MEETING REPORT

WILEY | Evolutionary Applications

Recent advances in conservation and population genomics data analysis

Sarah Hendricks[1] (iD) | Eric C. Anderson[2,3] (iD) | Tiago Antao[4] | Louis Bernatchez[5] |
Brenna R. Forester[6] | Brittany Garner[7,8] | Brian K. Hand[7] | Paul A. Hohenlohe[1] |
Martin Kardos[7] (iD) | Ben Koop[9] | Arun Sethuraman[10] | Robin S. Waples[11] (iD) |
Gordon Luikart[7,8]

[1]Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho

[2]Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Santa Cruz, California

[3]University of California, Santa Cruz, California

[4]Division of Biological Sciences, University of Montana, Missoula, Montana

[5]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada

[6]Department of Biology, Colorado State University, Fort Collins, Colorado

[7]Flathead Lake Biological Station, Montana Conservation Genomics Laboratory, Division of Biological Science, University of Montana, Missoula, Montana

[8]Wildlife Program, Fish and Wildlife Genomics Group, College of Forestry and Conservation, University of Montana, Missoula, Montana

[9]Department of Biology, Centre for Biomedical Research, University of Victoria, Victoria, British Columbia, Canada

[10]Department of Biological Sciences, California State University San Marcos, San Marcos, California

[11]NOAA Fisheries, Northwest Fisheries Science Center, Seattle, Washington

Correspondence
Sarah Hendricks, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844.
Email: shendri4@gmail.com

Abstract

New computational methods and next-generation sequencing (NGS) approaches have enabled the use of thousands or hundreds of thousands of genetic markers to address previously intractable questions. The methods and massive marker sets present both new data analysis challenges and opportunities to visualize, understand, and apply population and conservation genomic data in novel ways. The large scale and complexity of NGS data also increases the expertise and effort required to thoroughly and thoughtfully analyze and interpret data. To aid in this endeavor, a recent workshop entitled "Population Genomic Data Analysis," also known as "ConGen 2017," was held at the University of Montana. The ConGen workshop brought 15 instructors together with knowledge in a wide range of topics including NGS data filtering, genome assembly, genomic monitoring of effective population size, migration modeling, detecting adaptive genomic variation, genomewide association analysis, inbreeding depression, and landscape genomics. Here, we summarize the major themes of the workshop and the important take-home points that were offered to students throughout. We emphasize increasing participation by women in population and conservation genomics as a vital step for the advancement of science. Some important themes that emerged

during the workshop included the need for data visualization and its importance in finding problematic data, the effects of data filtering choices on downstream population genomic analyses, the increasing availability of whole-genome sequencing, and the new challenges it presents. Our goal here is to help motivate and educate a worldwide audience to improve population genomic data analysis and interpretation, and thereby advance the contribution of genomics to molecular ecology, evolutionary biology, and especially to the conservation of biodiversity.

KEYWORDS

bioinformatics pipeline, conservation genomics workshop, diversity in STEM, landscape genomics, population genomics

1 | INTRODUCTION

At this time, conservation and evolutionary geneticists can employ the power of genomic tools to answer questions in conservation that could not be answered using traditional genetics approaches (Allendorf, Hohenlohe, & Luikart, 2010; Bernatchez et al., 2017; Garner et al., 2016; Harrisson, Pavlova, Telonis-Scott, & Sunnucks, 2014; McMahon, Teeling, & Höglund, 2014; Shafer et al., 2015a, 2015b). Technological and analytical advances now allow us to use many thousands of loci, gene expression, or epigenetics to address basic questions of relevance for conservation, such as identifying loci associated with local adaptation or adaptive potential in species face changing environments (Bernatchez, 2016; Flanagan, Forester, Latch, Aitken, & Hoban, 2017; Harrisson et al., 2014; Hoban et al., 2016; Hoffmann et al., 2015; Jensen, Foll, & Bernatchez, 2016; Le Luyer et al., 2017; Wade et al., 2016). As conservation genomics matures, new challenges are arising. It is essential for researchers to keep up with the rapidly changing methods in appropriate study design, data quality assessment, and selecting appropriate analyses to obtain accurate results for conservation and management decisions (Benestan et al., 2016).

