately high for flesh firmness and fruit weight (0.60 and 0.53, respectively). They found that using all markers, rather than just those with significant associations as identified using GWAS, was more accurate (Iwata et al. 2013). In comparison, predictions of BVs in citrus for fruit weight and other fruit quality traits were high (> 0.7) across 106 cultivars (Iwata 2016). A corresponding GWAS detected the influence of major QTLs in all citrus fruit traits; the use of all markers was more accurate than using only significant SNPs (Iwata 2016).

Given the large amount of phenotypic data available for macadamia, and documented parentage since the first domesticated cultivars, this species is a strong candidate for GS. Macadamia has a selection cycle of 22 years and typical planting densities of 312 trees/ha (Topp et al. 2012) and would benefit greatly from this technology. Potentially, the first stage of progeny testing (Fig. 1) could be substantially reduced by genotyping seedlings, applying GS models, and only continuing further evaluations with those individuals predicted to be high yielding.

The reference population's size and structure, relationship between training and testing populations, choice of model, marker number and density, heritability of key traits, and LD span need to be assessed when employing GS in breeding. These have been considered in reviews conducted by Grattapaglia (2014) and Lin et al. (2014) on forestry and annual species. The accuracy of models can decline over subsequent generations, so it is necessary to recalibrate every few generations with new phenotypes and allelic frequencies (Goddard 2009; Viana et al. 2016).

The training population needs to be sufficiently large to enable accurate estimation of small effects across many loci and to capture all the genetic variation present in the breeding programme (Meuwissen et al. 2001). Ideally, the training and testing populations should be related or part of the same breeding programme for best results (Habier et al. 2007). Kumar et al. (2012b) divided 1120 apple seedlings into two groups for their GS evaluations: 90% of individuals for the training

population and 10% for validation. In their simulations of GS in oil palm, Wong and Bernardo (2008) used small population sizes of 30 to 70 individuals. Macadamia has a relatively small breeding population available for genomic studies; the mean number of seedlings per family in the Australian macadamia breeding programme's 'B1.2' population is 14 ($n = 1961$) (Topp et al. 2016). However, almost half of these trees have been removed from the field and are no longer available for genomic analysis. A larger second-generation progeny population ($n \approx 4000$) is currently being phenotyped and will be available for further genomic selection evaluation (B. Topp, pers. comm.).

A number of prediction models have been developed for GS: BLUP (best linear unbiased prediction), GBLUP, ridge regression, Lasso, reproducing kernel Hilbert space, and various Bayesian regressions (Jannink et al. 2010; Heslot et al. 2012). Heslot et al. (2012) suggested using a reduced set of models for implementing GS in breeding. These include a faster version of BayesB called weighted Bayesian shrinkage regression, Bayesian Lasso, and random forest. It is possible to combine different models to improve predictions; however, Heslot et al. (2012) found that combining different models did not always improve the accuracy of predictions. Cross validation is an essential step in GS to identify the accuracy of the model or to compare different models (Crossa et al. 2010; Heslot et al. 2012). Models such as GBLUP and BayesR (Erbe et al. 2012) are sound candidates for use in macadamia, assuming, respectively, a normal distribution of SNP effects using a genomic relationship matrix among candidates (GBLUP) and allowing for a small number of moderate to large effect QTL (BayesR). Testing both strategies is advisable given the paucity of data regarding the genetic nature of macadamia yield traits.

Models are affected by effective population size, heritability of the trait, and the size of the reference population (Daetwyler et al. 2008; Goddard 2009; Hayes et al. 2009). Effective population size is calculated using marker information and population heterozygosity; genetic gains are greater in species with smaller effective populations (Goddard 2009). Thus, the genetic diversity of the crop must be determined before genomics studies can begin. GEBVs are more accurately predicted when the trait is highly heritable (Hayes et al. 2009). Therefore, it is important to understand the heritability of the characteristics and its component traits. The accuracy of predicting BVs increases with the size of the reference population (Hayes et al. 2009), showing the importance of the training set in GS.

Accurate phenotyping is critical for GS; if the accuracy of phenotyping is poor, many more individuals will be need in the reference population (Desta and Ortiz 2014). Accurate phenotyping requires multiple well-characterised environments, stringent selection criteria, and large training

populations (Rikkerink et al. 2007; Xu and Crouch 2008; Resende Jr et al. 2012; Desta and Ortiz 2014). For many traits such as yield, data need to be collected across multiple years and sites and are costly and time consuming (Bernardo 2008; Stephens et al. 2009; Resende Jr et al. 2012; Xu et al. 2012; Isik 2014). This has been the case in macadamia where 2000 progeny from 47 families have been evaluated for 8 years at nine locations (Topp et al. 2016).

