

Waste not, want not: microsatellites remain an economical and informative technology for conservation genetics

Samantha Hauser¹, Giridhar Athrey², and Paul Leberg³

¹University of Wisconsin-Milwaukee College of Letters and Science

²Texas A&M University Department of Entomology

³University of Louisiana at Lafayette

January 27, 2021

Abstract

Comparisons of microsatellite and single-nucleotide polymorphisms (SNPs) have found that SNPs outperform microsatellites in population genetic analyses, calling into the question the continued utility of microsatellites in population and landscape genetics. Yet highly polymorphic markers may be of value in species that have reduced genetic variation. This study repeated analyses previously done using microsatellites with SNPs developed from ddRAD sequencing in the black-capped vireo source-sink system. SNPs provided greater resolution of genetic diversity, population differentiation, and migrant detection but could not reconstruct parentage relationships due to insufficient heterozygosities. The biological inferences made by both sets of markers were similar: asymmetrical gene flow from source populations to the remaining sink populations. With the landscape genetic analyses, we found different results between the two molecular markers, but associations of the top environmental features (riparian, open habitat, agriculture, and human development) with dispersal estimates were shared between marker types. Despite the higher precision of SNPs, we find that microsatellites effectively uncover population processes and patterns and are superior for parentage analyses in this species with reduced genetic diversity. This study illustrates the continued applicability and relevance of microsatellites in population genetic research.

Introduction

Molecular markers allow us to answer an array of population genetic questions about gene flow (Edelaar & Bolnick, 2012; Hudson, 1992), parentage (García et al., 2002), and population structuring (Cockerham & Weir, 1993; Narum et al., 2008). The toolbox of molecular markers has rapidly changed with advancing technologies in the last 30 years, progressing from mitochondrial markers to microsatellites to SNP (Single Nucleotide Polymorphisms) markers, each with progressively higher statistical power (Andrews & Luikart, 2014; Luikart et al., 2003; Morin et al., 2004). With this higher resolution of microsatellites and SNPs, their usage has extended to landscape genetics (Sork et al., 2010), adaptation and selection (Ahrens et al., 2018), hybridization (Toews et al., 2016), outbreeding and inbreeding depression (Steiner et al., 2013), and epigenetics (Harrisson et al., 2014). Of particular interest is the relatively recent interdisciplinary field of landscape genetics that combines the theory and methods from landscape ecology and population genetics to study how landscape (or seascape) features affect population processes such as gene flow (Zeller et al., 2012). Landscape genetics bridges the gap between environmental factors and species' responses, providing important insight into ecological and evolutionary processes. Information gleaned from molecular markers offers crucial insights for application in conservation and management efforts.

Microsatellites and SNPs are the most commonly used markers for population genetic studies, each with its pros and cons (Morin et al., 2009). Microsatellites are highly polymorphic, providing relatively high statistical power per locus, but suffer from null alleles, homoplasy, and complex and variable mutation

processes that can confound results (Defaveri et al., 2013; Putman & Carbone, 2014). The distribution of microsatellite markers genome-wide is also unknown across many species. Still, they are likely not distributed evenly across the genome, potentially yielding a poorer representation of overall genetic variation than SNPs (Narum et al., 2013). While microsatellites were once the most used marker in population genetics, SNPs are quickly replacing other molecular markers in ecological, evolutionary, and conservation studies (Baruch & Weller, 2008). SNPs are biallelic and thus have a simple mutation model but have less information per locus, requiring more loci than that for microsatellites to get equal statistical power (Helyar et al., 2011). SNPs also have lower error rates than microsatellites, and with next-generation sequencing, have lower genotyping costs per marker (Morin et al., 2009; Weinman, Solomon, & Rubenstein, 2015). SNPs occur across the genome, giving a better representation of genome-wide variation (Puckett & Eggert, 2016). Comparative assessments of the two markers found that SNPs outperform microsatellites with estimates of genetic diversity and population structure (Morin et al., 2009; Muñoz et al., 2017), and perform equally or poorer with parentage analyses (Flanagan & Jones, 2019; Thrasher et al., 2018; Weinman et al., 2015). Nevertheless, microsatellites are still useful in landscape genetics, and can yield comparable results (Liu et al., 2005; Väli et al., 2008). While many microsatellite and SNP comparisons exist for population genetic analyses, few papers (Hall & Beissinger, 2014; Puckett & Eggert, 2016) have compared them in a landscape genetics context. Furthermore, microsatellites have been demonstrated to perform well relative to older codominant markers in cases following bottlenecks (Spencer et al., 2000), but it is not clear if this advantage of these highly polymorphic loci persist when large numbers of SNP loci can be assessed (Morin et al., 2012; Zimmerman et al., 2020). We need such direct comparisons for understanding the biases, strengths, and weaknesses associated with different marker types both for accurate interpretation of data and to inform the adoption of new marker types in long-term studies.

