Applications in Plant Sciences

Botanical Society of America
leading scientists and educators since 1893

GENOMIC RESOURCES ARTICLE

Single-nucleotide polymorphisms that uniquely identify cultivars of avocado (*Persea americana*)

Wiebke Kämper[1,6], Stephen J. Trueman[1], Jade Cooke[2], Naga Kasinadhuni[3], Aaron J. Brunton[4,5], and Steven M. Ogbourne[4,5]

Manuscript received 14 January 2021; revision accepted 8 June 2021.

[1] Food Security Platform, Centre for Planetary Health and Food Security, School of Environment and Science, Griffith University, Nathan, Queensland 4111, Australia

[2] Australian Genome Research Facility, Melbourne, Victoria 3000, Australia

[3] Australian Genome Research Facility, Brisbane, Queensland 4072, Australia

[4] GeneCology Research Centre, University of the Sunshine Coast, Maroochydore DC, Queensland 4558, Australia

[5] School of Science, Technology and Engineering, University of the Sunshine Coast, Maroochydore DC, Queensland 4558, Australia

[6] Author for correspondence: w.kaemper@griffith.edu.au

Citation: Kämper, W., S. J. Trueman, J. Cooke, N. Kasinadhuni, A. J. Brunton, and S. M. Ogbourne. 2021. Single-nucleotide polymorphisms that uniquely identify cultivars of avocado (*Persea americana*). *Applications in Plant Sciences* 9(6): e11440.

doi:10.1002/aps3.11440

PREMISE: Progeny of avocado (*Persea americana*) are highly variable due to high levels of heterozygosity. Breeding programs need molecular resources to allow the assessment of genetic differences and the selection of genotypes. Polymorphisms that uniquely identify different avocado cultivars provide a valuable tool to accelerate avocado research and development, including, for example, genotype selection.

METHODS: A double-digest restriction site–associated DNA sequencing (ddRADseq) approach was used to screen 10 avocado cultivars for single-nucleotide polymorphisms (SNPs). The fragments were size selected with Blue Pippin and PCR using universal Illumina primers, and catalog tags were then created with de novo alignment using Stacks software. Catalog tags were tabulated and filtered to identify alleles unique to each cultivar.

RESULTS: A total of 104 million sequences were collected, and 52 homozygous SNPs were identified that uniquely distinguished nine avocado cultivars. The cultivars Carmen Hass and Hass have a strong genetic similarity and no homozygous SNPs distinguishing these cultivars could be identified; therefore, both cultivars were grouped together.

DISCUSSION: The resource described here for cultivars of *P. americana* presents a new and significant molecular resource that can enable targeted genotype selection, paternity analysis, germplasm genotyping, pollination dynamics investigation, and crop improvement.

KEY WORDS avocado; breeding; ddRADSeq; gene flow; pollination.

Avocado, *Persea americana* Mill. (Lauraceae), is a subtropical fruit tree native to Mexico and Central America (Schaffer et al., 2013). Mexico is the largest avocado producer with 34% of total global production (FAO, 2018). Growing avocados has become more popular as the market demand for avocado products has increased exponentially in recent years (FAO, 2018). A total of 90% of global avocado exports rely on a single cultivar, Hass (Crane et al., 2013). This cultivar is often preferred by producers because of its robustness in storage and shipping, and by producers and customers because of its change in peel color from green to black, which masks minor skin imperfections but also provides an index for ripeness (Crane et al., 2013). Avocado, like many tree crops, is propagated clonally through grafting, and thus trees of a single cultivar are clones (van Nocker and Gardiner, 2014).

The reliance of an entire industry on a single cultivar poses major risks due to the lack of genetic diversity and the resultant issues associated with susceptibilities to stressors such as disease, pests, and drought. Negative impacts such as disruption to supply chains are already being felt as a result of these risks. For example, overreliance on the Hass cultivar can result in suboptimal fruit-bearing patterns if conditions are not suitable (Garner and Lovatt, 2008), with many small fruit in one year and low numbers of large fruit in the following year. Evidence of the potentially catastrophic impact of low genetic diversity within a horticultural crop is currently being experienced in the banana industry, where Panama disease or Fusarium wilt (caused by *Fusarium oxysporum*) has caused extensive crop failure across all major banana-producing areas, essentially exacerbated by the extensive use of monocultures that facilitate rapid development of the disease (Ploetz, 2015; Li et al., 2020). Avocado production is currently threatened by laurel wilt, which causes the death of mature trees in some growing regions (Kendra et al., 2011; Ayala-Silva et al., 2012). Breeding programs are trying to identify alternative avocado cultivars that still present the sought-after Hass-like features but are genetically diverse, so that they can provide security to the avocado industry.