To implement GS in macadamia would involve growing progeny to their first leaf for DNA extraction and subsequent genotyping, hence reducing the labour and maintenance costs of growing trees to maturity. Topp et al. (2012) compared capital, maintenance, and evaluation costs, standardised to year of release of five cultivars for four breeding strategies. They found that full traditional assessment, involving the evaluation of 1200 hybrid seedlings for 9 years followed by regional variety trial for a total cycle length of 22 years, was much more expensive (\$1,545,922 net present value) with a low ratio of gain to breeding cost (\$570,000) compared to tandem selection, where seedlings are evaluated to age 7 only (\$795,508 and \$1,080,000). Their cloned seedling strategy, which evaluates 200 hybrid seedlings in a regional trial after only 2 years of initial measurements, also reduced the cycle length to 15 years, with improved breeding costs and gain to cost ratio than traditional assessment (\$986,075 and \$680,000) (Topp et al. 2012). Rapid phenotyping would be extremely useful in large tree crops with long juvenile periods like macadamia.

Genotyping costs continue to decline, with more data points becoming available, and therefore more markers likely to be in proximity to causal genes (Heffner et al. 2009; Khan and Korban 2012; Iwata et al. 2016). The cost of genotyping analysis varies with the volume of sequencing applied per sample, with the most popular services applying between one million to five million reads and the price per sample varying usually between US\$25 to US\$55 (A. Killian pers. comm.). Thus, with advancing technology, the accessibility of large numbers of molecular markers and the declining costs, the employment of and opportunity to use GS in breeding is increasing (Heffner et al. 2009; Iwata et al. 2016). Future macadamia breeding efforts should compare the costs and benefits of traditional breeding with selection strategies involving GWAS and GS.

Conclusions

The complex nature of yield in macadamia and its low heritability, as well as long cycle times, currently hinder cultivar development. Genetic improvement of yield by indirect selection for its component traits may improve breeding efficiency.

Characteristics, such as nut, kernel, and husk weight, and raceme length and width, may be more simply and accurately measured in breeding populations, especially in the years before yield per tree estimates are stable. Yield component traits can be investigated with GWAS to determine if any major markers are associated with each trait. If so, this information could be used in MAS. GS models are suitable for predicting complex traits like yield in macadamia seedlings, as well as to predict important related traits. It is essential to compare the genetic gains and the costs using these different breeding strategies, to determine which method or combination of methods are most efficient.

Acknowledgements We are grateful to Drs Craig Hardner, Mobashwer Alam, Mark Dieters, Jodi Neal, Chris Menzel, and Robert Henry for their comments. Thanks to two reviewers for helpful comments and suggestions. Thanks also to Todd Fox for illustrations and Nik Nieuwenhuis for technical support.

Data archiving statement There are no data in this review.

Funding information 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. KO received financial support from the Australian Postgraduate Award and Charles Morphet Peglar scholarship.