In this study, we repeated microsatellite-based population and landscape genetic analyses (Hauser & Leberg, 2020; Hauser et al., 2019) with SNPs developed from ddRAD sequencing (Peterson et al., 2012). The black-capped vireo (*Vireo atricapilla*; BCVI) source-sink metapopulation in central Texas (Hauser et al., 2019; Walker et al., 2016) serves as an ideal non-model system for marker comparison. The species is recovering from a demographic and genetic bottleneck (Athrey et al., 2012; Grzybowski et al., 1994; McFarland et al., 2013) that resulted in small fragmented populations. Further, the range of the species habitat is highly fragmented through land conversion from their breeding habitat, scrub habitat, to agriculture and human development, resulting in most of the remnant population restricted to protected habitats and military bases. The highest density of BCVI exists around Fort Hood in central Texas, where the species has been monitored carefully as a protected species (ESA Endangered from 1970-2018; Cimprich & Kostelke, 2006; Wilsey et al., 2014). The source-sink system comprises fragmented habitat patches driven by brown-headed cowbird parasitism (Walker et al., 2016) and mediated by riparian corridors (Hauser & Leberg, 2020).

Methods

We collected DNA samples (toenail clips and/or pin feathers) from 338 black-capped vireos from 6 sites throughout central Texas including Fort Hood [East Range (ER), Maxdale (MD), West Range (WR), San Saba Property (SS), Balcones Canyonlands National Wildlife Refuge (BC), and Colorado Bend State Park (CB; see Hauser et al. 2019 for more details)]. These sites span the source-sink system identified by demography (Walker et al., 2016) and microsatellite analyses (Hauser et al., 2019; Figure 1). We extracted DNA from the samples using the Qiagen QIAamp Micro DNA Kit (Qiagen Inc, Hilden, Germany) following the Protocol for Isolation of Genomic DNA from Small Volumes of Blood. We also used these 338 samples in the microsatellite analysis using 12 species-specific loci (Barr et al., 2007; Hauser et al., 2019), and to which we compared the results from the following SNP analysis.

We used 185 of the best quality BCVI samples for de-novo SNP discovery and genotyping. We followed the ddRAD library preparation using the restriction enzymes *speI* and *nlaIII* for paired-end 150 bp reads (Peterson et al., 2012) and sequenced the libraries on an Illumina HiSeq lane. Library preparation, quality control, and sequencing were performed at the Texas A&M AgriLife Genomics core facility in College Station, Texas. Sequence reads (total sequence reads = 802,466,640) were demultiplexed and filtered for poor quality

using the `process_radtags` function in Stacks v2.0 (Rochette et al., 2019), retaining 1,960,156 reads. We optimized parameters for the *de novo* pipeline, resulting in the following parameters: $m = 3$, $M = 2$, $n = 1$, $r = 0.80$, $\text{min_maf} = 0.05$. We filtered the dataset further in VCFtools for minor allele count ($\text{mac} = 3$) and genotyping rate (80%, Danecek et al., 2011). For direct comparison, the microsatellite dataset ($n = 338$) was subsampled to the same 185 individuals for which SNP data was produced. We designed the following analysis methods to parallel the microsatellite analyses described in Hauser et al. (2019) and Hauser & Leberg (2020) with minor modifications for large SNP datasets.

Population Genetics

We tested loci and samples for Hardy-Weinberg equilibrium (HWE) deviations and linkage disequilibrium (LD). Loci found to be in LD or deviating from HWE were omitted from further analysis. We calculated observed and expected heterozygosity (H_o and H_e , respectively) using `basic.stats` function in `hierfstat` R package (v 3.5.0) and allelic richness (A_r) using the `allel.rich` function in `hierfstat` R package (v 3.5.0) to estimate genetic diversity across the populations (Goudet, 2005). To evaluate how populations differed across these three metrics (H_o , H_e and A_r), we performed a randomized block ANOVA, blocking by locus, using the `'aov'` function in R with a post-hoc Tukey HSD test using the `TukeyHSD` R function. In these, and subsequent analyses, we corrected alpha levels for multiple comparisons using a standard Bonferroni correction (Hauser et al., 2019; Rice, 1989; Sethuraman et al., 2019). Wherever calculating p-values with iterations were computationally impossible for our resources, we assessed significance using 95% confidence intervals (Altman & Krzywinski, 2016; Gardner & Altman, 1986).

We estimated population genetic differentiation (pairwise F_{ST}) using the `pairwise.WCfst` function in R package `hierfstat` (v 3.5.0), estimating 95% confidence intervals with the `boot.ppfst` function in the same R package (Goudet, 2005). We assessed population structure using the Bayesian clustering program STRUCTURE (v 2.3.4). We used the admixture model with population as a prior to determine the number of unique genetic clusters (k) present within our system, testing k values ranging from 1 to 6. We performed these runs with 10 iterations, 500,000 burn-in period, and 500,000 MCMC (Monte Carlo Markov Chain) repetitions. We then submitted the STRUCTURE results to STRUCTURESELECTION and used the Evanno and Puechmaille methods to determine k (Li & Liu, 2018; Puechmaille, 2016).