Avocado is highly genetically heterozygous, resulting in unpredictable and highly variable progeny (Lahav and Lavi, 2013). Breeding programs are challenged by a scarcity of molecular resources and phenotypic data for cultivar identification. Advances in next-generation sequencing make screening for polymorphisms between different

 © 2021 The Authors. *Applications in Plant Sciences* published by Wiley Periodicals LLC on behalf of Botanical Society of America. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

avocado cultivars cost-effective and accessible (Kuhn et al., 2019; Talavera et al., 2019). Single-nucleotide polymorphisms (SNPs) are single-nucleotide differences in DNA sequences that can be used as molecular markers, and which can be profiled at a relatively low cost. Machine learning algorithms can then be used to link the polymorphisms to phenotypic variation, ideally leading to the identification of quantitative trait loci (QTL) that affect important crop traits (Voss-Fels et al., 2019). Testing these QTLs can significantly accelerate genotype selection compared with less targeted, conventional approaches where many generations are needed. Moreover, technologies such as the MassARRAY system (Agena Bioscience, San Diego, California, USA) can then be used to screen specific traits with large sample sizes at low cost once polymorphisms correlated to important phenotypes have been identified.

Creating tools that allow for the assessment of the genetic pool in breeding programs is critical for crop improvement. In this study, we used double-digest restriction site–associated DNA sequencing (ddRADseq) and Stacks software (Catchen et al., 2013) based on a de novo alignment to discover SNPs that identified 10 avocado cultivars and homozygous SNPs that uniquely distinguished nine avocado cultivars.

METHODS

Sample collection

Leaf samples from avocado trees were collected from multiple commercial farms, whenever possible, or from distant parts of the same farm near Childers, Queensland, Australia (Appendix 1). DNA was extracted with the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) from 42 leaf samples representing 10 cultivars. A ddRADseq approach was used to screen the leaf samples for private alleles, following the method of Peterson et al. (2012).

Establishment of a next-generation sequencing protocol

The ddRADseq methodology was optimized to determine the best restriction enzyme combination for avocado. A total of 150 ng of DNA template for each of three samples was pooled and digested with eight different restriction enzyme combinations (*Pst*I/*Msp*I, *Pst*I/*Mse*I, *Pst*/*Nla*III, *Pst*/*Hpy*Ch4IV, *Eco*RI/*Msp*I, *Eco*RI/*Mse*I, *Eco*RI/*Nla*III, and *Eco*RI/*Hpy*Ch4IV). Products were then ligated using barcoded adapters compatible with the restriction site overhang, purified using SPRIselect magnetic beads (Beckman Coulter, Indianapolis, Indiana, USA), and amplified with an indexed primer. The resulting libraries were assessed using TapeStation (Agilent Technologies, Santa Clara, California, USA) to provide the combination least likely to yield repetitive sequences.

Private allele marker discovery for avocado cultivars

A ddRADseq process was established with the optimal restriction enzyme combination (*Pst*I/*Mse*I) for three samples to establish the workflow, and then used for all leaf samples. A total of 50 ng of DNA from each of the 42 leaf samples was digested with *Pst*I/*Mse*I and cleaned using SPRIselect magnetic beads. Adapters containing a universal multiplex PCR primer (Australian Genome Research Facility, Melbourne, Victoria, Australia) and combinatorial inline barcode compatible to the restriction overhang were ligated. The digested-ligated products were pooled and purified using the QIAquick PCR Purification Kit (QIAGEN) and SPRIselect magnetic beads. Size selection was performed targeting fragments that were 280–375 bp in size using BluePippin (Sage Science, Beverly, Massachusetts, USA). Size-selected libraries were then PCR amplified with indexed sequencing primers, and finally sequenced (single end) on an Illumina NextSeq500 (Illumina, San Diego, California, USA) for 150 cycles in high-output mode.

Bioinformatics

The raw sequences were demultiplexed twice, first using the Illumina default demultiplexing pipeline and second using Stacks software (version 2.3d, http://catchenlab.life.illinois.edu/stacks/; Catchen et al., 2013). FASTQ files were automatically trimmed to the size of the shortest read minus two bases to compensate for differences in read length due to any variation in barcode sequences by Stacks software. Stacks of similar reads for each sample were created individually (also known as tags). The tags that appear across all samples were collated (i.e., catalog tags), and genotypes were calculated for the common polymorphic sites. Genotype data from Stacks were tabulated and filtered for all 42 samples to identify unique alleles present only within a specific cultivar. The following criteria were applied: (1) the private allele must be homozygous for all samples of a given cultivar, (2) there must be a call for every cultivar (i.e., no missing and/or failed data), (3) the variant position on the sequence must be ≥25 bp and ≤115 bp for downstream primer design for MassARRAY, and (4) the number of reads supporting the SNP must be >4. Sequences have been deposited in the Sequence Read Archive (SRA) on the National Center for Biotechnology Information (NCBI) database (accession: PRJNA683129).

Genetic relationships and structure

To assess the phylogenetic association among the 10 avocado cultivars, we constructed an unrooted neighbor joining (NJ) tree using the R package ape (Paradis and Schliep, 2019). Branch support for each node of the NJ tree was tested with 100 permutations, based on a pairwise Euclidean genetic distance matrix for each accession generated using the bitwise.dist function in the R package Poppr (Kamvar et al., 2014). The NJ tree topology was based on each individual accession collapsed into the respective 10 cultivars. Although there are a variety of tree-building methods (e.g., maximum likelihood and maximum parsimony), we selected the NJ tree to construct topology because it provides efficient and often similar or improved performance to the Bayesian maximum likelihood and maximum parsimony approaches (Yoshida and Nei, 2016).

