Selection of Genome-Wide SNPs for Pooled Allelotyping Assays Useful for Population Monitoring

Marielle Babineau[1], Eliza Collis[2], Angela Ruffell[3], Rowan Bunch[1], Jody McNally[1], Russell E. Lyons[2], Andrew C. Kotze[3], and Peter W. Hunt [ID][1],*

[1]CSIRO Agriculture and Food, Armidale, New South Wales, Australia

[2]School of Veterinary Science, The University of Queensland, Gatton, Queensland, Australia

[3]CSIRO Agriculture and Food, St-Lucia, Queensland, Australia

*Corresponding author: E-mail: peter.hunt@csiro.au.

Abstract

Parasitic worms are serious pests of humans, livestock, and crops worldwide. Multiple management strategies are employed in order to reduce their impact, and some of these may affect their genome and population allelic frequency distribution. The evolution of chemical resistance, ecological changes, and pest dispersal has allowed an increasing number of pests to become difficult to control with current management methods. Their lifestyle limits the use of ecological and individual-based management of populations. There is a need to develop rapid, affordable, and simple diagnostics to assess the efficacy of management strategies and delay the evolution of resistance to these strategies. This study presents a multilocus, equal-representation, whole-genome pooled single nucleotide polymorphisms (SNPs) selection approach as a monitoring tool for the ovine nematode parasite *Haemonchus contortus*. The SNP selection method used two reference genomes of different quality, then validated these SNPs against a high-quality recent genome assembly. From over 11 million high-quality SNPs identified, 334 SNPs were selected, of which 262 were species-specific, yielded similar allele frequencies when assessed as multiple individuals or as pools of individuals, and suitable to distinguish mixed nematode isolate pools from single isolate pools. As a proof-of-concept, 21 Australian *H. contortus* populations with various phenotypes and genotypes were screened. This analysis confirmed the overall low level of genetic differentiation between populations collected from the field, but clearly identifying highly inbred populations, and populations showing genetic signatures associated with chemical resistance. The analysis showed that 66% of the SNPs were necessary for stability in assessing population genetic patterns, and SNP pairs did not show linkage according to allelic frequencies across the 21 populations. This method demonstrates that ongoing monitoring of parasite allelic frequencies and genetic changes can be achieved as a management assessment tool to identify drug-treatment failure, population incursions, and inbreeding signatures due to selection. The SNP selection method could also be applied to other parasite species.

Key words: SNP discovery, *Haemonchus contortus*, pest management, allelotyping, resistance, population genetics.

Significance

Pest management rarely uses whole-genome population genetics to monitor the effects of management strategies. Here, we develop a pooled-individuals' single nucleotide polymorphism (SNP) selection pipeline in the ovine parasite, Barber's pole worm *Haemonchus contortus* and demonstrate its suitability to differentiate mixed and inbred populations due to selection pressures from management strategies. Rapidly and cheaply tracking changes in parasite population genetics as an ongoing diagnostic of management efficacy will be important for delaying the development of resistance to various management strategies. The widespread use of such SNP panels would allow the reliable detection of genomic changes in pest populations due to natural and anthropogenic selection.

Introduction

Parasites pose a serious risk to human, animal, and plant health in natural and agricultural environments. The practical application of domestic animal parasite management more often relies on managing host genetics and nutrition, minimizing exposure to infective stages, chemical control, and interventions timed to coincide with predicted lifecycle susceptibility. Season to season monitoring is often performed using visual methods including microscopy and is influenced by sampling strategies and perceptions of thresholds (Sréter et al. 1994; Besier et al. 2016). Many parasites are prophylactically controlled with chemical application or environment manipulation at set times of the year or in response to environmental circumstances (e.g., rainfall events). All management strategies have a long-term impact on pest and parasite population genetics. This is obvious with the evolution of chemical resistance, ecological shift to evade eradication, novel gene flow between populations following increased dispersal ability, hybridization, and the colonization of new environments (Wolstenholme et al. 2004; Emery et al. 2016; Hoberg and Zarlenga 2016; Rose et al. 2016). There are an increasing number of pests which have evolved the ability to evade management strategies (Tomasetto et al. 2017; Gould et al. 2018), and some parasite species persist despite globally coordinated efforts to control them (Phillips et al. 2017; Else et al. 2020). The genetic basis of many management resistance phenotypes remains elusive and complex, therefore limiting the development of genetic screening diagnostics. Integrated pest/parasite management dictates the use of multiple and diverse management solutions in order to diversify selective pressure on pests or parasites (Waller 1997; Gould et al. 2018). The effect of pest/parasite management strategies can be observed through population genetics and the evolutionary processes shaping their genomes. Several studies have argued for a better integration of population genetics into pest and parasite management (Roush and Daly 1990; Gould 1995; Rollins et al. 2006; Gilleard and Beech 2007; Porretta et al. 2007; Kirk et al. 2013; Cole and Viney 2018). Population genetic monitoring has the potential to identify shifts before the management methods become inefficient (Bailly et al. 2004; Cowled et al. 2012; Zheng et al. 2015).

The population genetics of many parasites, especially in agricultural settings, is not widely studied, and findings are often extrapolated from small and infrequent studies to devise management options without rigorous validation. Parasites present some additional challenges as they are difficult to sample due to their location within hosts, are often microscopic, and there are high number of co-occurring parasite species. Frequent genetic monitoring would allow the tailoring of personalized management advice, and assessment of the risk of parasites becoming management proof. However, the implementation of integrated pest management through population genetics requires a robust, rapid, cost-, and labor-effective diagnostic, from sampling strategy to molecular testing and manageable data analysis.

This study aimed to develop a diagnostic species-specific single nucleotide polymorphism (SNP) panel that would accurately track differences due to population mixing, management strategies, and geographical isolation. This study used the Barber's pole worm, *Haemonchus contortus*, as this species has become a model organism for parasitic nematodes (Gilleard 2013; Doyle et al. 2017). This nematode is a gastro-intestinal parasite with a worldwide distribution. Populations of this species have developed drug resistance to multiple chemical classes (Kotze and Prichard 2016). Its internal parasitic life cycle, and microscopic size for most of its development, its co-occurrence with multiple similar nematode parasites (collectively, gastrointestinal nematodes—GIN), render management through visual and ecological tracking difficult and labor intensive. The GIN most commonly observed in Australia alongside *H. contortus* are *Trichostrongylus vitrinus*, *Trichostrongylus colubriformis*, and *Teladorsagia circumcincta*, with *Oesophagostomum venulosum*, *Chabertia ovina*, *Cooperia oncophora*, and *Oesophagostomum columbianum* as the next-most common (Roeber et al. 2013).

Previous population genetics studies of *H. contortus* have shown strong population differentiation between worldwide lab-derived populations (Redman, Packard, et al. 2008) and between worldwide field populations (Troell et al. 2006; Sallé et al. 2019). Multiple studies have shown there was low population differentiation between *H. contortus* found between various mammalian host species (Cerutti et al. 2010). Australian populations of *H. contortus* have demonstrated low level genetic structure, with a geographical influence (Hunt et al. 2008). The population McMaster1931 is unique as it has been originally collected in Australia in 1931 and maintained in the laboratory since then, with passage through sheep host every couple of years followed by cryopreservation of larvae in between passages. This population predates the use of anthelmintics and is susceptible to all drugs. McMaster1931 is genetically distant from more recently collected *H. contortus* populations (Hunt et al. 2008). The population ChiswickAVRS is a laboratory-derived strain. It was originally collected in Australia in 1999 and was susceptible to anthelmintic drugs; it was then artificially selected for smooth vulval phenotype and avermectin resistance. Studies of intracontinental population structure in *H. contortus* have identified a fixation index (pairwise Fst) in the range of 0.00007 to 0.0757 (Hunt et al. 2008; Redman et al. 2015; Chaudhry et al. 2016). Population genetics studies of *H. contortus* have used a low number of loci including mitochondrial DNA, AFLPs, ITS regions of ribosomal RNA genes, and microsatellites (Otsen et al. 2001; Blouin 2002; Troell et al. 2006; Hunt et al. 2008; Redman, Grillo, et al. 2008; Gharamah et al. 2012; Yin et al. 2013). Recent whole-genome studies of intercontinental populations (including

Australian populations) have shown a pairwise Fst of 0.2–0.3 between Australian and Chinese or UK populations, and Fst of 0.16 ($\pm$0.02) between Australian and French populations (Khan et al. 2019; Sallé et al. 2019). Overall, studies have found high genetic diversity within populations, accompanied with low genetic differences between populations from the same country or continent.

The genomics of *H. contortus* is an evolving and dynamic area of research. Two genome assemblies (Laing et al. 2013; Schwarz et al. 2013) of greatly varying quality were produced, then the first genetic map (Doyle et al. 2017), followed by an additional genome assembly from a third strain (Palevich et al. 2019), and a highly resolved genome assembly in 2020 (Doyle et al. 2020). Whole-genome SNP studies with this species have revealed loci linked to ivermectin resistance (Luo et al. 2017; Sallé et al. 2019), as well as genetic differentiation between worldwide populations, and selective sweeps for anthelmintic resistance, reproduction pathways, and climate adaptation (Khan et al. 2019; Sallé et al. 2019).

Current diagnostics for infection levels with GINs including *H. contortus* involve the use of microscopy to enumerate the parasite eggs in fecal samples from the host animals (Faecal Worm Egg Count, FWEC). This method does not distinguish between *H. contortus* and at least five other genera of GIN. Therefore, where a risk of *H. contortus* is present, a culture of feces is conducted to obtain L3 stage larvae, and the genera distinguished from one another based on morphological characteristics (Love 2010; Van Wyk and Mayhew 2013) termed fecal culture and larval differentiation (FCLD). The testing for drug resistance in most cases uses the fecal egg count reduction test (FECRT) where groups of hosts are first tested to ensure they are infected, then treated with selected anthelmintics, usually in groups of 10–15 animals, and then FWEC conducted at 10-day posttreatment (Coles et al. 1992). Rarely, FCLD is used after FWEC in the FECRT test to identify which species survived the anthelmintic treatment and are therefore thought to be drug resistant. FECRT and the component FWEC and FCLD tests are labor intensive, require particular skills (especially FCLD), and have not been adopted to a high level in livestock production systems (Cabaret 2008; Vercruysse et al. 2018) due to costs and time constraints. Alternative tests are needed, and DNA-based tests have long been proposed as candidates to replace FWEC, FCLD, and FECRT (Kotze et al. 2020).

Specific mutations in target-site genes are known to confer resistance to a few drug classes in a number of parasitic nematodes, including *H. contortus*, but resistance diagnostics based on these mutations have not yet been developed for use with field populations (Kotze and Prichard 2016). Specifically, two SNPs in the β-tubulin-isotype 1 ($\beta tub1$) gene, E198A and F200Y, have been linked to increased benzimidazole drug resistance (Kotze et al. 2012) and are widespread in Australian populations. However, resistance mechanisms in *H. contortus* also involve diverse metabolic mechanisms associated with detoxification or efflux of drugs (Kotze et al. 2014). Metabolic resistance mechanisms are difficult to translate into reliable diagnostic (compared with point mutations) therefore favoring a population genetics approach such as that described by Otsen et al. (2016). A complete and unified genome assembly has only recently been made available (Doyle et al. 2017; 2020) offering the opportunity to access new molecular markers. Many studies have argued for a population genetics approach to *H. contortus* population monitoring (Hunt and Lello 2012; Emery et al. 2016).