We used several approaches to investigate patterns of gene flow among the populations, specifically to determine if there was directional gene flow. Using GENECLASS (v 2.0), we detected first generation migrants using `'L_home/L_max'` likelihood ratio, the Paetkau et al. (1995) criterion, 0.01 allelic frequency, and 0.01 p-value threshold. We could not estimate migration in BayesAss as it does not allow for large SNP datasets (maximum of 40 loci). We used parentage assignments in CERVUS (v 3.0.7) to directly observe migration among populations (Kalinowski et al., 2007). Second-year (SY) individuals disperse and establish their first breeding territories, while older individuals have strong site fidelity and stay in the same population for subsequent years. Therefore, we assigned parents after-second-year age or older (ASY) to SY individuals as offspring assuming that if a SY is in a different population than their assigned parent, then that the offspring was a migrant. SY individuals found to be in the same population as their parents were considered residents. For both programs, we assigned 81 of the SY offspring to candidate mothers ($n = 21$) and fathers ($n = 56$) using a SNP dataset with a 90% genotyping rate. A more stringent genotyping rate was used to avoid biases associated with missing data and parentage analyses (Hammerly et al., 2016) and to ensure that the program could accommodate the dataset (Kalinowski et al., 2007).

Landscape Genetics

For all landscape genetics analyses, we used the proportion of shared alleles (Dps) as a metric of gene flow, as Dps is more directly related to gene flow than other metrics of genetic differentiation (Landguth et al., 2010). We tested for isolation by distance at an individual and population level using the `mantel.randtest` function in the R package `adeigenet` (Jombart, 2008).

We used the same between-site and at-site variable database as Hauser et al (2020) including elevation, Euclidean distance, water, development, forest, scrub, open, agriculture, riparian, proportion of scrub habitat,

and brown-headed cowbird (BHCO) management at the sites. Between-site variables (elevation, Euclidean distance, water, development, forest, scrub, open, agriculture, and riparian) were transformed into resistance surfaces in CIRCUITSCAPE (McRae, 2006). We optimized the valuation of each resistance surface (see Hauser et al., 2020) for more details on optimization) using a linear-mixed effect model (R package lme4; Bates et al., 2015) with D_{ps} as the response variable, each resistance value as the fixed effect, and site as the random effect. Only the optimized resistance values for a given variable, the value with the lowest AICc score via the univariate linear-mixed effect models, were then used in subsequent hypotheses testing.

To investigate how landscape features influence gene flow in this system, we used a multivariate linear mixed effect model approach using candidate models that were driven by *a priori* hypotheses. All candidate models were checked for multicollinearity using a variance inflation factor (VIF) threshold of 4 before fitting the models. We used the linear mixed effect models in the R package lme4 using the full maximum likelihood with D_{ps} as the response variable, landscape features as fixed effects, and site as the random effect (Bates et al., 2015). We evaluated our candidate models with AICc, $\Delta AICc$, and AICcweights (R package GeNetIt). We considered models with a $\Delta AICc < 2$ to be competitive (Burnham & Anderson, 2002). Across all methods we compared results from the SNP data, the subsampled microsatellite data, and the full microsatellite data ($n = 338$) presented in Hauser et al. (2019).

Results

Population Genetics

After filtering, the genomic dataset included 11,507 SNP loci for 178 individuals with a mean coverage of 18.2x. The microsatellite dataset was also subsampled to the same 178 individuals. For both datasets, we found no deviations from HWE or LD after a Bonferroni correction at any of our study sites.

For the SNPs, we found significant differences for H_o , H_e and A_r among populations ($P < 0.001$; Figure 2, bottom panel). All sites except SS had significantly lower H_o than H_e . There were significant differences in H_o across populations in three broad groupings: CB had the lowest H_o values, BC, MD and WR had the intermediate values, and ER and SS had the highest H_o values. There were also significant differences in A_r among populations, namely that ER and WR were significantly higher than the rest of the populations (Figure 2). Values for all genetic diversity metrics and their variances calculated using microsatellites were much higher than those using SNPs (Table 1). For the microsatellites, we found significant differences for H_o and A_r among populations ($P < 0.001$) but not for H_e ($P = 0.549$). All sites except BC and CB had significantly lower H_o than H_e (Figure 2, top panel). MD was the only population with an estimate of H_o that was significantly different from the other 6 populations. Across populations, there were no significant differences in H_e or A_r , with the exception of a significant difference in A_r between BC and ER. The full microsatellite data from Hauser et al. (2019) exhibited no differences in any of the genetic diversity metrics across populations.

All pairwise F_{ST} values based on SNPs were statistically significant, except between ER and WR (Figure 3). We found the largest population differences between central Texas site CB and the Fort Hood sites. Central Texas sites were differentiated from WR, ER and MD (increasing in that order). There was no overall pattern that central Texas sites were more similar with other central Texas sites than with Fort Hood sites or vice versa. F_{ST} values calculated with microsatellites were an order of magnitude higher than those calculated with SNPs. The full microsatellite dataset showed that most genetic differentiation was between central Texas sites and Fort Hood sites (Table 2; Hauser et al., 2019). WR and ER, Fort Hood sites, were the only significantly differentiated populations relative to the rest of the source-sink system. Likewise, the full microsatellite analysis showed that most differentiation was between central Texas sites and Fort Hood sites (Hauser et al., 2019). The Puechmaile method, which accounts for uneven sampling (Li & Liu, 2018; Puechmaile, 2016), showed two unique genetics clusters for the microsatellite data and the full microsatellite dataset (Hauser et al., 2019) while the same approach using SNP markers identified only one cluster. All sets of STRUCTURE barplots based on SNPs showed no population subdivision and considerable mixing regardless of marker across k values ranging from 2 to 6 (Figure 4; Hauser et al., 2019).