We examined the genetic structure among cultivars by running a principal component analysis (PCA) from the genetic distance matrix, using the R package adegenet (Jombart and Ahmed, 2011). Additionally, we investigated admixture analysis using the R package LEA (Frichot and Francois, 2015), which applies a sparse nonnegative matrix factorization (snmf) algorithm to generate ancestry coefficients. To estimate the best representation of the ancestry coefficient (K), we ran a cross-entropy analysis between K = 1–10 values, with 10 runs for each value of K. The values of K were then visualized in a cross-entropy plot (Appendix S1) with the lowest value of K prior to an increase in the value selected as the closest estimate for the number of ancestral coefficients.

© 2021 Kämper et al.

Each of these analyses was run for all individual avocado accessions using an unfiltered variant call file (VCF), which contained the complete set of SNP data across the 10 cultivars. The result of using the complete, unfiltered SNP data to visualize the genetic associations among cultivars can be seen in the PCA plot (Fig. 1), where presumed clonal representatives of each cultivar can be individually visualized. However, this also validates the effectiveness of the genotyping platform given the tight clustering between clones in the PCA analyses.

RESULTS

A total of 104,485,411 sequences were generated using ddRADseq. From these sequences, we identified ~120,000 SNPs among the 10 cultivars, but only 52 homozygous SNPs (using the four criteria outlined above) were suitable for identifying differences among the 10 avocado cultivars (Table 1; sequences provided in Appendix S2). A total of 43 SNPs were homozygous for the respective target cultivar, while presenting a different homozygous SNP for all other cultivars.

The high similarity between the Carmen Hass and Hass cultivars resulted in no private alleles being produced among them, and therefore they were grouped together. These cultivars are likely isogenic or near isogenic, with Carmen Hass potentially being derived as an early flowering and fruiting somatic mutant of Hass (Illsley-Granich et al., 2011; Schaffer et al., 2013). We found nine homozygous SNPs for this cultivar pair, but although they presented a different homozygous SNP from most other cultivars, some cultivars presented a heterozygous SNP. The heterozygous alleles for these nine SNPs occurred, for example, in Lamb Hass and Maluma Hass, indicating that these two cultivars also have a high similarity to Hass.

Here, we provide not only the 52 identified SNPs, but also the corresponding cultivar that each SNP identified, as well as the sequence surrounding the SNP (Table 1, Appendix S2). Together,