In the present study, a pooled-samples equal-representation workflow was developed in order to select and validate a small number of randomly distributed and management-targeted whole-genome SNPs in *H. contortus*. We developed two SNPs panels, and tested their stability and reliability using 21 nematode populations, showing various resistance phenotypes, geographical origins, and inbreeding levels. As only a limited number of parasites species have high-quality genome assemblies available (International Helminth Genome Consortium 2018) the development of a pipeline with multiple and imperfect reference genomes may be useful in other circumstances. Next-generation whole-genome sequencing is not yet a convenient tool for diagnostics outside of human medicine; it requires large amount of high-quality DNA, expensive reagents, and equipment and skilled bioinformatics personnel to analyze, and generally laboratories capable of testing are located in major centers away from agricultural areas where the testing is required. In livestock breeding, allelotyping of SNPs to assess animal breeding value has become a routine procedure (ovineSNP50 and SNP600 BeadChip, Illumina, Inc). Therefore, in the interest of designing a rapid, readily usable parasite management diagnostic, the *H. contortus* SNP panel was developed for use with mass spectrometry allelotyping technology. The MassArray technology has recently been demonstrated on pooled human samples (Downes et al. 2004; Hellicar et al. 2015) but not previously in invertebrate eukaryotic organisms.

Results

SNP Panel Construction

Illumina genome sequencing was undertaken for two populations of *H. contortus*, McMaster1931 (McM) and Wallangra2003 (Wal). The Illumina sequence data aligned with a read depth of 10× or more in 26% (mean across the 11 libraries) of loci across the *H. contortus* MHCo3(ISE) reference genome, referred to as ISE (Laing et al. 2013) and in 11% of loci (mean across the 11 libraries) across the *H. contortus* McMaster genome assembly, referred to as MCM (Schwarz et al. 2013). Across the 11 libraries, 19% of loci did not have any coverage against the ISE genome, whereas 29% of loci did not have coverage against the MCM genome. The read trimming and filtering eliminated

Table 1

Alignment and Filtering Statistics for the Pooled McMaster1931 and Wallangra2003 Libraries of *Haemonchus contortus* Aligned to the ISE and MCM Reference Genomes

	Alignment to ISE		Alignment to MCM	
	Pooled 6 McMaster1931 Libraries	Pooled 5 Wallangra2003 Libraries	Pooled 6 McMaster1931 Libraries	Pooled 5 Wallangra2003 Libraries
Reads from Illumina	321,051,718	286,764,240	321,051,718	286,764,240
Remaining after trimming (%)	94	95	94	95
Remaining after alignment to reference genome (%)	87	88	88	89
Remaining after filtering (%)	60	56	69	70
Remaining after duplication removal (%)	94	95	94	94
Remaining after coverage screen (%)	80	82	73	72
Mean coverage	36.02	34.2	22.9	21.2

5% of reads on average, and approximately 88% of the remaining reads were aligned to the respective reference genome (table 1, see supplementary material S3, Supplementary Material online, for individual library statistics). The initial filtering removed 66% and 60% of reads in the samples aligned to ISE, whereas 75% of reads aligned to MCM genome were removed. After these steps the mean coverage was always above 20 reads per loci. The alignment to the ISE genome showed a higher alignment quality compared with alignment to the MCM genome demonstrated by a lower percentage of reads being rejected during filtering and a higher coverage using the ISE genome (table 1).

Alignment of the two population libraries to the ISE reference genome yielded a total of 15,804,346 variants of which 84.7% (13,397,305) were SNPs. The conservative filtering step eliminated 12% of SNPs leaving 11,843,880 high-quality SNPs, of which 11,638,195 were biallelic. The transition/transversion ratio was 1.85. Alignment of the two population libraries to the MCM reference genome yielded a total of 13,603,018 variants of which 85.1% (11,582,708) were SNPs. The conservative filtering step eliminated 15% of SNPs leaving 9,845,302 high-quality SNPs of which 9,636,253 were biallelic. The transition/transversion ratio was 1.91.

Due to the higher quality and number of SNPs discovered using the ISE reference genome, a higher number of randomly distributed SNPs were selected from this assembly, with 148 SNPs selected from the ISE assembly, and the remaining 96 selected from the MCM genome assembly. The 90 chemical-resistance candidate SNPs were selected from the ISE genome assembly. A minimum of two SNPs were selected from each genomic scaffold of the reference genome to allow comparison of SNP pairs known to be physically linked. The mean read coverage of selected SNPs was 40.84 (SD = 20.07) and 41.36 (SD = 22.38) in population Wallangra2003 and McMaster1931, respectively. Of the 244 randomly distributed SNPs, 74 had a difference in reference allele frequency between Wallangra2003 and McMaster1931 of ±0.0 to 0.2. The remaining SNPs had larger differences in reference allele frequency between the two populations, with the extremes exceeding 0.78 (fig. 1A). The mean allele frequency difference between the two populations in the randomly distributed panel was 0.31 (SD = 0.2). The SNPs also were selected to have an even distribution across the range of alternate allele frequencies in both populations from 0 to 0.5 (fig. 1B).

The 244 randomly selected SNPs and the 90 candidate drug-resistance gene SNPs were then mapped (fig. 2) back to the 2018 improved genome assembly (Doyle et al. 2017). Of the randomly distributed SNPs, 10% (25 SNPs) were located in coding regions (exons), whereas 81% (198 SNPs) were in noncoding regions (introns, intergenic, tandem repeat, and low complexity regions) (supplementary material S1, Supplementary Material online). An additional 8.6% (21 SNPs) did not map to the 2018 assembly and therefore do not have an annotation. Of the 90 candidate drug-resistance gene SNPs selected, 66% (60 SNPs) were found in coding regions of the target gene previously identified in the ISE assembly, and 27% (25 SNPs) were in introns of the target gene (supplementary material S1, Supplementary Material online). All chemical resistance candidate gene SNPs were successfully mapped to the improved genome assembly.

The mean difference in alternate allele frequency across all 313 SNPs (which successfully amplified and multiplexed) between the estimation from Illumina compared with Sequenom was −0.22 (SD = 0.39) and −0.23 (SD = 0.39) in population McMaster1931 and Wallangra2003 respectively (fig. 3). The next-generation sequence (NGS) to Sequenom allele frequency difference between the chemical-resistance SNPs and randomly distributed SNPs was not significant for either population (McMaster1931: $t = 0.36$, df = 93, $P = 0.71$, Wallangra2003: $t = 0.90$, df = 86, $P = 0.36$) and also not significant between the two populations ($t = -0.5$, df = 485, $P = 0.57$). For approximately 15% of the SNPs, the
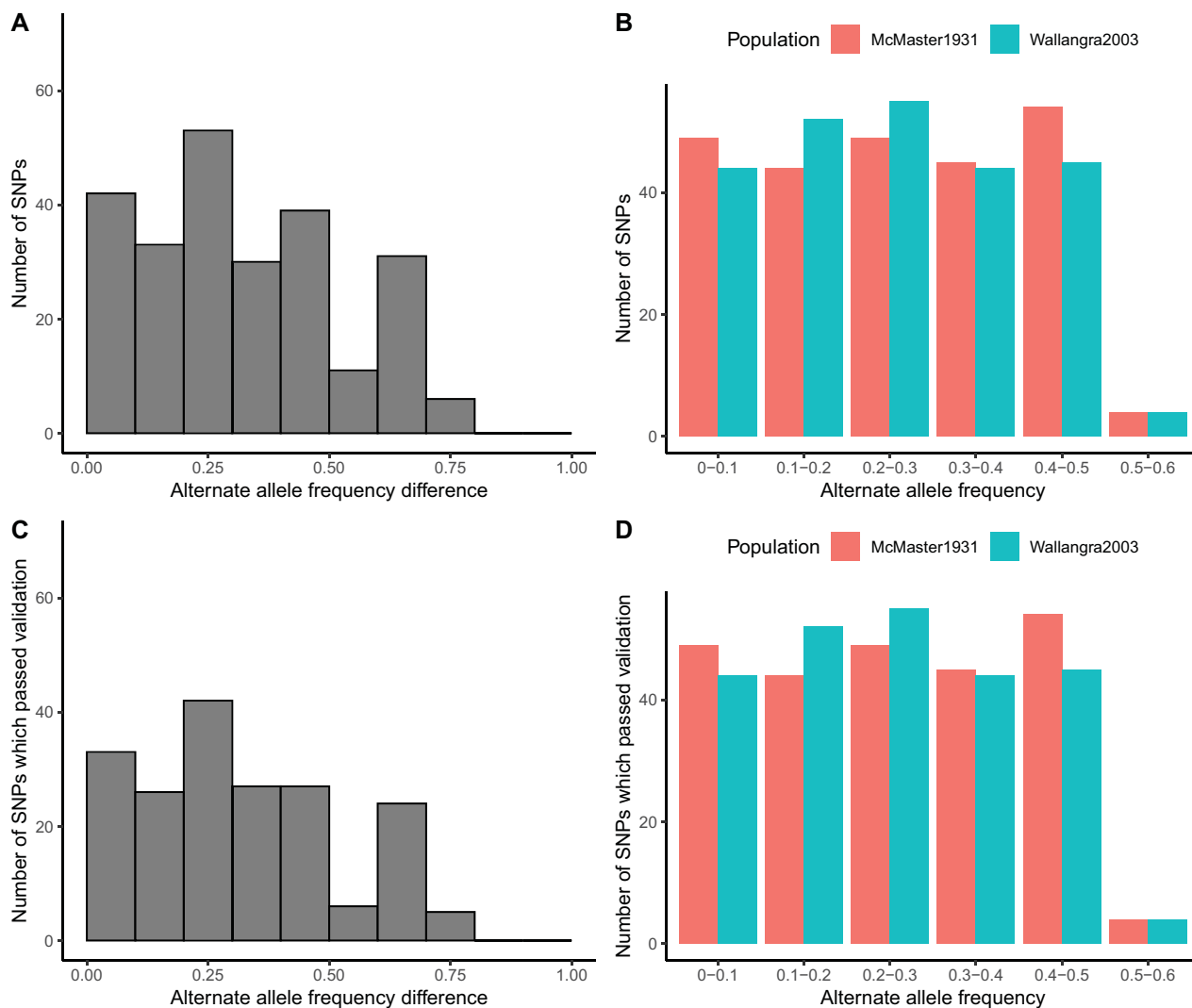
Fig. 1.—Comparison of the randomly selected SNPs across the two *Haemonchus contortus* populations. (A) Alternate allele frequency difference between population McMaster1931 and Wallangra2003 for the 244 SNPs. (B) Distribution of SNPs from populations McMaster1931 and Wallangra2003 across alternative allele frequency value for the 244 SNPs. (C) Alternate allele frequency difference between population McMaster1931 and Wallangra2003 for the 189 randomly selected SNPs which passed validation steps. (D) Distribution of SNPs from populations McMaster1931 and Wallangra2003 across alternative allele frequency values for the 189 SNPs which passed validation.

difference between the alternate allele frequency estimated from NGS compared with the allelotyping method was less than 0.1, and 30% of the SNPs had a less than 0.2 frequency difference (fig. 3). Both populations showed 33 SNPs that had a difference greater than −0.8. The high number of SNPs with a negative difference indicated the Illumina method tended to underestimate the pooled alternate allele frequency in both populations.