References

- Acquaah G (2012) Principles of plant genetics and breeding, Second edn. John Wiley & Sons, Chichester. <https://doi.org/10.1002/9781118313718>
- Aliyu O (2006) Phenotypic correlation and path coefficient analysis of nut yield and yield components in cashew (*Anacardium occidentale* L.) *Silvae Genetica* 55:19–24
- Australian Macadamia Society (2012) Global macadamia production. <http://australian-macadamias.org/industry/item/601-issue-five>. Accessed 15 Jul 2017
- Balding DJ (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet* 7(10):781–791. <https://doi.org/10.1038/nrg1916>
- Bazzaz FA, Chiariello NR, Coley PD, Pitelka LF (1987) Allocating resources to reproduction and defense. *Bioscience* 37(1):58–67. <https://doi.org/10.2307/1310178>
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48(5):1649. <https://doi.org/10.2135/cropsci2008.03.0131>
- Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. *Crop Sci* 47(3):1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Bodzon Z (2004) Correlations and heritability of the characters determining the seed yield of the long-raceme alfalfa (*Medicago sativa* L.) *J Appl Genet* 45:49–60
- Boyton S, Hardner C (2002) Phenology of flowering and nut production in macadamia. *Acta Hort* 575:381–387
- Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol* 12: 1–8
- Campbell NA, Reece JB (2002) Biology, 6th edition edn. Pearson Education, Inc., San Francisco
- Campbell AJ, Maddox CD, Morris SC (2005) Assessment protocols for nut-borer resistance—macadamia husk hardness 1999–2000. In: McConchie CA (ed) *Macadamia improvement by breeding stage 2*. Horticulture Australia Limited, Sydney, pp 205–242
- Cannell M (1985) Dry matter partitioning in tree crops. In: Cannell MGRJ, Jackson JE (eds) *Attributes of trees as crop plants*. Institute of Terrestrial Ecology, Huntington, pp 160–193
- Cantín CM, Gogorcena Y, Moreno MÁ (2010) Phenotypic diversity and relationships of fruit quality traits in peach and nectarine [*Prunus persica* (L.) Batsch] breeding progenies. *Euphytica* 171(2):211–226. <https://doi.org/10.1007/s10681-009-0023-4>
- Cao K, Wang L, Zhu G, Fang W, Chen C, Luo J (2012) Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. *Tree Genet Genomes* 8(5):975–990. <https://doi.org/10.1007/s11295-012-0477-8>
- Chagné D (2015) Chapter one—whole genome sequencing of fruit tree species. In: Christophe P, Anne-Françoise A-B (eds) *Advances in botanical research*, vol 74. Academic Press, pp 1–37
- Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142(1–2):169–196. <https://doi.org/10.1007/s10681-005-1681-5>
- Cros D, Denis M, Bouvet JM, Sánchez L (2015) Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. *BMC Genomics* 16(1):651–668. <https://doi.org/10.1186/s12864-015-1866-9>
- Crossa J, de los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186(2):713–724. <https://doi.org/10.1534/genetics.110.118521>
- Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3(10):e3395. <https://doi.org/10.1371/journal.pone.0003395>
- Denis M, Bouvet J-M (2013) Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genet Genomes* 9(1):37–51. <https://doi.org/10.1007/s11295-012-0528-1>
- Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci* 19(9):592–601. <https://doi.org/10.1016/j.tplants.2014.05.006>
- Donald C (1962) In search of yield. *J Aust Inst Agric Sci* 28:171–178
- Druet T, Macleod IM, Hayes BJ (2014) Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* 112(1):39–47. <https://doi.org/10.1038/hdy.2013.13>
- Duvick DN (1984) Genetic contributions to yield gains of U.S. hybrid maize, 1930 to 1980. In: Fehr WR (ed) *Genetic contributions to yield gains of five major crop plants*. CSSA Special Publication, vol 7. Crop Science Society of America and American Society of Agronomy, Madison, pp 15–47
- Eaton G, Kye T (1978) Yield component analysis in the cranberry. *J Am Soc Hortic Sci* 103:578–583
- Endresen DTF (2010) Predictive association between trait data and ecogeographic data for Nordic barley landraces. *Crop Sci* 50(6): 2418–2430. <https://doi.org/10.2135/cropsci2010.03.0174>
- Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA, Goddard ME (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci* 95(7):4114–4129. <https://doi.org/10.3168/jds.2011-5019>

- Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman Scientific & Technical, Essex
- Fraser J, Eaton GW (1983) Applications of yield component analysis to crop research. *Field Crops Abstracts* 36:787–797
- Goddard M (1991) Mapping genes for quantitative traits using linkage disequilibrium. *Genet Sel Evol* 23:131s–134s
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136(2):245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- Grattapaglia D (2014) Breeding forest trees by genomic selection: current progress and the way forward. In: *Genomics of plant genetic resources*. Springer, pp 651–682, doi:https://doi.org/10.1007/978-94-007-7572-5_26
- Grattapaglia D, Resende MD (2011) Genomic selection in forest tree breeding. *Tree Genet Genomes* 7(2):241–255. <https://doi.org/10.1007/s11295-010-0328-4>
- Gross C (1995) Macadamia. *Flora Aust* 16:419–425
- Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177(4):2389–2397. <https://doi.org/10.1534/genetics.107.081190>
- Hamilton RA, Fukunaga ET (1959) Growing macadamia nuts in Hawaii. *Hawaii Agr Expt Sta*
- Hamilton RA, Ito PJ (1984) Macadamia nut cultivars recommended for Hawaii. Information text series-College of Tropical Agriculture and Human Resources, University of Hawaii, Cooperative Extension Service (USA)
- Hansche PE (1983) Response to selection. In: Moore JN, Janick J (eds) *Methods in fruit breeding*. Purdue University Press, West Lafayette, pp 154–171
- Hansche P, Beres V, Forde H (1972) Estimates of quantitative genetic properties of walnut and their implications for cultivar improvement. *J Am Soc Hortic Sci* 97:279–285
- Hardner C (2015) Macadamia domestication in Hawai'i. *Genetic Resources and Crop Evolution*
- Hardner C, Winks C, Stephenson R, Gallagher E (2001) Genetic parameters for nut and kernel traits in macadamia. *Euphytica* 117(2):151–161. <https://doi.org/10.1023/A:1004016503740>
- Hardner CM, Winks CW, Stephenson RA, Gallagher EG, McConchie CA (2002) Genetic parameters for yield in macadamia. *Euphytica* 125(2):255–264. <https://doi.org/10.1023/A:1015857409317>
- Hardner C, Pisanu P, Boyton S (2004) National macadamia germplasm conservation program. Horticulture Australia,
- Hardner C, Peace C, Henshall J, Manners J (2005) Opportunities and constraints for marker-assisted selection in macadamia breeding. *Acta Hortic* 694:85–90
- Hardner C, Greaves B, Coverdale C, Wegener M Application of economic modelling to support selection decisions in macadamia. In: Mercer C (ed) 13th Australasian Plant Breeding Conference, Christchurch, New Zealand, 2006. pp 426–431
- Hardner CM, Peace C, Lowe AJ, Neal J, Pisanu P, Powell M, Schmidt A, Spain C, Williams K (2009) Genetic resources and domestication of *Macadamia*. *Hortic Rev* 35:1–126
- Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. *Genome* 53(11):876–883. <https://doi.org/10.1139/G10-076>
- Hayes B, Daetwyler H, Bowman P, Moser G, Tier B, Crump R, Khatkar M, Raadsma H, Goddard M (2009) Accuracy of genomic selection: comparing theory and results. *Proc Assoc Advmt Anim Breed Genet* 18:34–37
- He J, Zhao X, Laroche A, Z-X L, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci* 5:484
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Sci* 49(1):1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci* 50(5):1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci* 52(1):146–160. <https://doi.org/10.2135/cropsci2011.06.0297>
- Howlett BG, Nelson WR, Pattemore DE, Gee M (2015) Pollination of macadamia: review and opportunities for improving yields. *Sci Hortic* 197:411–419. <https://doi.org/10.1016/j.scienta.2015.09.057>
- Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol* 65(1):531–551. <https://doi.org/10.1146/annurev-arplant-050213-035715>
- Huett DO (2004) Macadamia physiology review: a canopy light response study and literature review. *Aust J Agric Res* 55(6):609. <https://doi.org/10.1071/AR03180>
- Igarashi M, Hatsuyama Y, Harada T, Fukasawa-Akada T (2016) Biotechnology and apple breeding in Japan. *Breed Sci* 66(1):18–33. <https://doi.org/10.1270/jsbbs.66.18>
- Isik F (2014) Genomic selection in forest tree breeding: the concept and an outlook to the future. *New For* 45(3):379–401. <https://doi.org/10.1007/s11056-014-9422-z>
- Isik F, Kumar S, Martínez-García PJ, Iwata H, Yamamoto T (2015) Acceleration of forest and fruit tree domestication by genomic selection. In: Plomion C, Adam-Blondon A-F (eds) *Advances in botanical research, land plants—trees*, vol 74. Elsevier, Oxford, pp 93–124
- Ito PJ (1980) Effect of style removal on fruit set in macadamia. *Hortscience* 15:520–521
- Iwata H (2016) Genomic selection in citrus breeding: accuracy of genomic prediction. *Plant and Animal Genome XXIV Conference*
- Iwata H, Hayashi T, Tsumura Y (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genet Genomes* 7(4):747–758. <https://doi.org/10.1007/s11295-011-0371-9>
- Iwata H, Hayashi T, Terakami S, Takada N, Sawamura Y, Yamamoto T (2013) Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breed Sci* 63(1):125–140. <https://doi.org/10.1270/jsbbs.63.125>
- Iwata H, Minamikawa MF, Kajiya-Kanegae H, Ishimori M, Hayashi T (2016) Genomics-assisted breeding in fruit trees. *Breed Sci* 66(1):100–115. <https://doi.org/10.1270/jsbbs.66.100>
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9(2):166–177. <https://doi.org/10.1093/bfgp/elq001>
- Kester DE, Asay R (1975) Almonds. In: Janick J, Moore JN (eds) *Advances in fruit breeding*. Purdue University Press, Indiana
- Khan MA, Korban SS (2012) Association mapping in forest trees and fruit crops. *J Exp Bot* 63(11):4045–4060. <https://doi.org/10.1093/jxb/ers105>
- Kumar S, Bink MC, Volz RK, Bus VG, Chagné D (2012a) Towards genomic selection in apple (*Malus × domestica* Borkh.) breeding programmes: prospects, challenges and strategies. *Tree Genet Genomes* 8(1):1–14. <https://doi.org/10.1007/s11295-011-0425-z>
- Kumar S, Chagne D, Bink MC, Volz RK, Whitworth C, Carlisle C (2012b) Genomic selection for fruit quality traits in apple (*Malus x domestica* Borkh.) *PLoS One* 7(5):e36674. <https://doi.org/10.1371/journal.pone.0036674>
- Kumar K, Nagi M, Singh D, Kaur R, Gupta R (2013a) Heritability estimates, correlation and path analysis studies for nut and kernel characters of Pecan (*Carya illinoensis* [Wang] K. Koch). *Afr J Agric Res* 8:1915–1919
- Kumar S, Garrick DJ, Bink MC, Whitworth C, Chagné D, Volz RK (2013b) Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* 14(1):393–406. <https://doi.org/10.1186/1471-2164-14-393>
- Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, Mayes S, Kulaveerasingam H, Yeoh SH, Harikrishna JA, Appleton DR