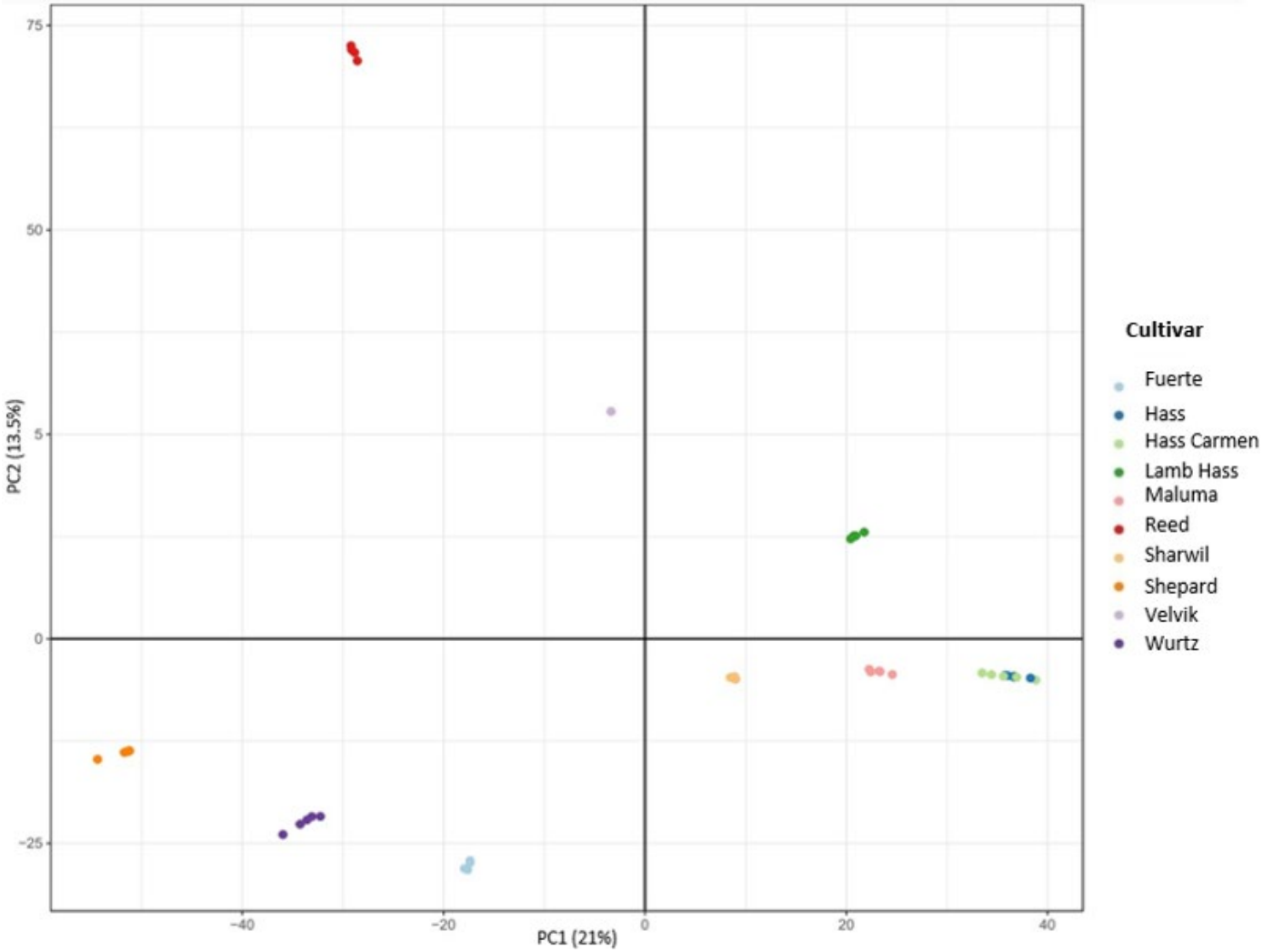

FIGURE 1. Principal component analysis (PCA) of 42 individual accessions from 10 avocado cultivars for the first two principal component axes, which account for 34.5% (axis 1 = 21%, axis 2 = 13.5%) of genetic variation.

© 2021 Kämper et al.

TABLE 1. Single-nucleotide polymorphisms (SNPs) for 10 avocado cultivars. A ddRADseq approach was used to screen 42 leaf samples from 10 avocado cultivars for private alleles: Hass (5 leaf samples), Carmen Hass (5), Lamb Hass (5), Maluma Hass (5), Reed (4), Sharwil (3), Shepard (5), Velvick (1), Wurtz (5), and Fuerte (4).[a]

Loci[b]	REF	ALT	Hass	Hass	Hass	Hass	Hass	Carmen Hass	Carmen Hass	Carmen Hass	Carmen Hass	Carmen Hass	Lamb Hass	Lamb Hass	Lamb Hass	Lamb Hass	Lamb Hass	Maluma	Maluma	Maluma	Maluma	Maluma
20474	C	G	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	CC	CC	CC	CC	CC	CG	CG	CG	CG	CG
23397	A	G	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	AG	AG	AG	AG	AG
25991	A	C	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	AA	AA	AA	AA	AC	AC	AC	AC	AC
32972	A	G	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	AG	AG	AG	AG	AG
46773	T	C	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT	TC	TC	TC	TC	TC
52608	T	A	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	TT	TT	TT	TT	TT	TA	TA	TA	TA	TA
73255	T	A	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	TA	TA	TA	TA	TA	TT	TT	TT	TT	TT
83491	G	A	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GA	GA	GA	GA	GA	GG	GG	GG	GG	GG
84299	G	T	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	GT	GT	GT	GT	GT	GG	GG	GG	GG	GG
12327	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT
20958	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA
228729	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA
31093	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA
35603	C	G	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	GG	GG	GG	GG	GG	CC	CC	CC	CC	CC
45259	A	T	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	TT	TT	TT	TT	TT	AA	AA	AA	AA	AA
75355	T	G	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	GG	GG	GG	GG	GG	TT	TT	TT	TT	TT
93019	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA
18696	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC
23841	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG
82866	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG
93949	C	T	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT
31271	T	A	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
81676	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
92177	A	T	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
116939	T	G	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
32517	T	A	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
43380	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
67950	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
111476	G	C	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
4390	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
91293	C	T	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
94477	T	G	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
94905	G	C	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
9837	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
21778	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
2603	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
2709	C	T	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
30101	G	A	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
36022	G	A	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
43359	C	A	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
5052	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
54042	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
60411	A	G	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
61596	C	A	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
67503	C	T	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
71330	G	T	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
81057	G	A	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
38381	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
46273	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
52667	T	C	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
64001	G	C	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
93565	C	T	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC

Note: REF = polymorphism in nine out of 10 cultivars; ALT = polymorphism in the target cultivar.
[a]Black cells indicate missing data. The nucleotides highlighted by black boxes represent the unique homozygous SNPs representing each cultivar (i.e., ALT).
[b]Locus provides a cross reference to the sequence surrounding the SNP (see Appendix S2).

© 2021 Kämper et al.

TABLE 1. (Continued)