The Illumina to Sequenom allele frequency difference for the SNPs which passed the validation steps (see below) yielded similar results: mean difference of 0.2 (SD = 0.4) with no significant difference between the two populations ($t = -0.7$, $df = 391$, $P = 0.5$) or between randomly distributed and chemical-resistance SNPs ($t = 1.3$, $df = 64$, $P = 0.2$). This result indicates that the SNPs which failed the three validation steps

had a wide range of allele frequency differences between the Illumina and Sequenom platforms. The whole panel Ti/Tv ratio was 0.87 across both SNPs sets, the randomly distributed had a ratio of 1.01 and the chemical resistance had a ratio of 0.57. The 189 randomly distribution SNPs which passed validation had a Ti/Tv of 1.16.

SNPs Panel Validation

The validation steps eliminated a total of 22% (72 SNPs) of the 334 selected SNPs evaluated (table 2): 55 from the randomly selected and 17 from the candidate resistance set. Eleven SNPs were eliminated for more than one criterion. The identity of the SNPs which passed, or failed quality control and were eliminated, is indicated in supplementary material
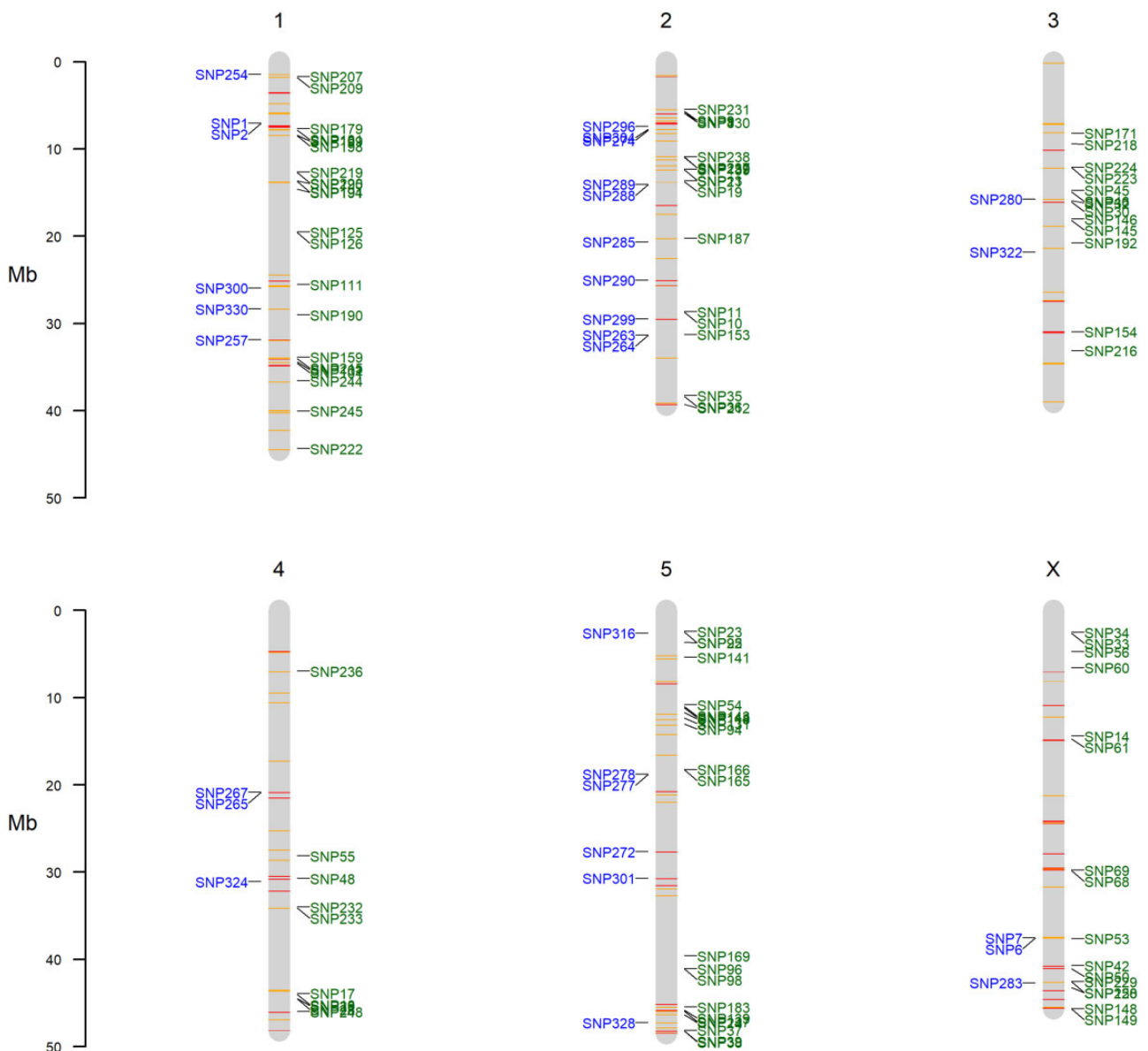
FIG. 2.—Genomic distribution of the randomly selected (green) and chemical resistance (blue) SNPs which passed the validation steps. SNPs that failed the validation steps (red) and SNPs that passed validation but did not amplify in 19 or more of the 21 Australian populations (orange) are also shown. Alignment is to *Haemonchus contortus* genome build as Doyle et al. (2017).

S1, Supplementary Material online. A total of 21 SNPs were eliminated as they failed the multiplexing or the amplification of the allelotyping reaction (table 2). A total of 20 SNPs were eliminated as they amplified in reactions using template DNA from a nontarget species. Many of the eliminated SNPs amplified in multiple co-occurring species. The species *Tel. circumcincta*, *T. colubriformis*, *O. columbianum*, and *C. ovina* all amplified in the 20 SNPs eliminated, whereas 14, 13, and 12 SNPs amplified for the species *O. venulosum*, *Ovis aries*, and *Coo. oncophora*, respectively.

A total of 13 SNPs were eliminated as they failed to accurately detect artificially admixed populations, showing a

frequency difference greater than 0.25 (supplementary material S4, Supplementary Material online) in the 50:50 validation test measuring the difference between the estimated 50%/50% admixed population and the observed 50%/50% population alternate allele frequency. The McMaster1931 and Wallangra2003 populations showed the highest mean 50:50 test difference with 0.076 (SD across all SNPs of 0.09), whereas GoldCoast2004 and Goondiwindi2011 showed the lowest difference of 0.048 (with a SD of 0.06 across SNPs), and Mackay2009 and Cannawigara2006 showed a 50:50 difference of 0.065 (SD = 0.091). Of the 13 eliminated SNPs, the mean of the 50:50 test difference was 0.40. The greatest
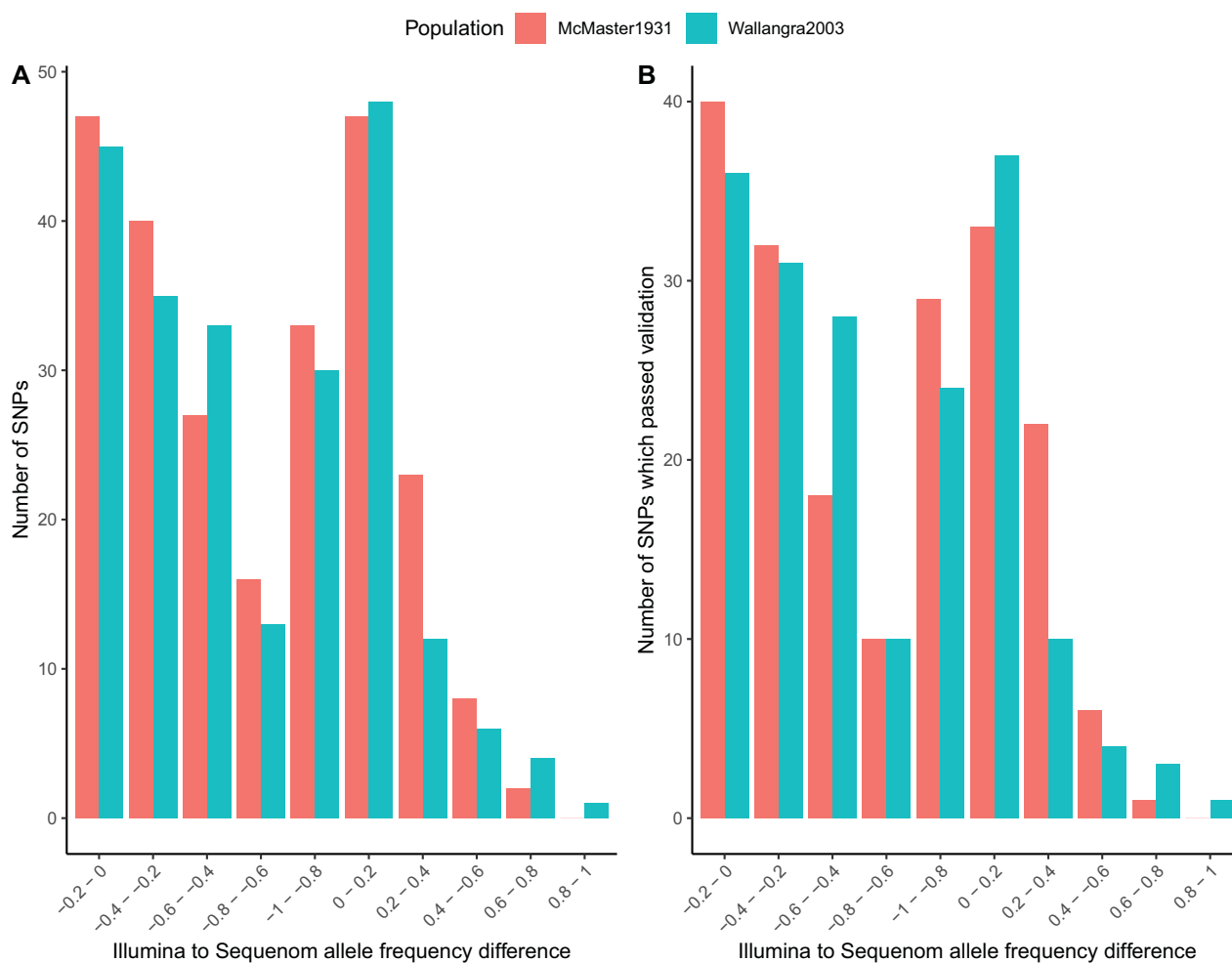
FIG. 3.—Alternate allele frequency difference between the NGS Illumina and Sequenom allelotyping methods for *Haemonchus contortus* populations McMaster1931 and Wallangra2003 for (*A*) 334 SNPs and (*B*) 262 SNPs that passed validation.

Table 2
Number of SNPs Eliminated due to Different Validation Steps

Validation Step	Randomly Selected SNPs	Chemical-Resistance SNPs	Total Number of SNP
Multiplex creation	13	0	13
Amplification	8	0	8
Pooling versus individual	24	3	27
Admixed population detection	7	6	13
Co-occurring species	9	11	20
Passed	189	73	262

NOTE.—Numbers include the SNPs that failed for multiple steps.

50:50 difference was 0.96. Detection of a gradient pattern from the 100%, 75%, 50%, 25%, and 0% artificially admixed populations was also assessed by eye for each SNP.