To address arising challenges, 15 experts from diverse areas of genomic data analysis came together to teach and exchange ideas about cutting-edge approaches for population genomic data analysis and interpretation. Students, postdocs, faculty, and agency researchers (e.g., museums, agency biologists) originating from 15 countries brought an assortment of data to work through various computational analyses. Of 31 students, 23 had restriction-site associated DNA (RAD) or genotyping by sequencing (GBS) data, four had exon capture data, and four students had whole-genome sequencing (WGS) data. Interestingly, of the 30 attendees at ConGen just 4 years ago, only a few students had RAD-seq data, only one had sequence capture data, and none had WGS data. The main focus of the 15 experts was on narrow-sense conservation genomics applications, which require use of conceptually novel approaches (Garner et al., 2016).

The week-long workshop, held at the University of Montana's Flathead Lake Biological Station, provided training in theory as well as empirical applications of NGS data production and analyses. Lectures, discussions, hands-on analysis of empirical data, and one-on-one assistance from instructors improved students' knowledge of conservation and evolutionary genomic projects. Many participants in the past have taken the knowledge and resources (PowerPoint slides, worksheets, video recorded lectures) acquired during the workshop and disseminated it to others in their laboratories, further extending the educational reach of ConGen among population genomic researchers (http://www.umt.edu/sell/cps/congen2017/).

In the opening keynote lecture, L. Bernatchez discussed several mechanisms that may enhance the maintenance of genetic variation and evolutionary potential in the face of a changing environment. Among these mechanisms that have been overlooked and should be considered in future theoretical development and predictive models, he discussed the prevalence of soft sweeps, the polygenic basis of adaptation, balancing selection, and transient polymorphisms, as well as epigenetic variation. A key message was that adaptive evolution in nature rarely involves the fixation of beneficial alleles. Instead, adaptation apparently proceeds most commonly by soft sweeps entailing shifts in frequencies of alleles being shared between differentially adapted populations. At last, L. Bernatchez argued that a new paradox seems to be emerging from recent studies whereby populations of highly reduced effective population sizes (N_e) and impoverished genetic diversity can sometimes retain their adaptive potential, and that epigenetic variation could account for this apparent contradiction (Bernatchez, 2016).

The remaining lectures focused mainly on approaches for data production or analysis. We discuss highlights from these lectures with the goal of motivating and educating a worldwide audience to improve population genomic data analysis and thereby advance the role of genomics in molecular ecology, evolutionary biology, and conservation. We describe (a) issues regarding recruiting and retaining a diverse workforce in conservation genomics, (b) impacts of genotyping error and data quality, and (c) improvements to downstream population genomic analyses.

2 | INCREASING CONTRIBUTIONS BY WOMEN (SARAH HENDRICKS AND BRENNA FORESTER)

Following the productive trend at recent ecology and evolutionary biology conferences, issues of gender bias were discussed at ConGen. When this important topic is not widely and openly examined, it can inhibit the advancement of science generally, and conservation and population genomics specifically. Diversity leads to better problem-solving, expands the talent pool, and promotes full inclusion of excellence across the social spectrum (Blackburn, 2017; Nielsen et al., 2017). Among the plethora of topics regarding increasing diversity in STEM fields (Blackburn, 2017; Wellenreuther & Otto, 2015), here we focus on overcoming the biases against women in computer sciences and the persistence of unconscious gender stereotypes that influence both male and female researchers.

Gender biases in computer science training may limit the effectiveness of efforts to attract and retain the best and most diverse workforce in conservation genomics. As of 2014, just 18.1% of computer science bachelor's degrees were awarded to women, and this proportion has declined by 10% over the last 10 years, further widening the gender gap (NCSES, 2016). This deficit in female computer scientists has been attributed to a lower sense of belonging by women than men due to a predominately male culture in the field (Cheryan, Ziegler, Montoya, & Jiang, 2017). There is also evidence of gender gaps in self-efficacy that may be due to a lack of sufficient early education in computer programming (Cheryan et al., 2017). Although not reported, these issues likely persist in bioinformatics and genomics. Efforts to maximize gender inclusion in computer science may benefit from changing masculine cultures in technological fields and providing early experiences for all students that signal a sense of belonging and ability to succeed in these fields. Efforts led by women, such as "Girls Who Code" (https://girlswhocode.com/) and "Learn to Code with Me" (https://learntocodewith.me/posts/13-places-women-learn-code/), aim to decrease the gender gap by targeting coding courses and workshops to girls and women. Likewise, short courses such as ConGen, which teach basics in linux, bash, and R scripting, act to support an inclusive community and address limitations due to gendered perceptions in the genomics era.

