Abstract. Current macadamia breeding programs involve a lengthy and laborious two-stage selection process: evaluation of a large number of unreplicated seedling progeny, followed by replicated trials of clonally propagated elite seedlings. Yield component traits, such as nut-in-shell weight (NW), kernel weight (KW), and kernel recovery (KR) are commercially important, are more easily measured than yield, and have a higher heritability. A genome-wide association study (GWAS) combined with marker-assisted selection offers an opportunity to reduce the time of candidate evaluation. In this study, a total of 281 progeny from 32 families, and 18 of their 29 parents have been genotyped for 7126 single nucleotide polymorphisms (SNP) markers. A GWAS was performed using ASReml with 4352 SNPs. We found five SNPs significantly associated with NW, nine with KW, and one with KR. Further, three of the top 10 markers for NW and KW were shared by combining GWAS with marker-assisted selection and MAS. It is worth noting, however, that candidate cultivars are selected indirectly based on genetic markers linked with desirable traits (Collard et al., 2005; Tester and Langridge 2010). Screening seedlings for markers significantly associated with important traits could predict these trait measurements years before they are actually expressed. As such, GWAS followed by MAS is a feasible way of accelerating selection cycles, by selecting candidate cultivars at an earlier stage, and thus improving genetic gain (Khan and Korban, 2012; Isik et al., 2015).

Long generation times of tree crops combined with large plant size generally means lengthy and laborious phenotyping to identify superior genotypes in tree crop breeding. Research into using GWAS for improving fruit and nut tree crop breeding is expanding, with significant associations found for fruit quality traits in Japanese pear (Iwata et al., 2013; Yamamoto et al., 2014), apple (Kumar et al., 2013), and peach (Cao et al., 2012). Recently, O’Connor et al. (2018) evaluated the potential to use genomics in macadamia breeding to improve varieties. The current study aimed to identify genetic markers associated with NW, KW, and KR in macadamia using GWAS for future use in MAS.

Materials and Methods

Plant material. The Australian macadamia breeding program’s B1.2 population is the focus of this study. The entire population included 2000 seedlings from 141 families, which were planted across nine locations in southeast Queensland and northeast New South Wales, Australia, between 2001 and 2003 (Topp et al., 2016). This study involved 281 progeny from 32 families across four of these sites in Queensland, as well as 18 of their 29 parents (n = 299).

Phenotypic analysis. Historical data were used in the study, taken from trees in 2010, when the trees were 7 to 9 years old. A sample of 50 nuts per tree was taken and dried to 1% moisture content. An average value for NW was obtained. Nuts were cracked mechanically, with kernel and shell weighed separately to calculate average KW and also KR per tree. NW and KW were both log transformed [log_{10}(X + 1)] due to the skewed nature of the data.

Genotyping and association analysis. Each tree was genotyped for 7126 SNP markers by Diversity Arrays Technology. Marker locations were unknown, as a complete reference genome is currently not available. SNPs were pruned for $\geq 50\%$ call rate and $\geq 2.5\%$ minor allele frequency across individuals, leaving 4352 markers for analysis. A genomic relationship matrix (GRM) was constructed using R to model the kinship of individuals. Association analysis was performed for the three traits using ASReml (Gilmour et al., 2009) in R, using a mixed model:

$$y = Wb + Xg + Zu + e,$$

where $y$ is a vector of phenotypes, $W$ is a matrix allocating fixed effects to individuals,
Table 1. Phenotypic minimum, maximum, mean, broad- and narrow-sense heritability for nut weight (NW), kernel weight (KW), and kernel recovery (KR).

|           | NW      | KW      | KR       |
|-----------|---------|---------|----------|
| Minimum   | 3.29 g  | 1.07 g  | 18.6%    |
| Maximum   | 12.43 g | 4.89 g  | 52.7%    |
| Mean      | 6.21 g  | 2.28 g  | 36.9%    |
| Broad-sense heritability \( (H^2) \) | 0.59    | 0.53    | 0.62     |
| Narrow-sense heritability \( (h^2) \) | 0.56 ± 0.08 | 0.51 ± 0.09 | 0.62 ± 0.10 |

Results and Discussion

NW ranged from 3.29 g to 12.43 g, with a mean of 6.21 g (Table 1; Fig. 1A). In comparison, weight of kernels varied between genotypes, from 1.07 g to 4.89 g, with an average KW of 2.28 g (Table 1; Fig. 1B). As a derivative of these two values, mean KR was 36.9%, and phenotypes ranged from 18.6% to 52.7% (Table 1).

Heritability ranged across traits from 0.51 for KW to 0.62 for KR (Table 1). Estimates of heritability for the three traits in this study were very similar to that of individual broad-sense heritability in a previous study of macadamia by Hardner et al. (2001), consisting of four replicates of 40 cultivars at three locations \( (H^2 = 0.63 \text{ for all three traits}) \). The replication of cultivars may have allowed more accurate estimates and smaller error, and hence higher heritability in the study by Hardner et al. (2001) compared with the current study. Our estimates of heritability were higher than that of NW and KW in pecan \( (h^2 = 0.35 \text{ and } 0.38, \text{ respectively}) \) (Thompson and Baker, 1993). In this study, estimates of narrow-sense and broad-sense heritability were similar for each trait (Table 1), indicating that the model could not detect any dominance. Alternatively, this could imply limited dominance exists for these traits.