- (2017) Genomic selection in commercial perennial crops: applicability and improvement in oil palm (*Elaeis guineensis* Jacq.) *Sci Rep* 7(1):2872–2881. <https://doi.org/10.1038/s41598-017-02602-6>
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124(3):743–756
- Lee SH, van der Werf JH, Hayes BJ, Goddard ME, Visscher PM (2008) Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet* 4(10):e1000231. <https://doi.org/10.1371/journal.pgen.1000231>
- Leverington RE (1962) Evaluation of macadamia nut varieties for processing. *Qld J Agr Sci* 19:33–46
- Li C (1975) Path analysis—a primer. Boxwood Press, California
- Lin Z, Hayes BJ, Daetwyler HD (2014) Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci* 65:1177–1191
- Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *Hortscience* 36:872–879
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits, vol 1. Sinauer Sunderland, MA
- McConchie C, Meyers N, Anderson K, Vivian-Smith A, O'Brien S, Richards S (1996) Development and maturation of macadamia nuts in Australia. In: RA Stephenson CW (ed) Proceedings of the third Australian Society of Horticultural Science and the first Australian Macadamia Society Research Conference, Gold Coast, Australia. pp 234–238
- McConchie CA, Meyers N, Vithanage V, Turnbull C (1997) Pollen parent effects on nut quality and yield in macadamia
- McFadyen L, Robertson D, Sedgley M, Kristiansen P, Olesen T (2012) Time of pruning affects fruit abscission, stem carbohydrates and yield of macadamia. *Funct Plant Biol* 39(6):481–492. <https://doi.org/10.1071/FP11254>
- Mehlenbacher SA (2002) Progress and prospects in nut breeding. In: XXVI International Horticultural Congress: Genetics and Breeding of Tree Fruits and Nuts 622, pp 57–79
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829
- Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T, Imai A, Hamada H, Hayashi T (2017) Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. *Sci Rep* 7(1):4721–4734. <https://doi.org/10.1038/s41598-017-05100-x>
- Moncur MW, Stephenson RA, Trochoulias T (1985) Floral development of *Macadamia integrifolia* Maiden & Betche under Australian conditions. *Sci Hortic* 27(1-2):87–96. [https://doi.org/10.1016/0304-4238\(85\)90058-5](https://doi.org/10.1016/0304-4238(85)90058-5)
- Muranty H, Jorge V, Bastien C, Lepoittevin C, Bouffier L, Sanchez L (2014) Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. *Tree Genet Genomes* 10(6):1491–1510. <https://doi.org/10.1007/s11295-014-0790-5>
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21(8):2194–2202. <https://doi.org/10.1105/tpc.109.068437>
- Nagao MA, Sakai WS (1990) Effects of gibberellic acid, ethephon or girdling on the production of racemes in *Macadamia integrifolia*. *Sci Hortic* 43(1-2):47–54. [https://doi.org/10.1016/0304-4238\(90\)90035-D](https://doi.org/10.1016/0304-4238(90)90035-D)
- Namkoong G, Lewontin RC, Yanchuk AD (2005) Plant genetic resource management: the next investments in quantitative and qualitative genetics. *Genet Resour Crop Evol* 51(8):853–862. <https://doi.org/10.1007/s10722-005-0776-0>
- de Nettancourt D (1977) Incompatibility in angiosperms. Springer-Verlag, Berlin. <https://doi.org/10.1007/978-3-662-12051-4>
- Nock CJ, Baten A, Barkla BJ, Furtado A, Henry RJ, King GJ (2016) Genome and transcriptome sequencing characterises the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* 17(1):937. <https://doi.org/10.1186/s12864-016-3272-3>