A total of 27 SNPs were eliminated as they failed the representation of pooled sample validation by showing a higher than 0.2 difference in allele frequency between pooled and individual samples from either the McMaster1931 or Wallangra2003 populations (table 2). *t*-Tests indicated that across all SNPs, the pooled versus individual's allele frequency was not significantly different in the McMaster1931 ($t = -0.81$, df $= 335$, $P = 0.41$) and Wallangra2003 populations ($t = -1.19$, df $= 443$, $P = 0.23$). Across both populations and across all SNPs, there was a significant correlation between the pooled sample allele frequency and the mean of
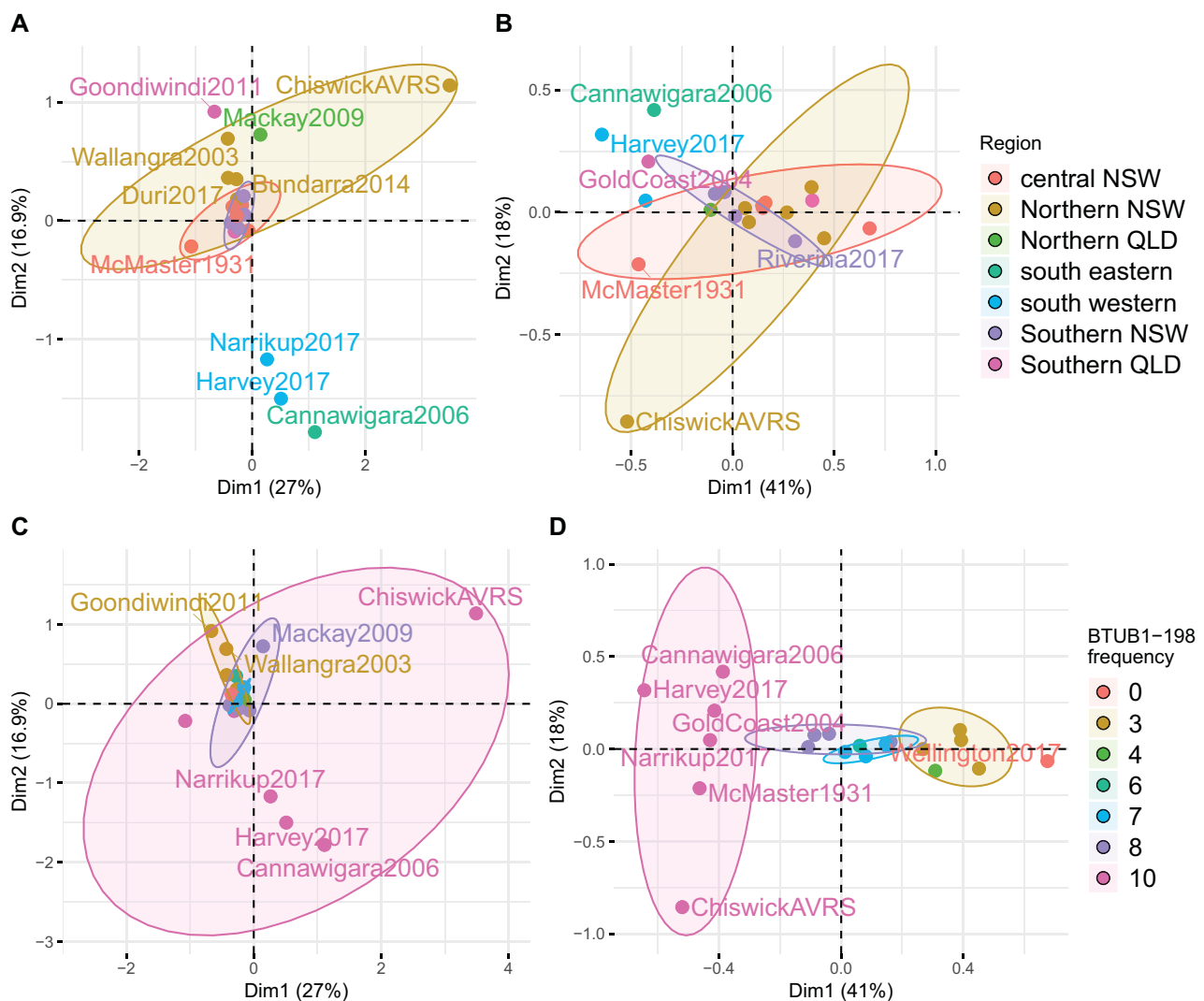
FIG. 4.—Principal component analysis (PCA) of 21 field populations. Derived using geographical region grouping from randomly selected SNPs (A) and chemical-resistance SNPs (B). Derived using BTUB1-198 allelic frequency in bins of 0.1 from randomly selected SNPs (C) and chemical-resistance SNPs (D).

the 30 individual allele frequencies ($R^2 = 0.78$, Fstat = 885, df = 539, $P = 1.3e-32$). The linear regression intercept and slope were 0.054 and 0.84 respectively. Approximately 50% of the SNPs (48% in McMaster1931 and 52% in Wallangra2003) showed that pooled worm samples underestimated the minor allele frequency compared with individual worm samples (negative allele differences), though these differences were not statistically significant. The mean pool to individuals sample difference across all SNPs was −0.02 (SD = 0.25) and −0.013 (SD = 0.26) in McMaster1931 and Wallangra2003, respectively. The highest pool to individuals sample difference was for SNP83 with a difference of 0.84.

Spatially Dispersed Population Differentiation

Amplification of the validated 262 SNPs across the 21 Australian populations demonstrated that 119 SNPs did not amplify reliably in three or more populations and were therefore eliminated from the population genetics analysis. Therefore 143 high-quality, reliable SNPs, including 113 randomly distributed and 30 putative resistance SNPs, were included in the final analysis. Of the randomly distributed SNPs retained, 4.5% (11) are located in exons, whereas 80% (90 SNPs) are in noncoding regions (introns or intergenic regions) (supplementary material S1, Supplementary Material online). An additional 5% (12) did not map to the 2018 assembly and therefore do not have an annotation. Of the 30 putative drug-resistance SNPs retained, 70% (21 SNPs) were found in coding regions of the target gene previously identified in the ISE assembly (supplementary material S1, Supplementary Material online). These 143 SNPs were distributed on all six scaffolds of the *H. contortus* genome (fig. 2).

There was no correlation between pairwise SNP physical distance on the chromosomes and the variance of alternate

allele frequency across the 21 populations ($R^2 = 0.004$, Fstat = 8.76, df = 1,714, $P = 0.003$). Variance ranged from 0 to 0.23 with 616 SNP pairs (36% of all possible SNP pairs) with variance lower than 0.05. These low variance SNP pairs were distributed across all chromosomes (chrom1: 156 pairs, chrom2:120, chrom3:80, chrom4:22, chrom5:154, chromX:84) and ranged in distance from 434 bases to 43,892,631 bases. The SNP pair with the lowest variance (SNP13 and SNP36) were 26,010,943 bases away from each other on chromosome 2. In contrast, the SNP pair with the highest variance (SNP125 and SNP126) were 14,500 bases away on chromosome 1.

The SNP allelotyping data from 21 populations were analyzed using a principal component analysis (PCA) for the entire set and for the randomly selected and the chemical resistance SNPs as two additional analyses (fig. 4). The first two axis components explained 43.9% and 59% of the population variation for the randomly selected and chemical-resistance SNPs sets respectively. A total of ten and six dimensions were necessary to encompass at least 90% of the variation in the randomly selected and chemical-resistance sets, respectively. The ten SNPs with the highest contribution to the population variation in the randomly selected SNPs allele frequency were an equal number of SNPs from the MCM and ISE genome assembly and ranged from 4.7% (SNP33) to 2.4% (SNP202). The ten SNPs with the highest contribution to the population variation for the putative drug resistance gene SNPs ranged from 41% (BTUB1-198) to 0.2% (SNP265) and included the BTUB1-200, two cytochrome P450 SNPs and two nAChR SNPs.

Geographical location and the BTUB1-198 allele frequencies were overlaid on the PCA pattern to attempt to identify single variables as the explanatory factor (fig. 4, see supplementary material S2, Supplementary Material online, for population detail). For the geographical origin variable, the NSW populations formed a cluster using both the randomly distributed and chemical-resistance SNPs sets (fig. 4A and B). The two WA and the two QLD populations were close together using the randomly distributed SNPs, whereas the QLD and WA/SA groups were the most distant. The laboratory-derived ChiswickAVRS population showed a distinct component location compared with all other populations using both SNP sets. The hierarchical clustering from the PCA coordinates of the two SNPs sets confirmed ChiswickAVRS as the most distant and unique population (supplementary material S5, Supplementary Material online). The pattern of variation seen in the populations using the putative drug resistance gene SNPs formed distinct clusters using the allele frequency of the BTUB1-198 SNP (fig. 4D). The BTUB1-198 frequency variation did not form distinct clusters in populations using the randomly distributed SNPs.

A total of 18 SNPs were homozygotes for the 21 populations (allele frequency 0 or 1) and 13 SNPs were heterozygous for all 21 populations (allele frequency between >0 and <1).

There was no significant difference in heterozygosity between the 21 populations across the 143 SNPs (data not shown).

Populations showed significant differences in Fst (fig. 5A). The pairwise Fst values ranged from 0 to 0.309 for the randomly selected markers and from 0 to 0.331 for the putative drug resistance gene SNPs (supplementary material S6, Supplementary Material online). The NSW and QLD populations grouped with short branch lengths using Fst, whereas the WA and SA populations were separated by longer branches on the tree. The laboratory-derived ChiswickAVRS population was separated by a very long branch, and this divergence was significant. Using the putative drug resistance gene SNPs, the populations were more segregated than for the randomly selected SNPs (fig. 5B). The two laboratory-derived populations were in a clade with the WA and SA populations. The populations with known drug resistance were scattered in different clades. The ChiswickAVRS population was significantly different from most populations (14 populations) using Fst derived from the analysis of the putative drug resistance gene SNPs.

The Shannon diversity index reports a higher diversity if there is a large variation of allele frequencies (from 0 to 1) across all SNPs in the given population. The Shannon allelic diversity ranged from 4.304 to 4.544 for the randomly selected SNPs and from 2.516 to 2.695 for the putative drug resistance gene SNPs (fig. 5C and D; supplementary material S7, Supplementary Material online). The Shannon index clustering mirrored the Fst clustering, but the patterns of significant differences were more pronounced, especially for the randomly distributed SNPs, with groups separated by higher degrees of significance. The ChiswickAVRS (4.304), Cannawigara2006 (4.379), and Harvey2017 (4.381) populations had significantly lower allelic diversity compared with most NSW populations, whereas the NSW populations Riverina2017 and Wongarbon2017 had the largest allelic diversity (supplementary material S7, Supplementary Material online). Using the putative drug resistance gene SNPs, ChiswickAVRS also had significantly lower allelic diversity, followed by Narrikup2017, whereas Cannawigara2006 had the highest diversity with 2.695. The Shannon index, from neutral and resistance SNPs, did not show a clustering pattern mirroring geography.