Reed	Reed	Reed	Reed	Sharwil	Sharwil	Sharwil	Shepard	Shepard	Shepard	Shepard	Shepard	Velvick	Wurtz	Wurtz	Wurtz	Wurtz	Wurtz	Fuerte	Fuerte	Fuerte	Fuerte
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
AA	AA	AA	AA	AC	AC	AC	AA	AA	AA	AA	AA	AC	AA	AA	AA	AA	AA	AA	AA	AA	AA
AA	AA	AA	AA	AG	AG	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
TT	TT	TT	TT	TA	TA	TA	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
TA	TA	TA	TA	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GT	GT	GT	GT	GT	GT	GG	GG	GG	GG
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
AA	AA	AA	AA	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
CC	CC	CC	CC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
TT	TT	TT	TT	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
TT	TT	TT	TT	GG	GG	GG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
TT	TT	TT	TT	AA	AA	AA	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
AA	AA	AA	AA	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
TT	TT	TT	TT	CC	CC	CC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
GG	GG	GG	GG	GG	GG	GG	CC	CC	CC	CC	CC	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	GG	TT	TT	TT	TT	TT	TT	TT	TT	TT
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	CC	GG	GG	GG	GG	GG	GG	GG	GG	GG
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	AA	AA	AA	AA	AA	AA	AA	AA
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	TT	TT	TT	TT
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	TT	TT	TT	TT
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	[redacted]	TT	TT	TT	TT	TT	CC	CC	CC	CC
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	[redacted]	AA	AA	AA	AA	AA	GG	GG	GG	GG
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	GG	GG	GG	GG
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	AA	AA	AA	AA	CC	CC	CC	CC
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	TT	TT	TT	TT
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	TT	TT	TT	TT
AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	AA	AA	AA
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	AA	AA	AA	AA	CC	CC	CC	CC
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT	CC	CC	CC	CC
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	TT	TT	TT	TT	TT	GG	GG	GG	GG
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	GG	GG	GG	GG
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC
TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC
GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	CC	CC	CC	CC
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT

© 2021 Kämper et al.

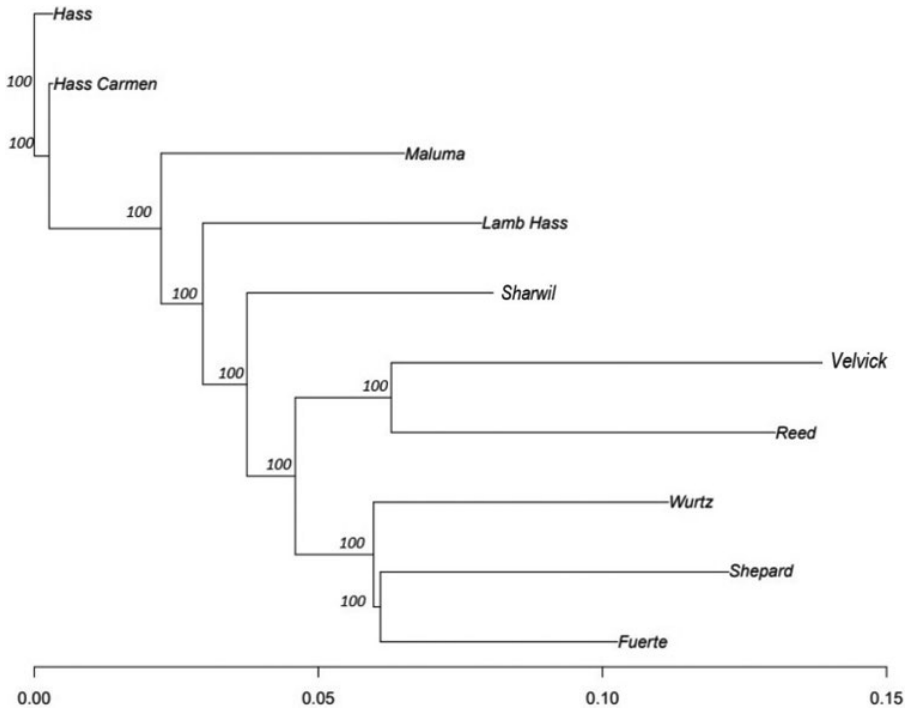

FIGURE 2. An unrooted neighbor joining tree of 10 avocado cultivars constructed from a pairwise Euclidian genetic distance matrix of 42 individual accessions.

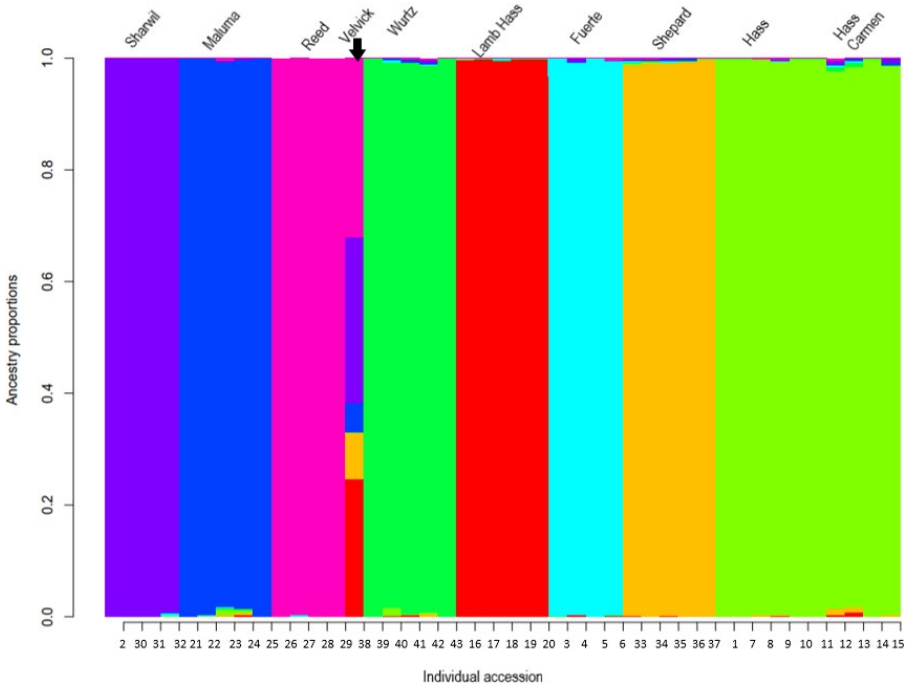

FIGURE 3. Admixture analysis barplots for $K = 8$ showing the genetic relationships within and among 42 individual accessions from 10 avocado cultivars.

these outputs provide significant cultivar-resolving power and value as a molecular resource to support research and development for the avocado industry.