- van Nocker S, Gardiner SE (2014) Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. *Hortic Res* 1:14022. <https://doi.org/10.1038/hortres.2014.22>
- O'Hare PTB (2010) Industry consultation helps guide macadamia breeding objectives vol 38
- O'Hare PJ, Stephenson RA, Quinlan K, Vock NT (2004) Macadamia growers handbook. Queensland Government
- Peace C (2005) Genetic characterisation of *Macadamia* with DNA markers. PhD, University of Queensland
- Peace CP (2017) DNA-informed breeding of rosaceous crops: promises, progress and prospects. *Hortic Res* 4:17006. <https://doi.org/10.1038/hortres.2017.6>
- Peace C, Ming R, Schmidt A, Manners J, Vithanage V (2008) Genomics of *Macadamia*, a recently domesticated tree nut crop. In: Moore P, Ming R (eds) Genomics of tropical crop plants, vol 1. Plant genetics and genomics: crops and models. Springer, New York, pp 313–332. https://doi.org/10.1007/978-0-387-71219-2_13
- Piepho H-P (1995) A simple procedure for yield component analysis. *Euphytica* 84(1):43–48. <https://doi.org/10.1007/BF01677555>
- Pisanu PC, Gross CL, Flood L (2009) Reproduction in wild populations of the threatened tree *Macadamia tetraphylla*: interpopulation pollen enriches fecundity in a declining species. *Biotropica* 41(3):391–398. <https://doi.org/10.1111/j.1744-7429.2008.00484.x>
- Porter G, Sherman W, Beckman T, Krewer G (2002) Fruit weight and shoot diameter relationship in early ripening peaches. *J Am Pomol Soc* 56:30
- Resende MF Jr, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MD, Kirst M (2012) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol* 193:617–624
- Rikkerink EH, Oraguzie NC, Gardiner SE (2007) Prospects of association mapping in perennial horticultural crops. In: Association mapping in plants. Springer, pp 249–269, doi:https://doi.org/10.1007/978-0-387-36011-9_11
- Rosengarten FJ (2004) The book of edible nuts. Courier Corporation
- Ru S, Main D, Evans K, Peace C (2015) Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. *Tree Genet Genomes* 11:1–12
- Samonte SOP, Wilson LT, McClung AM (1998) Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. *Crop Sci* 38(5):1130–1136. <https://doi.org/10.2135/cropsci1998.0011183X003800050004x>
- Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. *Curr Opin Plant Biol* 10(2):162–167. <https://doi.org/10.1016/j.pbi.2007.01.011>
- Seavey SR, Bawa KS (1986) Late-acting self-incompatibility in angiosperms. *Bot Rev* 52(2):195–219. <https://doi.org/10.1007/BF02861001>
- Sedgley M (1981) Early development of the *Macadamia* ovary. *Aust J Bot* 29(2):185–193. <https://doi.org/10.1071/BT9810185>
- Sedgley M, Blesing MA, Vithanage HIMV (1985) A developmental study of the structure and pollen receptivity of the macadamia pistil in relation to protandry and self-incompatibility. *Bot Gaz* 146(1):6–14. <https://doi.org/10.1086/337494>
- Sedgley M, Bell F, Bell D, Winks C, Pattison S, Hancock T (1990) Self- and cross-compatibility of macadamia cultivars. *J Hortic Sci* 65(2):205–213. <https://doi.org/10.1080/00221589.1990.11516048>
- de Souza VA, Byrne DH, Taylor JF (1998) Heritability, genetic and phenotypic correlations, and predicted selection response of quantitative traits in peach: I. An analysis of several reproductive traits. *J Am Soc Hortic Sci* 123:598–603
- Sparnaaij LD, Bos I (1993) Component analysis of complex characters in plant breeding: I. Proposed method for quantifying the relative