Using SNPs, we found 82 migrants, with 33 detected in central Texas sites and 49 detected in Fort Hood sites using GENECLASS2 (Table 3). All migrants detected were from WR. Migrants in central Texas sites comprised a much larger proportion of the total population (14.7 – 29.5%) compared to those detected in Fort Hood (<1% - 6.9%). Using the subsampled microsatellite data, we found 125 migrants, 25 detected in central Texas and 100 detected in Fort Hood. Of these detected migrations, 79 were between central Texas and Fort Hood, 9 were among central Texas sites, and 37 were among Fort Hood sites. Similar to the SNP data, migrants found in central Texas sites comprised a substantially greater portion of the estimated census population size (13.2 - 20.5%) than those in Fort Hood sites (1.8 - 3.0%). Hauser et al (2019) detected fewer migrants overall ($n = 22$), but similar patterns in proportion of migrants in populations were found. Regardless of dataset, proportions of migrants in central Texas populations were an order of magnitude higher than those in Fort Hood.

In CERVUS, the SNP-based analysis did not assign any candidate parents to offspring. The microsatellite analysis assigned 20 parent-offspring pairs at the 95% confidence interval (Table 4). We identified most offspring assigned to parents as migrants ($n = 16$), of which most were from Fort Hood ($n = 14$). We found directional migration from Fort Hood to central Texas ($n = 4$) compared to central Texas to Fort Hood ($n = 1$). The full microsatellite dataset assigned more parent-offspring pairs ($n = 21$) at 95% confidence interval (Hauser et al., 2019) and indicated similar patterns of directional migration from Fort Hood to central Texas as the subsampled dataset.

Landscape Genetics

For all datasets, there was no evidence of isolation by distance at either a population or individual level (microsatellites: $P = 0.92$, $P = 0.492$, respectively; SNPs: $P = 0.092$, $P = 0.946$, respectively; Hauser et al., 2019; Figure 5). The null model of isolation by distance in our multivariate linear mixed effect models only had low support via AIC evaluation for the SNPs and the full microsatellite dataset.

Top models with $\Delta AIC_c < 2$ for the SNP dataset were “Riparian + Water + Scrub” and “Development” (Table 5). From our analyses, riparian areas facilitated gene flow ($\beta = 0.022$, respectively), scrub habitat facilitated gene flow ($\beta = 0.0087$), water impeded gene flow ($\beta = -0.033$) and development impeded gene flow ($\beta = -0.0038$). The top models from Hauser et al. (2020) were ‘Agriculture + Development + Open’ and ‘Riparian + Agriculture + Open’. Hauser et al. (2020) indicated that agriculture and riparian areas facilitated gene flow while development and open habitat impeded gene flow. Common variables across the two datasets were riparian and development; the relationships of habitat with gene flow were similar. For the subsampled microsatellite data, 11 of the 20 candidate models had $\Delta AIC_c < 2$, including the null model, indicating a substantial loss of power using microsatellites with this reduced sample size (Table 5).

Discussion

There was overarching agreement in the inferences based on SNP and microsatellites datasets; both types of markers detected the BCVI source-sink system with WR and ER as putative source populations and the remaining populations as sinks. While we found agreement between the two marker types in the overall patterns (i.e., asymmetrical gene flow, weak structuring, and admixture), specific results differed between the datasets. Among population genetic estimates, we found statistically significant heterozygosity deficiencies in many populations, higher allelic richness in ER and WR, statistically significant pairwise F_{ST} values among population pairs, and detection of first-generation migrants. The microsatellite analyses found fewer differences in heterozygosity or allelic richness among populations, and few pairwise F_{ST} tests were significant. The SNP dataset was unsuccessful in reconstructing parentage, potentially due to insufficient power associated with biallelic markers compared to multiallelic microsatellites. For the landscape genetic results, the subsampled microsatellite data failed to identify any relevant top models. While the SNP and the full microsatellite analyses (Hauser et al., 2020) yielded two important landscape variables in common (riparian and developed), but the top models from these datasets were not in agreement for other variables.

Many of the discrepancies between the SNP and microsatellite results can be attributed to the higher loci number and thus greater statistical power associated with SNP datasets. A large number of biallelic

SNPs deflate and restrict the range of heterozygosity, allelic richness, and F_{ST} values compared to those of multi-allelic microsatellites (Weir & Hill, 2002). Regardless of marker used, most BCVI populations had lower heterozygosity than expected, and putative source populations ER and WR, had significantly higher allelic richness than the rest of the sites, but SNPs yielded fewer overlapping and smaller confidence intervals with these genetic diversity estimates (higher precision). Our ability to detect fine-scale genetic differentiation using SNPs improved with greater pairwise differentiation (F_{ST}). Previous studies have found that SNPs are more accurate at estimating genetic diversity metrics (Muñoz et al., 2017; Seddon et al., 2005) and genetic structuring (Liu et al., 2005; Morin et al., 2009; Seddon et al., 2005). However, the Bayesian clustering approach STRUCTURE was unable to detect fine-scale population structuring for either marker. This software has been found to perform poorly with fine-scale structure (Janes et al., 2017) and likely could not disentangle small levels of structuring in this metapopulation. BCVIs show strong fine-scale structuring (Athrey et al 2015), which may contribute to the results here. With high levels of gene flow characteristic of a metapopulation (Edelaar & Bolnick, 2012), we would not expect to see strong genetic structuring in the BCVI source-sink system.