Genetic relationships and structure

The NJ tree indicated that most of the 10 cultivars are genetically distinct (Fig. 2), with branch support values reported as 100% for each cultivar. Although Hass and Hass Carmen had strong bootstrap support (100%) as distinct cultivars, it is worth noting the close genetic distance between the two cultivars, which suggests that these cultivars are closely related. This is supported by the PCA analyses (Fig. 1), which showed that eight (Sharwil, Maluma, Reed, Wurtz, Lamb Hass, Fuerte, Velvick, and Shepard) of the 10 cultivars are genetically distinct, with Hass and Hass Carmen showing overlapping points in the PCA plot (Fig. 1). The close genetic relationship between the cultivars Hass and Hass Carmen is further supported by the admixture analysis, which showed the two cultivars have shared genetic structure (Fig. 3). Genetic variation in axes 1 and 2 of the PCA plot represented 34.5% of the total variation across the 10 avocado cultivars.

Application of the cross-entropy method estimated $K = 8$ (Appendix S1) as the best value to represent the ancestry coefficients among the 10 avocado cultivars. The admixture analysis of ancestry coefficients (Fig. 3) suggested that (1) eight of the 10 cultivars (Sharwil, Maluma, Reed, Wurtz, Lamb Hass, Fuerte, Velvick, and Shepard) are genetically distinct, (2) that Hass and Hass Carmen are genetically similar, and (3) that Velvick had genetic associations with Reed, Sharwil, Lamb Hass, Maluma, and Shepard.

DISCUSSION

We have demonstrated that partial genomic sequencing and bioinformatics can serve as an effective and efficient approach to detect SNPs in avocado. We identified a series of SNPs that were suitable for identifying 10 avocado cultivars and 52 homozygous SNPs that uniquely identified nine avocado cultivars. The identified SNPs provide a significant molecular resource for identifying and quantifying genetic variability in avocado cultivars and their progeny, which can be used as a tool to detect polymorphisms among different cultivars and to accelerate genotype selection (Voss-Fels et al., 2019). Large numbers of samples can be processed quickly at low cost if technologies such as MassARRAY panels are developed that target candidate SNPs linked to specific traits.

While the primary aim of this study was to identify unique SNPs that distinguish avocado cultivars, our results also

© 2021 Kämper et al.

provide a valuable contribution to understanding the genetic relationships among 10 commercially important avocado cultivars. Globally, Hass is the most important commercial cultivar (Crane et al., 2013). Hass is a hybrid between Guatemalan and Mexican landraces, with recent research revealing that Guatemalan sources represent ~39% of the Hass genome introgressed into a Mexican race background (Rendón-Anaya et al., 2019). However, international breeding and selection programs have resulted in numerous new cultivars that have become commercially significant (Schaffer et al., 2013). Several studies have reported phylogenetic relationships among a range of avocado landraces, cultivars, and accessions (Fiedler et al., 1998; Kuhn et al., 2019; Rendón-Anaya et al., 2019; Rubenstein et al., 2019) without clearly identifying the cultivars studied. As such, our study is the first to explicitly highlight the genetic relationships among the avocado cultivars Hass, Hass Carmen, Maluma, Lamb Hass, Sharwil, Velvick, Reed, Wurtz, Shepard, and Fuerte.

Our NJ tree, PCA, and admixture analyses, in combination, identified nine genetically distinct avocado cultivars. However, closer inspection suggested that eight (Maluma, Lamb Hass, Sharwil, Velvick, Reed, Wurtz, Shepard, and Fuerte) of the 10 avocado cultivars analyzed are genetically distinct. The analyses also clearly identified that Hass and Hass Carmen are closely related, and that Velvick shares genetic associations with the Reed, Sharwil, Lamb Hass, Maluma, and Shepard cultivars.

Hass Carmen was first reported as a distinct cultivar in 1986 (Illsley-Granich et al., 2011), being identified in a Hass orchard in Mexico. While appearing identical to its Hass parent, it was distinguished for its habit of asynchronous flowering and fruiting to the "original" Hass cultivar (Illsley-Granich et al., 2011). Hass Carmen became economically valuable because it produces fruit outside the regular avocado season and it therefore attracts higher prices (Illsley-Granich et al., 2011). Consequently, it was expected that the analyses would find that Hass Carmen shared genetic structure with Hass.