NW and KW were highly correlated both phenotypically \( (0.85, \text{ } P < 0.001) \) and genetically \( (0.80, \text{ } P < 0.001) \), whereas KW and KR were moderately but significantly correlated \( (r_p = 0.37, \text{ } r_g = 0.34, \text{ } P < 0.001) \) (Table 2). KR decreased with larger NW \( (r_p = −0.16, \text{ } r_g = −0.27) \) (Table 2). This may imply that nuts with thicker shells have smaller kernels and hence lower KR.

The Q-Q plot for NW (Fig. 2A) indicates that our model has accounted for population structure, because there were a similar number of observed and expected SNPs at low levels of significance [as suggested by Korte and Farlow (2013)]. However, genomic inflation has occurred, where the SNPs are rising above the one-to-one line as \( P \) values become more stringent, suggesting polygenic inheritance of the trait (Yang et al., 2011). Five SNPs were found to be significantly associated with NW after fitting 14 SNPs in a multiple regression model (Table 3), with the reduced number of significant markers in the multiple regression likely reflecting some linkage disequilibrium between the markers. Similar to NW, some genomic inflation was apparent in the Q-Q plot for KW (Fig. 2B). Nine markers remained nonredundant after simultaneously fitting 19 markers significantly associated with KW (Fig. 2B; Table 3). Three of these 10 SNPs were also
among the most significant markers associated with NW (Table 3). One SNP was significantly associated with KR (Fig. 2C; Table 3). Because KR is a derivative of NW and KW, KR could be estimated if significant markers for NW and KW are used in MAS.

One limitation with our research was that there is currently no completely assembled macadamia genome, so our markers could not be easily mapped; however, multiple regression results suggest that several significant markers are detecting the same QTL. Research by Nock et al. (2016) has sequenced 79% of the macadamia genome. When a more complete reference genome becomes available, the location of significant markers for macadamia NW, KW, and KR can be determined. Then, MAS could be used to increase genetic gain in macadamia breeding. Future progeny seedlings could be genotyped at target markers for each trait to determine the allelic state, and breeders predict NW and KW, and sequentially cull undesirable individuals before phenotypic expression of these traits.

In conclusion, we have identified markers linked to three commercially important nut traits in macadamia through a GWAS. This study is part of a larger project investigating genetic markers controlling target traits in macadamia. Future research will examine the economic benefits of using these markers to identify parents and progeny with desirable NW, KW, and high KR. GWAS combined with MAS for important nut characteristics would be an asset for the macadamia breeding program. The efficiency and accuracy of using genomic selection to improve nut yield also should be investigated by the Australia macadamia breeding program. Application of genomics may reduce the length of the selection cycle and increase genetic gain in macadamia breeding.

The efficiency and accuracy of using genomic selection to improve nut yield also should be investigated by the Australia macadamia breeding program. Application of genomics may reduce the length of the selection cycle and increase genetic gain in macadamia breeding.

**Literature Cited**

Australian Macadamia Society. 2018. Estimated World Macadamia Production. In: XXXVII International Nut and Dried Fruit Congress, Spain.

Cao, K., L. Wang, G. Zhu, W. Fang, C. Chen, and J. Luo. 2012. Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (Prunus persica) landraces in China. Tree Genet. Genomes 8:975–990.

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196.

 Falconer, D.S. 1989. Introduction to quantitative genetics. Longman Scientific & Technical, Essex, England.

Gilmour, A.R., B. Gogel, B. Cullis, R. Thompson, and D. Butler. (2009) ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Hardner, C., C. Winks, R. Stephenson, and E. Gallagher. 2001. Genetic parameters for nut and kernel traits in macadamia. Euphytica 117:151–161.

Hardner, C.M., C.W. Winks, R.A. Stephenson, E.G. Gallagher, and C.A. McConchie. 2002. Genetic parameters for yield in macadamia. Euphytica 125:255–264.

Hardner, C.M., C. Peace, A.J. Lowe, J. Neal, P. Pisanu, M. Powell, A. Schmidt, C. Spain, and K. Williams. 2009. Genetic resources and domestication of Macadamia. Hort. Rev. 35:1–126.

Isik, F., S. Kumar, P.J. Martinez-Garcia, H. Iwata, and T. Yamamoto. 2015. Acceleration of forest and fruit tree domestication by genomic selection, p. 93–124. In: C. Plomion and A.-F. Adam-Blondon (eds.). Advances in botanical research, land plants - trees, vol 74. Elsevier, Oxford, UK.

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

Khan, M.A. and S.S. Korban. 2012. Association mapping in forest trees and fruit crops. J. Expt. Bot. 63:4045–4060.
Korte, A. and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 9:29.