The comparison of the pairwise Fst values and tree topology between data sets with varying numbers of SNPs indicated that sets of 75 and 100 SNPs were stable and yielded similar results as the complete 113 SNPs panel. However, sets of 10, 25 and 50 SNPs were not stable enough to accurately represent the 21 populations in this study, even when including the top ten informative SNPs from the PCA. The Fst tree from the three replicates of the 75 SNPs and 100 SNPs sets yielded the same topology, apart from the unstable position of population Wongarbon between replicates of both sets and with the topology of the tree from the complete SNPs panel. The 75 SNP and 100 SNP panels showed the highest similarity of the fixation index analysis compared with the
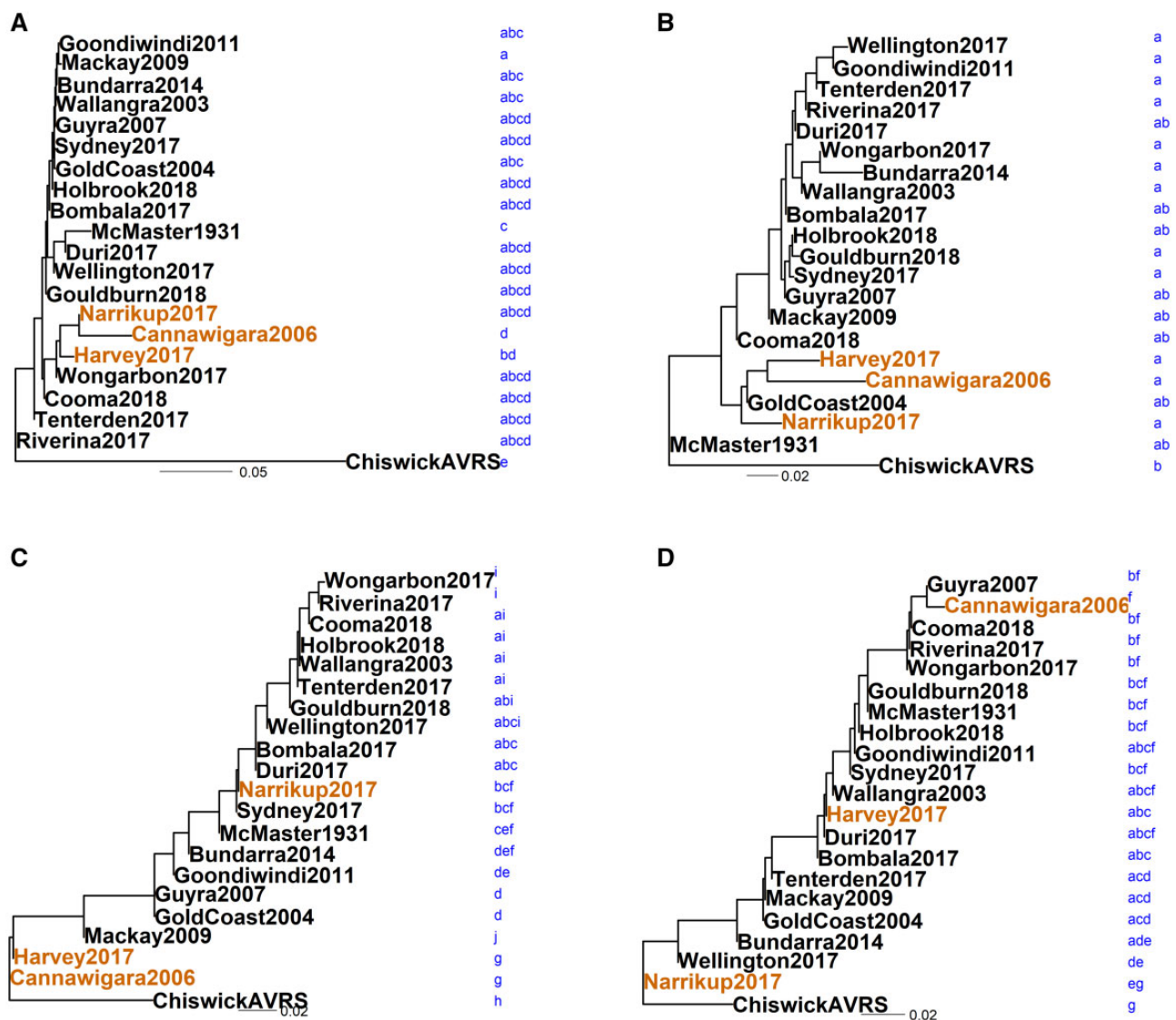
FIG. 5.—Neighbor-joining tree based on the euclidean distance between the 21 *Haemonchus contortus* populations. Pairwise Fst values were used to generate the trees in (*A*) and (*B*), and Shannon's allelic diversity index was used to generate the trees in (*C*) and (*D*). Panels (*A*) and (*C*) are from analyses using the randomly selected SNPs, and panels (*B*) and (*D*) are from analyses using the putative chemical-resistance SNPs. Different letters indicate significant differences between populations ($P<0.05$). The three population from South Australia and Western Australia colored orange.

complete 113 SNPs panel. The difference between the Fst values of the 75 SNPs and 100 SNPs panels was within 0.03 and 0.01 compared with the complete panel. Additionally, the 75 SNPs panel only had five (16%) fewer significant population pair compared with the complete panels, whereas the 100 SNPs panel identified the same populations pairs with significantly different Fst. By contrast, the 10, 25, and 50 SNPs sets yielded unstable Fst tree topologies, and large variations in maximum Fst values (>0.05), and large variations in the number (>8 pairs) and identity of significant pairwise Fst. Noteworthy, the analysis of the ten PCA-informative SNPs yielded a somewhat similar Fst tree topology, but showed the largest differences in mean, maximum, SD, and number

of significant population pairs compared with the complete panel. The above result from this sensitivity test indicates that only 66% of SNPs from the complete panel selected here are necessary to achieve the same level of population monitoring.

Discussion

SNPs Panel Construction

The SNP discovery workflow used in the present study is suitable for situations where the reference genome is of varying quality and/or there are multiple reference genome assemblies. We selected 334 SNPs for evaluation from over 13 million SNPs identified from alignments of sequence data with

two partial genome assemblies. Our SNP selection strategy performed well, with only a low percentage (2%) of non-amplifying markers, whereas 94% of the markers were mapped with high accuracy (i.e., BLAST E-value) to a recent high-quality *H. contortus* genome assembly. The mean read coverage we used (20× each in two sequencing projects) exceeds the 10× coverage considered to be ideal for discovery of SNPs variants from multisample data (Jiang et al. 2019). We did not determine if a lower coverage requirement would have been sufficient. Rather, with a very large number of potential SNP markers to choose from, we decided on a more stringent approach. The earlier *H. contortus* genome assemblies we utilized during marker selection were not equally useful for the task, reflecting findings that the ISE genome better represents *H. contortus* than the McM genome. The design of the SNP panel based on the two earlier genomes, followed by validation on the more complete genome assembly allowed a post hoc evaluation of the original alignment for the genome regions which were the focus of the SNP panel for *H. contortus*. The SNPs that did not map to the recent assembly (6%), but amplified well, could be due to gaps in the genomic assembly, or poor BLAST accuracy against the genome due to structural or sequence variants. The genome of *H. contortus* remains a focus of research, and refinements will continue to be made (Laing et al. 2013; Palevich et al. 2019; Doyle et al. 2020).

Pooled individual samples were selected because 1) field derived samples of *H. contortus* can only be obtained from nematode eggs in feces or larvae from fecal cultures, larvae are preferred because of a lower risk of polymerase chain reaction (PCR)-inhibiting contaminants, 2) *H. contortus* larvae are <1 mm long, therefore time consuming to separate and do not reliably yield DNA, and finally 3) in order to obtain a reliable estimate of population genetics statistics in populations that contain millions of individuals, several hundred to a thousand individuals would be needed. Pooled genotyping has recently been acknowledged to be a cost-effective and reliable estimate of individual allele frequency data for population genetics applications (Futschik and Schlötterer 2010; Boitard et al. 2012; Rellstab et al. 2013; Lynch et al. 2014).

The estimated allele frequency divergence between that predicted from NGS alignment and that predicted from pooled individual Sequenom allelotyping was greater than observed in the one study to our knowledge that has also transferred NGS data to the Sequenom platform (Smith et al. 2018). Smith et al. (2018) reported 36% of SNPs were predicted to be heterozygous from NGS but were homozygous from Sequenom, resulting in only 50% of SNPs being informative after transfer to Sequenom. Our study corroborates this finding, with 40% of potential SNPs homozygous when estimated using pooled allelotyping from Sequenom assays. Interestingly, an additional 6–12% SNPs, in the McMaster1931 and Wallangra2003 populations, respectively, were homozygous for the nucleotide predicted to be the minor allele from NGS. These results show that the estimation of allelic frequency from NGS can differ largely from the estimation by Sequenom (based on pooled base weight). The differences in allelic frequencies between the two methods observed here could have been influenced by different quantitative population sampling as the Illumina sequencing was undertaken with pools of DNA from 120 adult individuals, whereas the Sequenom allelotyping was carried out on DNA from a pool of at least 5,000 individual larvae.

The number of high-quality SNPs identified in this study indicates a high heterogeneity and genetic diversity within populations. This has been extensively reported from previous *H. contortus* genetic studies (Hunt et al. 2008; Gilleard and Redman 2016). Comparison of high-quality SNP density in the *H. contortus* genome with other studies is difficult as previous studies in *H. contortus* have mainly identified SNPs in specific loci linked to chemical resistance and have used related susceptible populations as the reference (Bagnall et al. 2017; Luo et al. 2017). Using highly reduced representation 2 b-RAD sequencing, the study from Khan et al. (2019) identified 86k whole-genome SNPs from various worldwide *H. contortus* populations. The 2 b-RAD sequencing method has been flagged as potentially problematic in highly heterogeneous genomes due to the short reads limiting locus determination (Wang et al. 2012). Given the 11 million SNPs identified in our study, and the 86k SNPs identified by Khan et al. (2019), this would imply that 2 b-RAD sequencing in *H. contortus* can uncover less than 1% of the total number of biallelic SNPs, or that the populations used here are significantly more polymorphic than the one used by Khan et al. (2019). The study by Sallé et al. (2019) discovered 23 million SNPs using a similar sequencing platform as the present study, Illumina HiSeq2500, but alignment was made to the 2017 genome assembly. This could indicate that the lower assembly quality of the ISE and McM genomes used in the present study, compared with the 2018 assembly, results in a considerable underestimation of 12 million SNPs (52%). The high number of shorter, noncontiguous scaffolds in the ISE and McM genomes, compared with the six chromosomes of the 2018 assembly, could contribute to the significant masking of SNPs observed here. This nematode species, with its multiple worldwide genome assemblies of various quality, would be a good model to test SNP discovery robustness compared with reference genome characteristics.

This large variation in the number of SNPs identified could be due to technical or potentially biological issues. As a species, *H. contortus* has a large, complex, and heterogeneous genome evidenced by the observation of much more inter-genome variation than expected between Australian and South African genomic assembly comparison. Future whole-genome SNP discovery studies in *H. contortus* will help obtain a more complete estimate of the level of polymorphism in this species.

Similar SNP discovery and validation pipelines have been published for animals, crops, pests, and humans (Paschou et al. 2007; Van Tassell et al. 2008; Kofler et al. 2011; Ozerov et al. 2013; Mimee et al. 2015; Melo et al. 2016; Torkamaneh et al. 2016; Kleinman-Ruiz et al. 2017; Wright et al. 2019; Catanese et al. 2021). However, some were not designed for pooled sequencing (Paschou et al. 2007; Melo et al. 2016; Torkamaneh et al. 2016), or did not aim to select a subset of representative SNPs (Van Tassell et al. 2008; Kofler et al. 2011; Mimee et al. 2015). A similar approach to ours was used by Kleinman-Ruiz (2017) and Wright (2019), in lynx and Tasmanian devil, which contrary to *H. contortus* have strong population structure and known pedigrees. These two studies also used only NGS and did not transfer results onto another platform. The validation of SNPs predicted from NGS genotyping on the Sequenom platform allows a possible integrated monitoring approach, where combined sets of SNPs can be used for pathogen monitoring (such as *H. contortus*) and livestock genetics, to achieve efficiencies of scale (also using SNP-Chip approach) as well as reduced cost. The monitoring of widespread agricultural pests generally necessitating larger populations at large spatial scale compared with endangered species.