- contribution of individual components to variation of the complex character. *Euphytica* 70(3):225–235. <https://doi.org/10.1007/BF00023763>
- Stephens MJ, Scalzo J, Alspach PA, Beatson RA, Connor AM (2009) Genetic variation and covariation of yield and phytochemical traits in a red raspberry factorial study. *J Am Soc Hortic Sci* 134:445–452
- Stephenson R, Cull B, Mayer D (1986) Effects of site, climate, cultivar, flushing, and soil and leaf nutrient status on yields of macadamia in south east Queensland. *Sci Hortic* 30(3):227–235. [https://doi.org/10.1016/0304-4238\(86\)90101-9](https://doi.org/10.1016/0304-4238(86)90101-9)
- Stephenson RA, Gallagher EC, Rasmussen TS (1989) Effects of growth manipulation on carbohydrate reserves of macadamia trees. *Sci Hortic* 40(3):227–235. [https://doi.org/10.1016/0304-4238\(89\)90115-5](https://doi.org/10.1016/0304-4238(89)90115-5)
- Thomas R, Grafius J (1976) Prediction of heterosis levels from parental information. In: Proc. Seventh Congress of Eucarpia, pp 173–180
- Thompson T, Baker J (1993) Heritability and phenotypic correlations of six pecan nut characteristics. *J Am Soc Hortic Sci* 118:415–418
- Topp B, Hardner C, Kelly A (2012) Strategies for breeding macadamias in Australia. *Acta Hortic* 935:47–53
- Topp B, Hardner CM, Neal J, Kelly A, Russell D, McConchie C, O'Hare PJ (2016) Overview of the Australian macadamia industry breeding program
- Trueman SJ (2013) The reproductive biology of macadamia. *Sci Hortic* 150:354–359. <https://doi.org/10.1016/j.scienta.2012.11.032>
- Trueman SJ, Turnbull CGN (1994) Effects of cross-pollination and flower removal on fruit set in *Macadamia*. *Ann Bot* 73(1):23–32. <https://doi.org/10.1006/anbo.1994.1003>
- Trueman S, Richards S, McConchie C, Turnbull C (2000) Relationships between kernel oil content, fruit removal force and abscission in macadamia. *Anim Prod Sci* 40(6):859–866. <https://doi.org/10.1071/EA00004>
- Urata U (1954) Pollination requirements of macadamia. Hawaii Agric Exp Station Tech Bull 22:1–40
- Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 10(12):621–630. <https://doi.org/10.1016/j.tplants.2005.10.004>
- Viana AP, Resende MDV, Riaz S, Walker MA (2016) Genome selection in fruit breeding: application to table grapes. *Sci Agric* 73(2):142–149. <https://doi.org/10.1590/0103-9016-2014-0323>
- Westwood M (1993) Temperate-zone pomology: physiology and culture. Third edition edn. Timber Press, Inc., Portland
- Wilkie J (2009) Interactions between the vegetative growth, flowering and yield of macadamia (*Macadamia integrifolia*, *M. integrifolia* × *M. tetraphylla*), in a canopy management context. Univ of New England, Armidale, Australia, PhD Diss
- Wong C, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor Appl Genet* 116(6):815–824. <https://doi.org/10.1007/s00122-008-0715-5>
- Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci* 48(2):391–407. <https://doi.org/10.2135/cropsci2007.04.0191>
- Xu Y, Lu Y, Xie C, Gao S, Wan J, Prasanna BM (2012) Whole-genome strategies for marker-assisted plant breeding. *Mol Breed* 29(4):833–854. <https://doi.org/10.1007/s11032-012-9699-6>
- Yamamoto T, Terakami S (2016) Genomics of pear and other Rosaceae fruit trees. *Breed Sci* 66(1):148–159. <https://doi.org/10.1270/jsbbs.66.148>
- Yao Q, Mehlenbacher S (2000) Heritability, variance components and correlation of morphological and phenological traits in hazelnut. *Plant Breed* 119(5):369–381. <https://doi.org/10.1046/j.1439-0523.2000.00524.x>