The inability to reconstruct parentage using SNPs in the present study may be due to the lower information content of SNP markers compared to the multiallelic microsatellites. Several studies have shown that microsatellites outperform SNPs with parentage analyses because of their high polymorphism information content per locus (Defaveri et al., 2013; Weinman et al., 2015). Further, parentage depends primarily upon heterozygosity values to reconstruct relationships. As SNPs have lower heterozygosity values, they consequently lose the ability to reconstruct relationships (Kaiser et al., 2017; Morin et al., 2004; Tokarska et al., 2009; Weinman et al., 2015). Morin et al (2004) indicated that a heterozygosity minimum of 0.20 is required for paternity exclusion analyses, but Blouin et al (1996) has found that even higher values ($H_e = 0.60 - 0.75$) would be necessary to accurately reconstruct 1st order relationships. As the maximal heterozygosity value possible with SNP loci is 0.50 (Tokarska et al., 2009), it is unsurprising that SNPs often provide insufficient information to reconstruct parentage. Currently there is no consensus on the superior marker for parentage analyses (Flanagan & Jones, 2019; Thrasher et al., 2018). Our other population genetic results support the assertion that while SNP data have substantially higher statistical power than microsatellite data, these benefits do not extend to parentage analysis due to the low heterozygosity values (0.153 – 0.176). For this system and many other non-model systems in which low genetic diversity and/or bottlenecks have occurred (i.e., threatened or endangered species), microsatellites may be the superior marker for parentage reconstruction.

Despite SNPs' purported higher resolution into population genetic processes, as seen here and many other comparisons (Kaiser et al., 2017; Kleinman-Ruiz et al., 2017; Seddon et al., 2005), significant findings do not necessarily translate to biologically relevant differences. Statistically significant differences found in genetic diversity (heterozygosity and allelic richness) and structure metrics (F_{ST}) among the BCVI study sites were extremely small (on the order of thousandths) and may lack biological significance. When calculating population genetic metrics, large SNP datasets, such as ours, increase the chance of statistically significant results (using p-values or 95% confidence intervals) and Type I error of results (Wigginton et al., 2005). The insights gleaned from the microsatellite and SNP data were similar in terms of population dynamics: high connectivity among sites and asymmetrical gene flow from source populations (WR and ER) to sink populations.

This study serves as one of the first direct marker comparisons in a landscape genetic context showing varying results between SNPs and microsatellites. Neither SNPs nor microsatellites found any evidence for isolation by distance, as would be expected for a metapopulation with considerable admixture as found here (Gaggiotti, 1996; Jenkins et al., 2010). Isolation-by-distance as a model in the linear mixed model analysis consistently showed low support for the SNP and full microsatellite datasets. While both datasets identified overlapping landscape variables (riparian and human development) we found some discrepancies between the top models of each marker. The landscape genetic analyses with SNPs identified additional landscape variables to the full microsatellite dataset (Hauser & Leberg, 2020): scrub and water, while not identifying agriculture and open habitat in top models. Scrub is the breeding habitat of the BCVI and would be expected

as both a top model and an important landcover type for facilitating dispersal. Large water bodies in this area, not associated with riparian areas, are likely driving the negative relationship between water and gene flow. The full microsatellite analysis found that agricultural areas facilitated gene flow, opposite to predictions and BCI observations. The subsampled microsatellite data did not yield any significant top models as it indicated that 11 of the 20 *a priori* models were equally informative and as equally as informative as the null model, isolation by distance, for which formal testing showed no relationship. While we cannot say which marker produces the more accurate results in this system, landscape genetic analyses using microsatellites require higher sampling compared to SNP analyses. Nevertheless, it is promising that both markers identify similar landscape variables (riparian and scrub) that have been corroborated by observational and telemetry data (Dittmar et al., 2014).

Formal comparisons between SNPs and microsatellites have been lacking in landscape genetics, especially in populations recovering from bottlenecks. While our comparison helps to fill said gap, it is not satisfactory in a complete investigation of marker performance. Genetic distance metrics are often more precisely estimated using SNPs (Morin et al., 2009; Muñoz et al., 2017) and therefore could yield more accurate landscape genetic inferences. However, metrics such as *Dps* as used in the present study have not been used in formal comparison and simulation studies. We need further investigations in landscape genetics to understand the respective accuracy and precision of microsatellites and SNPs, especially as many contemporary landscape genetics research is being done with one marker or the other.

We show that overall SNP and microsatellite data can infer similar biological processes and patterns. Microsatellites can still be used for a wide variety of population or conservation questions despite a large adoption of genomics techniques in the field. We especially want to make this assertion for systems with existing or legacy microsatellite panels, in which development of new markers would be costly, piecemeal genotyping is commonplace (as found in management), or where bioinformatics expertise or computational power is not accessible. Further, microsatellites often are the marker of choice for parentage assignment as they are (at present) better suited than SNPs, economical, and demonstrably repeatable. In species with low genetic diversity or that have experienced bottlenecks, especially prevalent in conservation genetics, microsatellites will provide the necessary power in parentage analyses that SNPs cannot. Nevertheless, in developing new molecular markers for a population genetic study, SNPs are less expensive per locus than microsatellites and have substantially more statistical power than microsatellites for most comparisons, yielding a cost-effective approach over microsatellites. SNPs also allow for investigation into adaptive variation with loci under selection whereas microsatellites cannot (Ahrens et al., 2018; Helyar et al., 2011). We urge researchers to thoroughly consider the utility of each marker based on their system and urge reviewers and editors to not disregard research using microsatellites. This comparison serves as an illustration of such a case where microsatellites and SNPs results converge in conclusions and microsatellites still maintain a function in population genetics.