SNPs Validation

The SNP panel selected following the validation steps can be used to accurately estimate allele frequency from pooled DNA and track admixed population variations exclusively in *H. contortus*, even in samples containing DNA from mixed-GIN species. The 20 SNPs that amplified in co-occurring species could provide the basis of community-level assessment for future studies, especially in species where a reference genome is not available. Given the genetic profile of two or more *H. contortus* populations, the SNP panel can identify the levels of admixture (0%, 25%, 50%, or 75%) between populations with an accuracy of ± 0.08. This capacity will enable us to consider detection of quarantine-drench failure or other sources of between-property transfer of parasites. The validation steps performed on the two populations in the present study was not biased toward specific values of alternate allele frequency as shown by the similar range of allele frequency (and allele frequency differences) before and after validation.

The validation step eliminating the highest number of SNPs was the examination of allele frequency from pooled samples compared with the summation of assays conducted using DNA from individual nematodes. In agreement with an earlier study using *Arabidopsis thaliana* (Rellstab et al. 2013), the pool versus individual allele frequency difference in the present study was not significantly different, at below 4%, and showed a strong correlation (R^2). In line with Lynch et al. (2014), the lowest allele frequency estimation of pooled sample was close to the $5/N$ threshold with an expected value of 0.05, with the observed value being 0.06 in our study. The

mean read coverage of our pooled sample SNP estimation was $40\times$, a value which was identified by Kofler et al. (2011) as the threshold to estimate allele frequency for population genetic application from pooled samples, but lower than the value of $50\times$ identified by Schlötterer et al. (2014).

We attempted to identify factors which might predict failure of quality control for the SNP which we analyzed in this study. None of the factors evaluated, such as sequencing coverage, loci duplication in either reference genome, frequency uncertainty values, low difference between alternate allele frequency of Wallangra2003 and McMaster1931, scaffold bias, homozygote to heterozygote ratio, or transversion versus transition (data not shown) could successfully predict which SNPs failed or passed the different validation steps. This demonstrates the requirement for laboratory-based quality control steps rather than relying solely upon currently available bioinformatics analysis methods.

Australian Populations Genetic Diversity

The PCA and Fst methods have proven successful for assessing population genetics following whole-genome SNP discovery (Paschou et al. 2007; Mimee et al. 2015; Ballesteros et al. 2018; Wright et al. 2019). Differentiation between the field and lab-derived populations, and between the various field populations was achieved in the present study. Similar to previous results, this analysis indicated the presence of low levels of population structure and high intrapopulation genetic diversity in Australian *H. contortus* (Hunt et al. 2008) compared with comparisons between populations from different continents (Yin et al. 2013, 2016; Dey et al. 2019; Khan et al. 2019; Portanier et al. 2019). The SNPs that did not reliably amplify in three or more of the 21 populations are possibly further evidence of the high-density of polymorphisms in *H. contortus*, as the amplification of SNPs of interest relies on the absence of other polymorphisms within a 150-bp genomic region. The randomly distributed SNPs overwhelmingly (80%) annotated to noncoding regions of the genome compared with the chemical-resistance SNPs set. The randomly selected SNPs have the potential to be selectively neutral, and adequately represent population genetic shifts due to a combination of management strategies. Each SNP evaluated can be considered independent as there was no evidence of pairwise SNP linkage across the 21 populations. We examined the data to investigate if the SNP panel has the flexibility to be reduced to fit technical or financial constraints. As long as a minimum of 75 random SNPs is selected from the neutral panel constructed here, the population genetic structure based on Fst results was stable, this was the best estimate of the minimum useful SNP density we could deduce using our methods. If a much greater number of SNP markers were developed into high throughput tests a full-scale power analysis might be possible, refining the estimate of minimum useful SNP density. SNP selection based on PCA contribution to

subsequently assess population genetics has been shown to be a successful and tested method in humans (Paschou et al. 2007). This method does not hold in the nematode *H. contortus* based on our results; the most informative SNPs from the PCA on their own did not accurately represent the relationships between populations in *H. contortus*.

The three population genetics metrics we calculated from the allelotyping data reveal no genetic structure based on geographic location or known chemical resistance for the randomly distributed SNPs. The putative chemical-resistance SNP set also shows no genetic structure based on geographic origin but does separate the 21 populations by categories of BTUB1-198 frequency. The PCA and Fst analysis show the three western populations (SA and WA) together but also include eastern (NSW) populations within the grouping. This indicates that there is no large-scale geographic pattern in the Australian *H. contortus* populations. the pattern of Australian populations based solely on the BTUB1-198 mutation (E198A) is highly similar to the pattern shown by the chemical-resistant SNP panel, which is not the case for the BTUB1-200 (F200Y) mutation. The importance of the E198A mutation compared with the F200Y supports earlier analyses of benzimidazole resistance in Australian isolates in which the presence of the E198A mutation conferred a higher level of resistance (Kotze et al. 2012). The BTUB1-198 pattern holds even though some of the populations are resistant to drugs other than the benzimidazoles. This could indicate the widespread baseline resistance of field populations in Australia to the benzimidazoles due to long standing use of this chemical group (Emery et al. 2016). As expected, the inbred laboratory stain ChiswickAVRS was easily differentiated from the 20 other populations. Compared with the randomly distributed SNPs, the chemical-resistance SNPs show larger Fst values accompanied with lower differences in allelic diversity, which is to be expected from selective polymorphisms as also observed previously (Luo et al. 2017).

Conclusion

The two sets of SNPs developed here—randomly selected and putative chemical-resistance—provide a cost-effective, flexible, and rapid monitoring tool for population characterization and potentially detecting genetic change over time for an important parasite of livestock. This tool can be used to identify new population incursions, varying levels of admixed population, assess quarantine-drench efficacy, and monitor changes in populations due to management interventions. The strategic use of such genetic tools could help slow the evolution of drug-resistant populations. The workflow developed here is flexible enough to be applied to other parasites and pest population control. The approach of SNP selection presented here will be applicable to other species with high genetic diversity and uncertainty in genomic assembly.

Materials and Methods

Parasite Culture

To generate parasite material for use in this work, sheep were housed indoors on slatted floors so that no transmission between animals via pasture was possible and were given heat treated feed rations which do not contain viable parasite material. Before infection, sheep were treated with anthelmintic drugs (naphthalophos, monepantel, levamisole, albendazole, abamectin, triclabendazole, and oxfendazole) according to the manufacturer's dosing instructions and using the individual animal's liveweight. The doses were administered over two days to avoid possible antagonisms between products. Levamisole, albendazole, and abamectin were provided using a combination product, as was the triclabendazole and oxfendazole. Adult parasite material was harvested postmortem, and larvae were cultured from collected feces according to standard procedures. Animal ethics was approved for these experiments under New South Wales (Australia) legislation by the CSIRO FD McMaster Laboratory Animal Ethics Committee under animal research authorities 10/11, 12/02, 13/23, 14/10, 15/07, 16/14, 17/12, and 18/09.

SNP Discovery

DNA was extracted from five and six pools of 20 sexually mature adult female *H. contortus* worms from the Wallangra2003 and McMaster1931 isolates, respectively. As *H. contortus* is polyandrous, each pool would be expected to comprise DNA from more than 20 individuals due to the likely presence of DNA from the female parent, the fertilized eggs in utero, and stored spermatozoa within the spermatheca of each individual. These populations were selected as genetically distant Australian populations (Hunt et al. 2008). A total of 11 paired-end 100-bp fragment libraries were sequenced using MiSeq technology (Illumina Australia, VIC) and the data are available through NCBI (BioProject ID PRJNA784704). The 11 individual libraries were aligned to the two previously available *H. contortus* reference genome assemblies (Laing et al. 2013; Schwarz et al. 2013). SNP were chosen from these alignments and analyzed according to the variant discovery pipeline (fig. 6). The SNP panel was subsequently screened against the improved 2017 genome assembly for *H. contortus* (Doyle et al. 2017) in order to assess SNP validity, position, and functionality.

Following the alignment of individual libraries to the two reference genomes, reads were filtered based on the following criteria; 1) high alignment quality (MAPQ > 60), 2) both reads in a pair were aligned (samtools v1.3.1, Li et al. 2009), and 3) optical and PCR read duplicates identified and removed (Picard v2.9.2, http://broadinstitute.github.io/picard, last accessed February 21, 2022). The variant discovery consisted of two steps; 1) variants (SNPs and Indels) were identified in each of the 11 samples individually (intrapopulation), and 2)
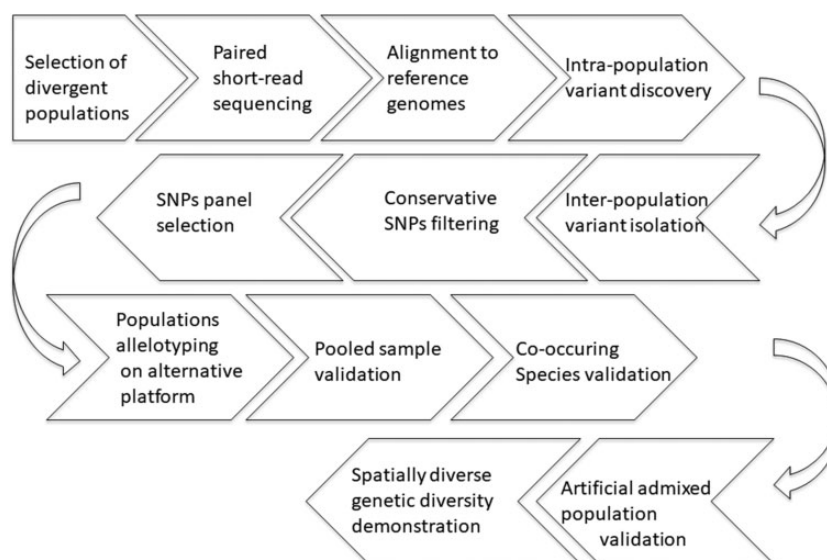
Fig. 6.—SNP discovery, selection, and validation workflow.

the two populations were genotyped (interpopulation) (fig. 6). The variant discovery steps were performed using GATK (Van der Auwera et al. 2013).

Variant filtering was used to remove Indels and SNPs which were not biallelic. Remaining SNPs were removed if they had a low quality by coverage score (QD < 5), were found only on the same read orientation (FisherStrand >60), had differing variant call quality between the reference and alternate alleles (MQRankSumtest < −5), had variants that were found only within the extremities of reads (ReadPosRankSum < −8), and finally if variants were found more commonly on one strand than the other (StrandOddRatio > 3). The conservative filtering followed the best practice guideline of DePristo et al. (2011).

From the filtered SNPs a panel was selected. First, an equal number of SNP from alignments to the two draft genomes were selected, including 244 randomly distributed SNPs, and 90 SNPs specifically located within or near genes implicated in drug resistance. The 244 randomly distributed SNPs were selected based upon the following criteria; 1) equal number of transition and transversion substitutions, 2) located on the largest 50% of all genomic scaffolds within each of the two draft genomes, 3) representing the range of possible allele frequency differences between populations (0–1), and 4) across the range of possible minor allele frequencies within populations (0–0.5). The selection process was designed to find a set of markers that would be likely to display genetic variation between populations, be unlikely to have gene specific associations with drug selection and, when analyzed as a set, would reflect overall genetic changes in the population under study, or reflect differences between populations. A further 90 SNPs were selected specifically from the scientific literature to enable comparisons in genes implicated in drug

resistance in *H. contortus*. Protein sequences from genes with known and putative effects on drug susceptibility in *H. contortus* (supplementary material S1, Supplementary Material online) were retrieved from NCBI and screened against the list of high-quality SNPs discovered. These included known SNPs which are putatively involved in drug resistance and were from the isotype 1 of the β-tubulin gene *βtub1* (benzimidazole resistance), monepantel target gene *acr* (monepantel resistance), the glutamate gated chloride ion channel-3, and other glutamate gated chloride ion channel receptors (macrocyclic lactone resistance), the amphid dendrite dye-filling deficient gene *dyf-7* (macrocyclic lactone resistance), and nicotine acetylcholine receptors (levamisole resistance). Further, genes putatively involved in drug resistance such as cytochrome P450s, P-glycoproteins, UDP-glucuronosyltransferases, ATP-binding cassette transporters, and haf-transporters were also screened for presence of high-quality SNPs (Neveu et al. 2010; Williamson et al. 2011; Martin et al. 2012; Janssen et al. 2013; Kotze et al. 2014; Romine et al. 2014).