Unconscious stereotypes persist in the minds of male and female researchers, as evident in the studies of reference letters for postdoctoral fellowships and other academic positions (Dutt, Pfaff, Bernstein, Dillard, & Block, 2016; Madera, Hebl, & Martin, 2009; Trix & Psenka, 2003). One study of recommendation letters for medical faculty positions found that letters written on behalf of females differed from those written on behalf of men in length, negative language, and gender-linked terms. Overall, the study found that the letters, regardless of the gender of the recommender, reinforced stereotypes that portray men as researchers and professionals and women as teachers and students (Trix & Psenka, 2003). Another study found that men, more than women, were described as having agentic leadership traits, such as being in control of subordinates, speaking assertively, working independently and competitively, and

initiating tasks (Madera et al., 2009). Furthermore, women were described as having more communal characteristics, which had a negative association for women with employment decisions (Madera et al., 2009). Letters of recommendation have been shown to greatly affect hireability ratings of applicants (Madera et al., 2009). On the level of personal action, we suggest recommenders edit their own letters to avoid gender bias (http://www.csw.arizona.edu/LORbias).

Despite similar proportions of women and men awarded doctoral degrees in science and engineering disciplines, women are less likely to obtain tenure-track positions in academia than their male counterparts. Although there are many reasons for this "leaky pipeline" (Gasser & Shaffer, 2014; Goulden, Mason, & Frasch, 2011; Holmes, OConnell, & Dutt, 2015), increasing training and avoiding biases in reference letters may benefit not only women, but also the greater scientific community by promoting innovation through diversity and inclusion. Further, there are many topics such as referee opportunity bias (Lerback & Hanson, 2017), the childcare-conference conundrum (Calisi & A Working Group of Mothers in Science, 2018), and misconceptions around hiring preferences (Williams & Ceci, 2015) that should also be addressed to reduce disadvantages to women. With the brief mention of this topic, we hope to stimulate future studies of gatekeeping practices in the field of conservation, so institutions can develop initiatives to recruit, retain, and advance women in STEM fields as mentorship will be essential for eliminating gender bias in computer science, bioinformatics, and by extension, conservation biology. We ask our readers to initiate discussions regarding the persistence of stereotypes and how these stereotypes affect excellence across our community. We wonder: Can the active and intentional cultivation of inclusivity help to expand the role of genomics in molecular ecology, population genomics, and nature conservation?
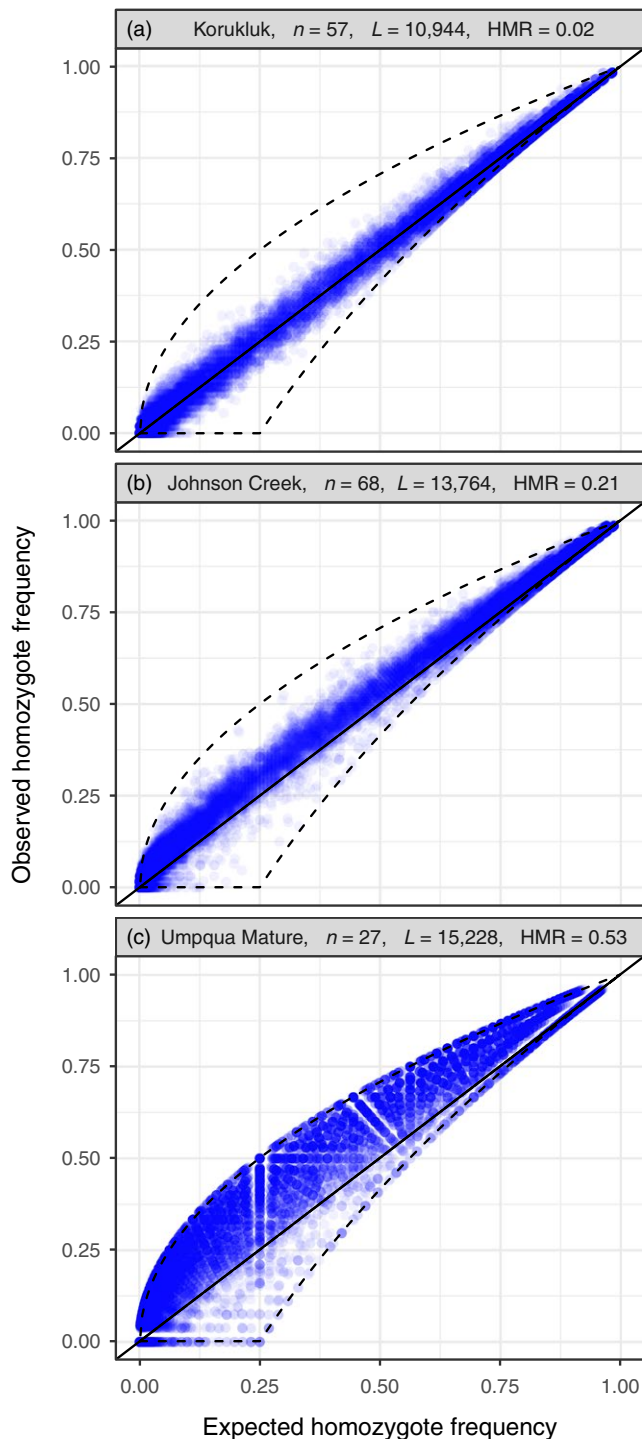
3 | GENOTYPING ERROR AND IMPROVING DATA QUALITY