Velvick was supported as a distinct cultivar on the NJ tree and the PCA, yet the admixture analysis indicated this cultivar has been developed from five other cultivars (Reed, Sharwil, Lamb Hass, Shepard, and Maluma). Therefore, although our cross-validation estimate of the number of ancestral coefficients indicated eight distinct cultivars, we suggest the unique genetic structure of Velvick may support this cultivar as being genetically distinct. Mixed genetic structure of avocado cultivars was also reported by Rubenstein et al. (2019), who highlighted that three avocado landraces were relatively genetically distinct groups with genetic associations both within and between each landrace.

Overall, our analysis of the genetic relationships among 10 avocado cultivars indicated close genetic associations among cultivars, and highlighted the significance of a panel of homozygous SNPs that can uniquely identify each cultivar. The SNP panel identified in this study therefore provides an important tool for the precise delineation of avocado cultivars and a significant boost to efforts to increase genetic diversity in avocado breeding programs.

The avocado industry relies heavily on a single cultivar, Hass, which poses major risks from a lack of genetic diversity. This is further exacerbated by the recent addition of Hass Carmen, which is genetically indistinct from Hass, to the avocado industry. These risks can be reduced if the industry establishes new cultivars with higher levels of genetic diversity. Access to genetic markers that uniquely identify the genotype of germplasm and progeny will assist in achieving that aim. The SNP markers identified here allow assessment of genetic diversity in current breeding programs and provide a resource of significant value to facilitate research and development into the genetics of avocado, including paternity analysis, germplasm genotyping, investigation of pollination dynamics, genotype selection, and crop improvement.

ACKNOWLEDGMENTS

This work was funded by project PH16001 of the Hort Frontiers Pollination Fund, part of the Hort Frontiers strategic partnership initiative developed by Hort Innovation, with co-investment from Griffith University, University of the Sunshine Coast, Plant & Food Research Ltd., and contributions from the Australian Government.

AUTHOR CONTRIBUTIONS

W.K., S.J.T., and S.M.O. conceived of and designed the study. W.K., J.D., N.K., and A.J.B. performed the analysis and produced the figures and tables. W.K., A.J.B., and S.M.O. wrote the original draft and assisted W.K. in data curation. S.J.T. and S.M.O. supervised the project. All authors reviewed and edited the draft manuscript and approved the final version of the manuscript. S.J.T. and S.M.O secured funding for the work.

DATA AVAILABILITY STATEMENT

Sequences have been deposited in the Sequence Read Archive (SRA) on the National Center for Biotechnology Information (NCBI) database (accession: PRJNA683129).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Cross-entropy graph for 10 runs of each *K* value (K1–10) applied using the sparse non-negative matrix factorization (snmf) algorithm to generate ancestry coefficients of 42 individual accessions from 10 avocado cultivars.

APPENDIX S2. Sequence information for 52 single-nucleotide polymorphism (SNP) markers.

LITERATURE CITED

Ayala-Silva, T., G. G. Gordon, R. Schnell, and M. Winterstein. 2012. Application of propiconazole in management of Laurel wilt disease in avocado (*Persea americana* Mill.) trees. *Acta Horticulturae* 948: 71–78. https://doi.org/10.17660/ActaHortic.2012.948.7.

Catchen, J., P. Hohenlohe, S. Bassham, A. Amores, and W. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22(11): 3124–3140. https://doi.org/10.1111/mec.12354.

Crane, J. H., G. Douhan, B. A. Faber, M. L. Arpaia, G. S. Bender, C. F. Balerdi, and A. F. Barrientos-Priego. 2013. Cultivars and rootstocks. *In* B. Schaffer, B. N. Wolstenholme, and A. W. Whiley [eds.], The avocado: Botany, production and uses, 200–233. CABI, Wallingford, Oxfordshire, United Kingdom.

 © 2021 Kämper et al.

FAO. 2018. Statistics Division of Food and Agriculture Organization of the United Nations (FAOSTAT). Website https://www.fao.org/faostat/en/#data/QC [accessed 9 June 2021].

Fiedler, J., G. Bufler, and F. Bangerth. 1998. Genetic relationships of avocado (*Persea americana* Mill.) using RAPD markers. *Euphytica* 101: 249–255.

Frichot, E., and O. François. 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* 6: 925–929.

Garner, L. C., and C. J. Lovatt. 2008. The relationship between flower and fruit abscission and alternate bearing of 'Hass' avocado. *Journal of the American Society for Horticultural Science* 133: 3–10. https://doi.org/10.21273/jashs.133.1.3.

Illsley-Granich, C., R. Brokaw, and S. Ochoa-Ascencio. 2011. Hass Carmen®, a precocious flowering avocado tree. Proceedings of the VII World Avocado Congress, Cairns, Australia, 5–9 September 2011.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281. https://doi.org/10.7717/peerj.281.

Kendra, P. E., W. S. Montgomery, J. Niogret, J. E. Peña, J. L. Capinera, G. Brar, N. D. Epsky, and R. R. Heath. 2011. Attraction of the redbay ambrosia beetle, *Xyleborus glabratus*, to avocado, lychee, and essential oil lures. *Journal of Chemical Ecology* 37: 932–942. https://doi.org/10.1007/s10886-011-9998-0.