Allelotyping Reaction

The *H. contortus* SNP panel was developed for use with mass spectrometry allelotyping technology. The 334 SNPs were multiplexed to be allelotyped using the Sequenom platform (Gabriel et al. 2009). An average of 25–40 SNPs was included in each reaction. PCR primers and extension primers were designed for each SNP. The markers were multiplexed in as few sets as possible depending on the mass of the allele sequenced, ensuring each marker's amplicon size did not overlap on the chromatogram. Each SNP marker was analyzed after primer extension using the Agena Bioscience

Sequenom MassArray system in triplicate (Neogen Australasia [formerly UQ Animal Genetics Laboratory], Gatton, QLD).

SNP Panel Validation

The validation for each SNP involved three steps determining if, 1) the pooled sample allele frequency estimation was similar to frequency estimates from analysis of multiple separate individuals from the same population, 2) the PCR amplification was successful for *H. contortus* but not co-occurring nematode species, and 3) the SNP could be used to identify artificially admixed populations (fig. 6).

Pooled versus Individual Sample Allelotyping

DNA was extracted from 30 individual adult males from the McMaster1931 and the Wallangra2003 isolates using the DirectPCR Lysis (MouseTail) buffer (Viagen Biotech, CA) following the manufacturer's protocol. DNA was quantified using spectrophotometry (NanoDrop, Thermo Fisher Scientific) and the 30 individuals per population were pooled in equimolar concentration to create three samples: MCM_100%, WAL_100%, and a pool of 50% MCM and 50% WAL. Each of the 60 individuals and three pooled samples were allelotyped for each SNP marker in triplicate.

Detection of Artificially Admixed Populations

Artificially admixed populations were created using three population pairs: 1) McMaster1931 and Wallangra2003, 2) Mackay2009 and Cannawigara2006, and 3) Goondiwindi2011 and GoldCoast2004 (supplementary material S2, Supplementary Material online). The WAL/MCM pair was selected as it was initially used in genomic discovery of the SNPs, the Mackay/Cannawigara pair were from the most distant in North to South axis, and the Goondiwindi/GoldCoast pair were from the most divergent coast/inland distance. Pooled samples from each population were mixed at different ratios in a pairwise fashion: 0:100, 25:75, 50:50, 75:25, and 100:0. The alternate (e.g., nonreference) allele frequencies from these admixed populations were evaluated in two ways. First, the observed gradient in allele frequency across the samples was determined and compared with that expected based on analysis of the individual populations. Second, the alternate allele frequency observed for the 50:50 population was compared with the observed alternate allele frequency mean of the 0:100 and 100:0 populations.

Co-occurring GIN Species

This validation step was designed to ensure the *H. contortus*-specificity of the SNP panel so that the test would be useful to analyze samples from coinfected animals. DNA of adult individuals from *T. colubriformis*, *Tel. circumcincta*, *C. ovina*, *Coo. oncophora*, and *O. columbianum* was extracted from

samples collected in 2012, 2007, 2007, 2013, and 2007, respectively. Samples from the species *O. venulosum* were acquired from the South Australian Museum (original collection in 1981). Although not available as a single species sample, a 20:80 sample of *H. contortus*: *T. vitrinus* was used (collection in 2007) to establish SNPs affected by this level of *T. vitrinus* inclusion in *H. contortus* samples. Sheep (*Ovis aries*) DNA was also used to ensure no cross-reaction with host genetic material. Except for *O. venulosum*, all samples came from the CSIRO F.D. McMaster laboratory parasite collection (Armidale, NSW, Australia).

Using the SNP Panel to Examine Spatially Diverse *H. contortus* Populations from Australia

A total of 21 *H. contortus* populations collected across Australia were used (supplementary material S2, Supplementary Material online). Ten populations from commercial properties were sampled in 2017/18; populations including those from Holbrook2018, Goulbourn2018, Sydney2017, Tenterden2017, Duri2017, Riverina2017, Bombala2017, Cooma2018, Wongarbon2017, and Wellington2017. Eleven historical samples came from the CSIRO F.D. McMaster laboratory parasite collection (supplementary material S2, Supplementary Material online). McMaster1931 and ChiswickAVRS are considered laboratory populations; they originate from field collections that are older than 20 years and have been subjected to strong genetic bottlenecks due to phenotype selection (ChiswickAVRS) or random selection of low numbers of individuals throughout the years (McMaster1931). The population Wallangra2003 is resistant to the benzimidazole, macrocyclic lactone, imidazothiazole and salicylanilide drug classes, GoldCoast2004, and ChiswickAVRS are resistant to macrocyclic lactones, and Cannawigara2006 is resistant to benzimidazoles (detailed in Hunt et al. [2008]). The drug resistance phenotype is unknown for the other populations. The BTUB1-198 mutation is shown as it is one of the well-known mutations conferring high level benzimidazole resistance in *H. contortus*.

DNA was extracted from a pool of larvae (2017/2018 populations) or adult individuals and population-level SNP allele frequencies were obtained for each population using the extracted DNA. The mean frequency of the nonreference alternate allele of each SNP was calculated along with the respective allele frequency uncertainty from the Sequenom allelotyping platform. The mean frequency uncertainty from allelotyping was added to any frequency difference between individual and pool to obtain an overall frequency uncertainty value across all SNPs. These overall frequency uncertainty values were used to create a distribution of potential alternate allele frequencies. To evaluate the extent of linkage groups of the SNPs selected from pooled population amplification, the correlation between pairwise SNPs distance on each of the

chromosome with the variance across the 21 population's alternate allele frequency was calculated.

PCA, pairwise fixation index (Fst) estimation, and allelic frequency diversity were used to investigate the relationships between populations based on genetic diversity. The PCA was performed in R using the *dudi.pca* function using the mean alternate allele frequency of each population for each SNP (R v.3.4.0, R Development Core Team 2019). The Fst between population pair was calculated using the alternate allele frequency distribution input (described above) using the R function *stamppFst* (Pembleton et al. 2013). Pairwise Fst was visualized using a neighbor-joining tree using the Fst distance between each population pair, as a distance matrix. Allelic diversity was calculated using the Shannon H index (Shannon and Weaver 1949) and visualized using a neighbor-joining tree using the euclidean distance between each population pair as a distance matrix.

To assess the stability of the population genetics analysis used to monitor populations, Fst analysis was performed on three replicates of four sizes of randomly selected SNP panel: 25, 50, 75, and 100 SNPs. Noninformative SNPs were removed beforehand. The contribution of the top ten most informative SNPs as identified by the PCA to the Fst analysis were also used as a SNP set on their own.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Data Availability

The DNA sequence data underlying this article are available in the NCBI Bioproject database at https://www.ncbi.nlm.nih.gov/bioproject/, and can be accessed with the project number PRJNA784704. The primer sequences underlying these allelo-typing reactions data in this article are the shared property of Meat and Livestock Australia and can only be shared by permission. Data will be shared on request to the corresponding author with permission of Meat and Livestock Australia.

Literature Cited

Bagnall NH, et al. 2017. Mutations in the Hco-mptl-1 gene in a field-derived monepantel-resistant isolate of *Haemonchus contortus*. Int J Parasitol. 7(2):236–240.

Bailly X, Migeon A, Navajas M. 2004. Analysis of microsatellite variation in the spider mite pest *Tetranychus turkestani* (Acari: Tetranychidae) reveals population genetic structure and raises questions about related ecological factors. Biol J Linn Soc. 10:69–78.

Ballesteros C, et al. 2018. Clinical validation of molecular markers of macrocyclic lactone resistance in *Dirofilaria immitis*. Int J Parasitol Drugs Drug Resist. 8(3):596–606.

Besier RB, Kahn LP, Sargison ND, Van Wyk JA. 2016. Diagnosis, treatment and management of *Haemonchus contortus* in small ruminants. In: Gasser RB, Samson-Himmelstjerna GV, editors. Advances in parasitology. Vol. 93. San Diego: Elsevier. p. 181–238.

Blouin MS. 2002. Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. Int J Parasitol. 32(5):527–531.

Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A. 2012. Detecting selective sweeps from pooled next-generation sequencing samples. Mol Biol Evol. 29(9):2177–2186.

Cabaret J. 2008. Pro and cons of targeted selective treatment against digestive-tract strongyles of ruminants. Parasite 15(3):506–509.

Catanese G, Trotta JM, Iriondo M, Grau A, Estonba A. 2021. Discovery of SNP markers of red shrimp *Aristeus antennatus* for population structure in Western Mediterranean Sea. Conserv Genet Resourc. 13:21–25.

Cerutti MC, et al. 2010. Genetic variability of *Haemonchus contortus* (Nematoda: Trichostrongyloidea) in alpine ruminant host species. J Helminthol. 84(3):276–283.

Chaudhry U, et al. 2016. Microsatellite marker analysis of *Haemonchus contortus* populations from Pakistan suggests that frequent benzimidazole drug treatment does not result in a reduction of overall genetic diversity. Parasit Vectors. 9(1):349.

Cole R, Viney M. 2018. The population genetics of parasitic nematodes of wild animals. Parasit Vectors. 11(1):590.

Coles GC, et al. 1992. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. Vet Parasitol. 44(1–2):35–44.

Cowled BD, et al. 2012. Integrating survey and molecular approaches to better understand wildlife disease ecology. PLoS One 7(10):e46310.

DePristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5):491–498.

Dey AR, et al. 2019. Genetic diversity patterns of *Haemonchus contortus* isolated from sheep and goats in Bangladesh. Infect Genet Evol. 68:177–184.

Downes K, et al. 2004. SNP allele frequency estimation in DNA pools and variance components analysis. Biotechniques 36(5):840–845.

Doyle SR, et al. 2017. A genome resequencing-based genetic map reveals the recombination landscape of an outbred parasitic nematode in the presence of polyploidy and polyandry. Genome Biol Evol. 10(2):396–409.

Doyle SR, et al. 2020. Extensive genomic and transcriptomic variation defines the chromosome-scale assembly of *Haemonchus contortus*, a model gastrointestinal worm. Commun Biol. 3(1):656.

Else KJ, et al. 2020. Whipworm and roundworm infections. Nat Rev Dis Primers. 6(1):44.

Emery DL, Hunt PW, Le Jambre LF. 2016. *Haemonchus contortus*: the then and now, and where to from here? Int J Parasitol. 46(12):755–769.

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics 186(1):207–218.

Gabriel S, Ziaugra L, Tabbaa D. 2009. SNP genotyping using the sequenom MassARRAY iPLEX platform. In: Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Smith DR, editors. Current protocols in human genetics. Hoboken (NJ): John Wiley & Sons, Inc.