Acknowledgments

We would like to acknowledge our funding through the Department of Defense, Strategic Environmental Research and Development Program (SERDP), through project RC-2120 to J. Lawler. Special thanks to the private landowners, Texas State Parks and Wildlife, Colorado Bend State Park, Balcones Canyonlands National Wildlife Research, and the Natural Resources Division, particularly D. Cimprich and his team at Fort Hood for providing us access and resources for this research.

References:

- Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and progress. *Molecular Ecology*, 27 (6), 1342–1356.
- Altman, N., & Krzywinski, M. (2016). Points of significance: P values and the search for significance. *Nature Methods*, 14 (1), 3–4.
- Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis.

Molecular Ecology ,23 (7), 1661–1667.

Athrey, G., Barr, K. R., Lance, R. F., & Leberg, P. L. (2012). Birds in space and time: Genetic changes accompanying anthropogenic habitat fragmentation in the endangered black-capped vireo (*Vireo atricapilla*). *Evolutionary Applications* , 5 (6), 540–552.

Barr, K. R., Dharmarajan, G., Rhodes, O. E., Lance, R. L., & Leberg, P. L. (2007). Novel microsatellite loci for the study of the black-capped vireo (*Vireo atricapillus*). *Molecular Ecology Notes* ,7 (6), 1067–1069.

Baruch, E., & Weller, J. I. (2008). Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics* , 39 (5), 474–479.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* , 67 (1), 1–29.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*.

Cimprich, D. A., & Kostecke, R. M. (2006). Distribution of the black capped vireo at Fort Hood, Texas. *Southwestern Naturalist* ,51 (1), 99–102.

Clark-Cockerham, C., & Weir, B. S. (1993). Estimation of gene flow from F statistics. *Evolution* , 47 (3), 855–863.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., et al. (2011). The variant call format and VCFtools. *Bioinformatics* , 27 (15), 2156–2158.

Defaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Molecular Ecology Resources* ,13 (3), 377–392.

Dittmar, E. M., Cimprich, D. A., Sperry, J. H., & Weatherhead, P. J. (2014). Habitat selection by juvenile black-capped vireos following independence from parental care. *Journal of Wildlife Management* ,78 (6), 1005–1011.

Edelaar, P., & Bolnick, D. I. (2012). Non-random gene flow: An underappreciated force in evolution and ecology. *Trends in Ecology and Evolution* , 27 (12), 659–665.

Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology* , 28 (3), 544–567.

Gaggiotti, O. (1996). Population Genetic Models of Source-Sink Metapopulations. *Theoretical population biology* , 50 , 178–208.

García, D., Carleos, C., Parra, D., & Cañ, J. (2002). Sib-parentage testing using molecular markers when parents are unknown. *Animal Genetics* , 33 , 364–371.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P values: Estimation rather than hypothesis testing. *British Medical Journal (Clinical research ed.)* , 292 (6522), 746–750.

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* , 5 (1), 184–186.

Grzybowski, J. A., Tazik, D. J., & Schnell, G. D. (1994). Regional Analysis of Black-Capped Vireo Breeding Habitats. *The Condor* ,96 (2), 512–544.

Hall, L. A., & Beissinger, S. R. (2014). A practical toolbox for design and analysis of landscape genetics studies. *Landscape Ecology* ,29 (9), 1487–1504.

Hammerly, S. C., de la Cerda, D. A., Bailey, H., & Johnson, J. A. (2016). A pedigree gone bad: increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population. *Animal Conservation* , 19 (3), 296–303.

- Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* , 7 (9), 1008–1025.
- Hauser, S. S., & Leberg, P. L. (2020). Riparian areas potentially provide crucial corridors through fragmented landscape for black-capped vireo (*Vireo atricapilla*) source-sink system. *Conservation Genetics* , 1 , 3.
- Hauser, S. S., Walker, L., & Leberg, P. L. (2019). Asymmetrical gene flow of the recently delisted passerine black-capped vireo (*Vireo atricapilla*) indicates source-sink dynamics in central Texas. *Ecology and Evolution* , 9 (1), 463–470.
- Hauser, S., Wakeland, K., & Leberg, P. (2019). Inconsistent use of multiple comparison corrections in studies of population genetic structure: Are some type I errors more tolerable than others? *Molecular Ecology Resources* , 19 (1), 144–148.
- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., et al. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources* , 11 (SUPPL. 1), 123–136.
- Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* , 589 , 583–589.
- Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum. *Molecular Ecology* , 26 (14), 3594–3602.
- Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., et al. (2010). A meta-analysis of isolation by distance: Relic or reference standard for landscape genetics? *Ecography* , 33 (2), 315–320.
- Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* , 24 (11), 1403–1405.
- Kaiser, S. A., Taylor, S. A., Chen, N., Sillett, T. S., Bondra, E. R., & Webster, M. S. (2017). A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources* , 17 (2), 183–193.
- Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* , 16 (5), 1099–1106.
- Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J., et al. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics* , 18 (1), 556–568.
- Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* , 19 (19), 4179–4191.
- Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources* , 18 (1), 176–177.
- Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* , 6 (Suppl. 1), S26.
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* , 4 (12), 981–994.
- McFarland, T. M., Mathewson, H. A., Groce, J. E., Morrison, M. L., & Wilkins, R. N. (2013). A Range-Wide Survey of the Endangered Black-Capped Vireo in Texas. *Southeastern Naturalist* , 12 (1), 41–60.
- McRae, B. H. (2006). Isolation by Resistance. *Evolution* , 60 (8), 1551–1561.