Kumar, S., D.J. Garrick, M.C. Bink, C. Whitworth, D. Chagné, and R.K. Volz. 2013. Novel genomic approaches unravel genetic architecture of complex traits in apple. BMC Genomics 14:393–406.

Nock, C.J., A. Baten, B.J. Barkla, A. Furtado, R.J. Henry, and G.J. King. 2016. Genome and transcriptome sequencing characterises the gene space of *Macadamia integrifolia* (Proteaceae). BMC Genomics 17:937.

O'Connor, K., B. Hayes, and B. Topp. 2018. Prospects for increasing yield in macadamia using component traits and genomics. Tree Genet. Genomes 14:7.

Tester, M. and P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. Science 327:818–822.

Table 3. List of nonredundant significant single nucleotide polymorphisms (SNPs) and their significance values ($P$) for nut weight (NW), kernel weight (KW), and kernel recovery (KR). Bolded markers are also significant for another trait. See Supplemental Table 1 for marker sequences. SNP ID does not imply chromosome location.

| SNP    | NW          | SNP | KW          | SNP | KR          |
|--------|-------------|-----|-------------|-----|-------------|
| SNP2607 | 1.31 × 10$^{-8}$ | SNP0289 | 1.58 × 10$^{-9}$ | SNP4603 | 8.17 × 10$^{-5}$ |
| SNP4565 | 7.68 × 10$^{-7}$ | SNP4668 | 6.94 × 10$^{-9}$ |
| SNP5959 | 9.06 × 10$^{-7}$ | SNP3809 | 9.57 × 10$^{-8}$ |
| SNP6113 | 1.78 × 10$^{-5}$ | SNP2607 | 7.96 × 10$^{-8}$ |
| SNP4829 | 3.65 × 10$^{-5}$ | SNP2026 | 4.19 × 10$^{-5}$ |

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

White, I. 2013. Pin function for asreml-R: 2017.

Yang, J., M.N. Weedon, S. Purcell, G. Lettre, K. Estrada, C.J. Willer, A.V. Smith, E. Ingelsson, J.R. O’Connell, M. Mangino, R. Magi, P.A. Madden, A.C. Heath, D.R. Nyholt, N.G. Martin, G.W. Montgomery, T.M. Frayling, J.N. Hirschhorn, M.I. McCarthy, M.E. Goddard, and P.M. Visscher. 2011. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19:807–812.

Thompson, T. and J. Baker. 1993. Heritability and phenotypic correlations of six pecan nut characteristics. J. Amer. Soc. Hort. Sci. 118:415–418.

Topp, B., C.M. Hardner, J. Neal, A. Kelly, D. Russell, C. McConachie, and P.J. O’Hare. 2016. Overview of the Australian macadamia industry breeding program. Acta Hort. 1127:45–50.
Supplemental Table 1. Single nucleotide polymorphisms (SNPs) significantly associated with nut traits, Allele ID, Allele sequences for reference allele, and SNP allele, and trimmed sequence for the reference allele.

| SNP Code | Allele ID | Allele sequence reference | Allele sequence SNP | Trimmed sequence reference |
|----------|-----------|---------------------------|---------------------|----------------------------|
| SNP0289  | 19933185F0| TGCAGAGTCTAGAGACCAGTGGAGATGACGATG | AATCCATACGAGACGAGATCGTGTAGTG | AATCCATACGAGACGAGATCGTGTAGTG |
| SNP0803  | 1993521F0| TGCAGGTAGTGAAAGCTCTCTTATAAGGACGACGATG | TGCAGGTAGTGAAAGCTCTCTTATAAGGACGACGATG | TGCAGGTAGTGAAAGCTCTCTTATAAGGACGACGATG |
| SNP1196  | 1993605F0| TGCAGAAATGGGACGACACTCTTCTTTACCCAATATAG | TGCAGAAATGGGACGACACTCTTCTTTACCCAATATAG | TGCAGAAATGGGACGACACTCTTCTTTACCCAATATAG |
| SNP2026  | 1993197F0| TGCAGGATCGGAGCCGAGGAAAGGGAAGGAAACCCAACAGGAGA | TGCAGGATCGGAGCCGAGGAAAGGGAAGGAAACCCAACAGGAGA | TGCAGGATCGGAGCCGAGGAAAGGGAAGGAAACCCAACAGGAGA |
| SNP2607  | 1993670F0| TGCAGAGTGTGGAAGATGCGAGATCGGAAAGAAGGAA | TGCAGAGTGTGGAAGATGCGAGATCGGAAAGAAGGAA | TGCAGAGTGTGGAAGATGCGAGATCGGAAAGAAGGAA |
| SNP3809  | 1993341F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |
| SNP4565  | 1993540F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |
| SNP4603  | 1993197F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |
| SNP4668  | 1993091F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |
| SNP4829  | 1993082F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |
| SNP5959  | 1993268F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |
| SNP6113  | 1993091F0| TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA | TGCAGAGCTTAGGGTGGTGCTGCGAGATCGGAAAGAAGGAA |