Gharamah AA, Azizah MNS, Rahman WA. 2012. Genetic variation of *Haemonchus contortus* (Trichostrongylidae) in sheep and goats from Malaysia and Yemen. Vet Parasitol. 188(3–4):268–276.

Gilleard JS, Beech RN. 2007. Population genetics of anthelmintic resistance in parasitic nematodes. Parasitology 134(Pt 8):1133–1147.

Gilleard JS, Redman E. 2016. Genetic diversity and population structure of *Haemonchus contortus*. In: Gasser RB, Samson-Himmelstjerna GV, editors. Advances in parasitology. Vol. 93. San Diego: Elsevier. p. 31–68.

Gilleard JS. 2013. *Haemonchus contortus* as a paradigm and model to study anthelmintic drug resistance. Parasitology 140(12):1506–1522.

Gould F. 1995. Comparisons between resistance management strategies for insects and weeds. Weed Technol. 9(4):830–839.

Gould F, Brown ZS, Kuzma J. 2018. Wicked evolution: can we address the sociobiological dilemma of pesticide resistance? Science 360(6390):728–732.

Hellicar AD, Rahman A, Smith DV, Henshall JM. 2015. Machine learning approach for pooled DNA sample calibration. BMC Bioinformatics 16(1):214.

Hoberg EP, Zarlenga DS. 2016. Evolution and biogeography of *Haemonchus contortus*: linking faunal dynamics in space and time. In: Gasser RB, Samson-Himmelstjerna GV, editors. Advances in parasitology. Vol. 93. San Diego: Elsevier. p. 1–30.

Hunt PW, Knox MR, Le Jambre LF, McNally J, Anderson LJ. 2008. Genetic and phenotypic differences between isolates of *Haemonchus contortus* in Australia. Int J Parasitol. 38(8–9):885–900.

Hunt PW, Lello J. 2012. How to make DNA count: DNA-based diagnostic tools in veterinary parasitology. Vet Parasitol. 186(1–2):101–108.

Janssen IJI, Krücken J, Demeler J, von Samson-Himmelstjerna G. 2013. *Caenorhabditis elegans*: modest increase of susceptibility to ivermectin in individual P-glycoprotein loss-of-function strains. Exp Parasitol. 134(2):171–177.

Jiang Y, Jiang Y, Wang S, Zhang Q, Ding X. 2019. Optimal sequencing depth design for whole genome re-sequencing in pigs. BMC Bioinformatics 20(1):556.

Khan S, et al. 2019. Analysis of genome-wide SNPs based on 2b-RAD sequencing of pooled samples reveals signature of selection in different populations of *Haemonchus contortus*. J Biosci. 44(4):97.

Kirk H, Dorn S, Mazzi D. 2013. Molecular genetics and genomics generate new insights into invertebrate pest invasions. Evol Appl. 6(5):842–856.

Kleinman-Ruiz D, et al. 2017. Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. BMC Genomics. 18(1):556.

Kofler R, et al. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6(1):e15925.

Kotze AC, et al. 2012. Relative level of thiabendazole resistance associated with the E198A and F200Y SNPs in larvae of a multi-drug resistant isolate of *Haemonchus contortus*. Int J Parasitol Drugs Drug Resist. 2:92–97.

Kotze AC, et al. 2014. Recent advances in candidate-gene and whole-genome approaches to the discovery of anthelmintic resistance markers and the description of drug/receptor interactions. Int J Parasitol Drugs Drug Resist. 4(3):164–184.

Kotze AC, Gilleard JS, Doyle SR, Prichard RK. 2020. Challenges and opportunities for the adoption of molecular diagnostics for anthelmintic resistance. Int J Parasitol Drugs Drug Resist. 14:264–273.

Kotze AC, Prichard RK. 2016. Anthelmintic resistance in *Haemonchus contortus*. In: Gasser RB, Samson-Himmelstjerna GV, editors. Advances in parasitology. Vol. 93. San Diego: Elsevier. p. 397–428.

Kotze AC, Ruffell AP, Ingham AB. 2014. Phenobarbital induction and chemical synergism demonstrate the role of UDP-glucuronosyltransferases in detoxification of naphthalophos by *Haemonchus contortus* larvae. Antimicrob Agents Chemother. 58(12):7475–7483.

Laing R, et al. 2013. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. Genome Biol. 14(8):R88.

Love S. 2010. Worms – what's new: some recent developments in worm control in ruminants and camelids. Australasian Farm Business Management Network 7(2):1–8.

Luo X, et al. 2017. Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in *Haemonchus contortus*. Parasit Vectors. 10(1):31.

Lynch M, Bost D, Wilson S, Maruki T, Harrison S. 2014. Population-genetic inference from pooled-sequencing data. Genome Biol Evol. 6(5):1210–1218.

Martin RJ, Buxton SK, Neveu C, Charvet CL, Robertson AP. 2012. Emodepside and SL0-1 potassium channels: a review. Exp Parasitol. 132(1):40–46.

Melo ATO, Bartaula R, Hale I. 2016. GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics 17(1):29.

Mimee B, et al. 2015. A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. Mol Ecol Resour. 15(6):1356–1365.

Neveu C, et al. 2010. Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance. Pharmacogenet Genomics. 20(7):414–425.

Otsen M, et al. 2001. Amplified fragment length polymorphism analysis of genetic diversity of Haemonchus contortus during selection for drug resistance. Int J Parasitol. 31(10):1138–1143.

Ozerov M, et al. 2013. Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. PLoS One 8(12):e82434.

Palevich N, Maclean PH, Baten A, Scott RW, Leathwick DM. 2019. The genome sequence of the anthelmintic-susceptible New Zealand *Haemonchus contortus*. Genome Biol Evol. 11(7):1965–1970.

Paschou P, et al. 2007. PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genet. 3(9):e160.

Phillips MA, et al. 2017. Malaria. Nat Rev Dis Primers. 3(1):17050.

Porretta D, Canestrelli D, Bellini R, Celli G, Urbanelli S. 2007. Improving insect pest management through population genetic data: a case study of the mosquito *Ochlerotatus caspius* (Pallas): pest management through population genetic data. J Appl Ecol. 44(3):682–691.

Portanier E, et al. 2019. Does host socio-spatial behavior lead to a fine-scale spatial genetic structure in its associated parasites? Parasite 26:64.

R Development Core Team. 2019. R: a language and environment for statistical computing [Internet]. Vienna (Austria): R Foundation for Statistical Computing.

Redman E, et al. 2015. The emergence of resistance to the benzimidazole anthelmintics in parasitic nematodes of livestock is characterised by multiple independent hard and soft selective sweeps. PLoS Negl Trop Dis. 9(2):e0003494.

Redman E, Grillo V, et al. 2008. Genetics of mating and sex determination in the parasitic nematode *Haemonchus contortus*. Genetics 180(4):1877–1887.

Redman E, Packard E, et al. 2008. Microsatellite analysis reveals marked genetic differentiation between *Haemonchus contortus* laboratory isolates and provides a rapid system of genetic fingerprinting. Int J Parasitol. 38(1):111–122.

Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC. 2013. Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. PLoS One 8(11):e80422.

Roeber F, Jex AR, Gasser RB. 2013. Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for

exploring epidemiology and drug resistance – an Australian perspective. Parasit Vectors. 6(1):153.

Rollins LA, Woolnough AP, Sherwin WB. 2006. Population genetic tools for pest management: a review. Wildl Res. 33(4):251.

Romine NM, Martin RJ, Beetham JK. 2014. Transcriptomic evaluation of the nicotinic acetylcholine receptor pathway in levamisole-resistant and -sensitive *Oesophagostomum dentatum*. Mol Biochem Parasitol. 193(1):66–70.

Rose H, et al. 2016. Climate-driven changes to the spatio-temporal distribution of the parasitic nematode, *Haemonchus contortus*, in sheep in Europe. Glob Change Biol. 22(3):1271–1285.

Roush RT, Daly JC. 1990. The role of population genetics in resistance research and management. In: Roush RT, Tabashnik BE, editors. Pesticide resistance in arthropods. Boston: Springer.

Sallé G, et al. 2019. The global diversity of *Haemonchus contortus* is shaped by human intervention and climate. Nat Commun. 10(1):4811.

Schlötterer C, Tobler C, Kofler R, Nolte V. 2014. Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. Nat Rev Genet. 15(11):749–763.

Schwarz EM, et al. 2013. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. Genome Biol. 14(8):R89.

Shannon CE, Weaver W. 1949. The mathematical theory of communication. Urbana (IL): University of Illinois Press.

Smith HM, et al. 2018. SNP markers tightly linked to root knot nematode resistance in grapevine (*Vitis cinerea*) identified by a genotyping-by-sequencing approach followed by Sequenom MassARRAY validation. PLoS One 13(2):e0193121.

Sréter T, Molnár V, Kassai T. 1994. The distribution of nematode egg counts and larval counts in grazing sheep and their implications for parasite control. Int J Parasitol. 24(1):103–108.

Tomasetto F, Tylianakis JM, Reale M, Wratten S, Goldson SL. 2017. Intensified agriculture favors evolved resistance to biological control. Proc Natl Acad Sci U S A. 114(15):3885–3890.

Torkamaneh D, Laroche J, Belzile F. 2016. Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. PLoS One 11(8):e0161333.

Troell K, Engström A, Morrison DA, Mattsson JG, Höglund J. 2006. Global patterns reveal strong population structure in Haemonchus contortus, a nematode parasite of domesticated ruminants. Int J Parasitol. 36(12):1305–1316.

Van der Auwera GA, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline: the genome analysis toolkit best practices pipeline. In: Bateman A, Pearson WR, Stein LD, Stormo GD, Yates JR, editors. Current protocols in bioinformatics. Hoboken (NJ): John Wiley & Sons, Inc. p. 11.10.1–11.10.33.

Van Tassell CP, et al. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods. 5(3):247–252.

Van Wyk JA, Mayhew E. 2013. Morphological identification of parasitic nematode infective larvae of small ruminants and cattle: a practical lab guide. Onderstepoort J Vet Res. 80(1):539.

Vercruysse J, et al. 2018. Control of Helminth ruminant infections by 2030. Parasitology 145(13):1655–1664.

Waller P. 1997. Nematode parasite control in livestock of tropic/subtropic: the need for novel approach. Int J Parasitol. 27(10):1193–1201.

Wang S, Meyer E, McKay JK, Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Methods. 9(8):808–810.

Williamson SM, et al. 2011. Candidate anthelmintic resistance-associated gene expression and sequence polymorphisms in a triple-resistant field isolate of *Haemonchus contortus*. Mol Biochem Parasitol. 180(2):99–105.

Wolstenholme AJ, Fairweather I, Prichard R, von Samson-Himmelstjerna G, Sangster NC. 2004. Drug resistance in veterinary Helminths. Trends Parasitol. 20(10):469–476.

Wright B, et al. 2019. From reference genomes to population genomics: comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. BMC Genomics 20(1):453.

Yin F, et al. 2013. Genetic variability within and among *Haemonchus contortus* isolates from goats and sheep in China. Parasit Vectors. 6(1):279.

Yin F, et al. 2016. Population structure of *Haemonchus contortus* from seven geographical regions in China, determined on the basis of microsatellite markers. Parasit Vectors. 9(1):586.

Zheng Y, Qiao X, Wang K, Dorn S, Chen M. 2015. Population genetics affected by pest management using fruit-bagging: a case study with *Grapholita molesta* in China. Entomol Exp Appl. 156(2):117–127.

Associate editor: Maria Costantini