- Morin, P. A., Archer, F. I., Pease, V. L., Hancock-Hanser, B. L., Robertson, K. M., Huebinger, R. M., Martien, K. K., et al. (2012). Empirical comparison of single nucleotide polymorphisms and microsatellites for population and demographic analyses of bowhead whales. *Endangered Species Research* , 19 (2), 129–147.
- Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* , 19 (4), 208–216.
- Morin, P. A., Martien, K. K., & Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* , 9 (1), 66–73.
- Muñoz, I., Henriques, D., Jara, L., Johnston, J. S., Chávez-Galarza, J., De La Rúa, P., & Pinto, M. A. (2017). SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*). *Molecular Ecology Resources* , 17 (4), 783–795.
- Narum, S. R., Banks, M., Beacham, T. D., Bellinger, M. R., Campbell, M. R., Dekoning, J., Elz, A., et al. (2008). Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* , 17 (15), 3464–3477.
- Narum, Shawn R, Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* , 22 (11), 2841–2847.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* , 7 (5), e37135.
- Puckett, E. E., & Eggert, L. S. (2016). Comparison of SNP and microsatellite genotyping panels for spatial assignment of individuals to natal range: A case study using the American black bear (*Ursus americanus*). *Biological Conservation* , 193 , 86–93.
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* , 16 (3), 608–627.
- Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution* , 4 (22), 4399–4428.
- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* , 43 (1), 223–225.
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* , 28 (21), 4737–4754.
- Seddon, J. M., Parker, H. G., Ostrander, E. A., & Ellegren, H. (2005). SNPs in ecological and conservation studies: A test in the Scandinavian wolf population. *Molecular Ecology* , 14 (2), 503–511.
- Sethuraman, A., Gonzalez, N. M., Grenier, C. E., Kansagra, K. S., Mey, K. K., Nunez-Zavala, S. B., Summerhays, B. E. W., et al. (2019). Continued misuse of multiple testing correction methods in population genetics—A wake-up call? *Molecular Ecology Resources* , 19 (1), 23–26.
- Sork, V. L., Davis, F. W., Westfall, R., Flint, A., Ikegami, M., Wang, H., & Grivet, D. (2010). Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology* , 19 (17), 3806–3823.
- Spencer, C. C., Neigel, J. E., & Leberg, P. L. (2000). Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology* , 9 (10), 1517–1528.
- Steiner, C. C., Putnam, A. S., Hoeck, P. E. A., & Ryder, O. A. (2013). Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences* , 1 (1), 261–281.

- Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., & Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources* , 18 (5), 953–965.
- Toews, D. P. L., Brelsford, A., Grossen, C., Milá, B., & Irwin, D. E. (2016). Genomic variation across the Yellow-rumped Warbler species complex. *The Auk* , 133 (4), 698–717.
- Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J. M., Pertoldi, C., Kristensen, T. N., Loeschke, V., et al. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: The case of European bison. *Heredity* , 103 (4), 326–332.
- Väli, Ü., Einarsson, A., Waits, L., & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* , 17 (17), 3808–3817.
- Walker, L. E., Marzluff, J. M., & Cimprich, D. A. (2016). Source-sink population dynamics driven by a brood parasite: A case study of an endangered songbird, the black-capped vireo. *Biological Conservation* , 203 , 108–118.
- Weinman, L. R., Solomon, J. W., & Rubenstein, D. R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources* , 15 (3), 502–511.
- Weir, B. S., & Hill, W. G. (2002). Estimating F-Statistics. *Annual Review of Genetics* , 36 , 721–50.
- Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A Note on Exact Tests of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics* , 76 (5), 887–893.
- Wilsey, C. B., Lawler, J. J., Cimprich, D., & Schumaker, N. H. (2014). Dependence of the endangered black-capped vireo on sustained cowbird management. *Conservation Biology* , 28 (2), 561–571.
- Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology* , 27 (6), 777–797.
- Zimmerman, S. J., Aldridge, C. L., & Oyler-McCance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics* , 21 (1), 382–382.

Data Accessibility:

Primer Genbank accession numbers: EF363782- EF36378295

Genotype data will be available on Dryad in VCF and genepop file formats.

Author Contributions:

S. Hauser, G. Athrey, and P. Leberg conceived the project. S. Hauser carried out the laboratory protocols and analyzed the data with supervision by G. Athrey. S. Hauser took the lead in writing the manuscript with critical feedback from G. Athrey and P. Leberg.

Tables

Table 1. Summary of sample size, expected heterozygosity (H_e), observed heterozygosity (H_o), and allelic richness (A_r) over 12 microsatellite loci. 95% confidence intervals are in parentheses. The subsampled microsatellite analysis is featured in the top panel ($n = 178$) and the SNP analysis is featured in the bottom panel ($n = 178$).