Kuhn, D. N., A. Groh, J. Rahaman, B. Freeman, M. L. Arpaia, N. Van den Berg, N. Abeysekara, et al. 2019. Creation of an avocado unambiguous genotype SNP database for germplasm curation and as an aid to breeders. *Tree Genetics & Genomes* 15: 71. https://doi.org/10.1007/s11295-019-1374-1.

Lahav, E., and U. Lavi. 2013. Genetics and breeding. *In* B. Schaffer, B. N. Wolstenholme, and A. W. Whiley [eds.], The avocado: Botany, production and uses, 51–85. CABI, Wallingford, Oxfordshire, United Kingdom.

Li, Z., T. Wang, C. He, K. Cheng, R. Zeng, and Y. Song. 2020. Control of Panama disease of banana by intercropping with Chinese chive (*Allium tuberosum* Rottler): Cultivar differences. *BMC Plant Biology* 20: 432. https://doi.org/10.1186/s12870-020-02640-9.

Paradis, E., and K. Schliep. 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3): 526–528.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7: e37135. https://doi.org/10.1371/journal.pone.0037135.

Ploetz, R. C. 2015. Management of Fusarium wilt of banana: A review with special reference to tropical race 4. *Crop Protection* 73: 7–15. https://doi.org/10.1016/j.cropro.2015.01.007.

Rendón-Anaya, M., E. Ibarra-Laclette, A. Méndez-Bravo, T. Lan, C. Zheng, L. Carretero-Paulet, C. A. Perez-Torres, et al. 2019. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proceedings of the National Academy of Sciences, USA* 116(34): 17081–17089. https://doi.org/10.1073/pnas.1822129116.

Rubenstein M., R. Eshed, A. Rozen, T. Zvirian, D. N. Kuhn, V. Irihimovitch, A. Sherman, and R. Ophir. 2019. Genetic diversity of avocado (*Persea americana* Mill.) germplasm using pooled sequencing. *BMC Genomics* 20: 379. https://doi.org/10.1186/s12864-019-5672-7.

Schaffer, B., B. N. Wolstenholme, and A. W. Whiley. 2013. Introduction. *In* B. Schaffer, B. N. Wolstenholme, and A. W. Whiley [eds.], The avocado: Botany, production and uses. CABI, Wallingford, Oxfordshire, United Kingdom.

Talavera, A., A. Soorni, A. Bombarely, A. J. Matas, and J. I. Hormaza. 2019. Genome-wide SNP discovery and genomic characterization in avocado (*Persea americana* Mill.). *Scientific Reports* 9: 20137. https://doi.org/10.1038/s41598-019-56526-4.

van Nocker, S., and S. E. Gardiner. 2014. Breeding better cultivars, faster: Applications of new technologies for the rapid deployment of superior horticultural tree crops. *Horticulture Research* 1: 14022. https://doi.org/10.1038/hortres.2014.22.

Voss-Fels, K. P., A. Stahl, and L. T. Hickey. 2019. Q&A: Modern crop breeding for future food security. *BMC Biology* 17: 18. https://doi.org/10.1186/s12915-019-0638-4.

Yoshida, R., and M. Nei. 2016. Efficiencies of the NJp, maximum likelihood, and Bayesian methods of phylogenetic construction for compositional and noncompositional genes. *Molecular Biology and Evolution* 33: 1618–1624. https://doi.org/10.1093/molbev/msw042.

APPENDIX 1. Voucher information for leaf samples collected in Central Queensland, Australia. Voucher specimens (USC18055–USC18096) were deposited at the University of the Sunshine Coast Herbarium.

Cultivar	Collection locality	Geographic coordinates	No. of individuals
Carmen Hass	Childers, QLD	25°13′27″S, 152°17′55″E	2
		25°13′11″S, 152°18′30″E	2
		25°13′22″S, 152°19′6″E	1
Fuerte	North Isis, QLD	25°11′5″S, 152°15′23″E	4
Hass	Goodwood, QLD	25°7′41″S, 152°22′33″E	1
		25°7′43″S, 152°22′48″E	1
	Childers, QLD	25°13′20″S, 152°18′58″E	1
	North Isis, QLD	25°11′2″S, 152°15′38″E	2
Lamb Hass	Goodwood, QLD	25°8′10″S, 152°22′29″E	2
		25°7′32″S, 152°22′7″E	3
Maluma Hass	Childers, QLD	25°13′16″S, 152°19′10″E	2
		25°13′4″S, 152°18′39″E	3
Reed	North Isis, QLD	25°11′2″S, 152°15′38″E	1
	Childers, QLD	25°13′31″S, 152°18′9″E	3
Sharwil	North Isis, QLD	25°11′2″S, 152°15′38″E	3
Shepard	Goodwood, QLD	25°7′44″S, 152°22′43″E	1
		25°7′29″S, 152°22′24″E	1
	Childers, QLD	25°13′12″S, 152°18′42″E	1
		25°13′14″S, 152°18′55″E	1
	North Isis, QLD	25°11′2″S, 152°15′38″E	1
Velvick	North Isis, QLD	25°11′2″S, 152°15′38″E	1
Wurtz	Goodwood, QLD	25°7′39″S, 152°22′19″E	2
		25°7′36″S, 152°22′11″E	1
	North Isis, QLD	25°11′2″S, 152°15′38″E	2

 © 2021 Kämper et al.