Microsatellites

SNPs

Note. BC: Balcones Canyonlands National Wildlife Refuge; CB: Colorado Bend State Park; ER: East Range (Fort Hood); M

Table 2. Genetic differentiation between sites sampled for black-capped vireos. Pairwise F_{ST} values are depicted on the lower left and 95% confidence intervals are depicted on the upper right. Values that are significant, i.e., the 95% confidence interval overlaps with 0, are in bold. The subsampled microsatellite analysis is featured in the top panel (n = 178) and the SNP analysis is featured in the bottom panel (n= 178).

Microsatellites	
SS	
BC	
CB	
ER	
MD	
WR	
SNPs	
SS	
BC	
CB	
ER	
MD	
WR	
Note.SS: San Saba Property; BC: Balcones Canyonlands National Wildlife Refuge; CB: Colorado Bend State Park; ER: Eas	

Table 3. The total number of detected 1st generation migrants (# M), proportion of total migrants detected (% M), estimated population size (N), and proportion of abundances that are migrants (%N) in each population (GENECLASS2). The subsampled microsatellite analysis is featured in the top panel (n = 178) and the SNP analysis is featured in the bottom panel (n= 178).

Microsatellites	
Population	
Central Texas	
Fort Hood	
SNPs	

Population
Central Texas

Fort Hood

Note. BC: Balcones Canyonlands National Wildlife Refuge; CB: Colorado Bend State Park; ERc: East Range (Fort Hood);

Table 4. The number of offspring assigned to candidate offspring (CERVUS) and designated as Migrants or Residents (N) using the subsampled microsatellite data. Directional movement between regions (Fort Hood (FH) and central Texas (CT) (ex: FH to CT denotes movement from Fort Hood to central Texas), percentage of each subcategory, migrants or residents, (%) and percentage of total number of assigned offspring (% Total) are also shown. Assignment data using SNP data not shown as there were no successful parent-offspring assignments.

Movement	Movement	%	% Total
Migrants	16	-	80
FH to CT	4	25	20
CT to FH	1	6	5
CT to CT	1	6	5
FH to FH	10	63	50
Residents	4	-	20
CT	0	0	0
FH	4	100	20

Table 5. Summary of linear mixed effects models results including AICc, delta AICc values and AICc weights for the candidate models. Bolded values indicate delta AICc values less than 2. Scores for our null model, isolation by distance are italicized. Numbers next to variable names (e.g., Ag100) indicate the optimized value parameterized for the associated variable in CIRCUITSCAPE. The subsampled microsatellite analysis is featured in the left panel (n = 178) and the SNP analysis is featured in the right panel (n= 178).

Microsatellites
Model
Open2
Wetland001
Ag100
Euclidean Distance (Null)
Elevation
Water100
Scrub05
Dev100
Forest001
Elevation + Wetland001
Elevation + Scrub05
Wetlands001 + Water100 + Scrub05
Ag100 + Dev100 + Open2
From_BHCO
From_Scrub
Wetland001 + Ag100 +Open2

Wetlands001 + Water100 + Ag100 + Open2

From_Scrub + From_BHCO

Water100 + Wetland001 + Scrub05 + From_BHCO + From_Scrub

Water100 + Wetland001 + Ag100 + Open2 + From_Scrub + From_BHCO

Note: Ag: agriculture, Dev: developed, Distance: Isolation-by-distance, From_BHCO: the level of BHCO control at the emi

Figure Captions

Figure 1. Black-capped vireo study sites in central Texas (black circles) including Balcones Canyonlands (BC), Colorado Bend State Park (CB), San Saba Property (SS) and on Fort Hood (black triangles) including East Range combined (ER), Maxdale (MD), and West Range combined (WR). The six landscape cover types depicted as follows: agricultural croplands in brown, human development in magenta, forest in green, open habitat (including grazing lands) in yellow, scrub in orange, water bodies in navy blue, and wetlands in light blue.

Figure 2. Genetic diversity estimates (dots) with 95% confidence intervals (error bars) per BCVI population: BC, CB, ER, MD, SS, WR. Observed and expected heterozygosity (blue and orange, respectively) per population in the left panel and allelic richness (Ar, in black) per population on the right panel. Estimates in which their 95% confidence intervals overlap are not statistically different.

Figure 3. Pairwise F_{ST} estimates (dots) with 95% confidence intervals (error bars) between BCVI populations: BC, CB, ER, SS, MD, WR. Estimates that overlap with 0 are not statistically significant and estimates in which 95% confidence intervals overlap are not statistically different from one another.

Figure 4. Weak to no population structuring among BCVI populations (BC, CB, SS, ER, MD, and WR). STRUCTURE barplots for k values (number of unique clusters) 2 through 4. Each vertical line represents the genetic signature of an individual with colors representing each cluster.

Figure 5. No signature of isolation by distance at an individual (left panel) or population level (right panel). The relationship between genetic similarity (proportion of shared alleles; Dps) on the y-axis and Euclidean distance (in meters; m) on the x-axis.

Hosted file

BCVI_SNP_Figure1_v2.pdf available at <https://authorea.com/users/392014/articles/506031-waste-not-want-not-microsatellites-remain-an-economical-and-informative-technology-for-conservation-genetics>