On a more technical level, several authors discussed ways to assess and prevent genotyping errors and improve data quality. We discuss several of these here.

3.1 | Back to the basics: finding and visualizing erroneous data (Eric Anderson and Robin Waples)

3.1.1 | Genotyping errors

Systematic departures from Hardy–Weinberg equilibrium (HWE) in datasets where HWE is expected can indicate genotyping errors in which heterozygotes are miscalled as homozygotes. A simple visualization of expected and observed frequencies of homozygote genotypes across single nucleotide polymorphisms (SNPs) can be effective in identifying data problems (Figure 1). A simple model for estimating the heterozygote miscall (dropout) rate was applied to 12 publicly available RAD-seq datasets (Fernández et al., 2016; Hecht, Matala, Hess, & Narum, 2015; Laporte et al., 2016; Larson

FIGURE 1 Observed (*y*-axis) versus expected (*x*-axis) homozygote frequencies at SNPs in three RAD studies of Chinook salmon. The solid black line is at *y* = *x*, and the dotted lines show the maximum and minimum possible observed values given the expected values computed from the observed allele frequencies. *n* is number of individuals, L is number of SNPs, and HMR is the heterozygote miscall rate estimated from the dataset. (a) Korukluk River, Western Alaska (Larson et al., 2014): a carefully filtered dataset showing almost no distortions from HWE and with a low estimated HMR of 0.02. (b) Johnson Creek (Hecht et al., 2015): Most of the points lie above the *y* = *x* line and HMR is estimated to be 0.17. (c) Low-read-depth data from mature-migrating Umpqua river Chinook (Prince et al., 2017). Genotypes were called using ANGSD's doGeno option assuming a uniform prior on genotypes. Profound homozygote excesses are observed with HMR = 0.52

this provides a cautionary note that it is good practice to visualize your data to ascertain if more homozygotes are called than expected under Hardy–Weinberg equilibrium.

3.1.2 | Probabilistic genotype calling

Probabilistic genotype calling, as conducted by the software program *ANGSD* (Korneliussen, Albrechtsen, & Nielsen, 2014), is a principled method for dealing with low-coverage sequencing data; however, it should be applied carefully. With low-coverage sequencing, because there is so little information at any individual site, the statistical model and the prior distributions are relatively more influential than they are with high-read-depth data. A good example can be seen in a recent paper by Prince et al. (2017) which features lower-depth sampling than many other contemporary RAD-seq studies. In analyses of their RAD-seq data, Prince et al. used ANGSD to integrate over the genotype uncertainty rather than directly calling genotypes. Even more importantly, when they were able to, they were careful to use population-specific allele frequency-based genotype priors for their analyses rather than a simple uniform prior distribution on genotypes. The choice of prior is important: If one uses ANGSD to call genotypes from the Prince et al. data using the uniform prior on genotypes, the result shows a strong tendency to incorrectly infer heterozygotes as homozygotes (Figure 1c). This is not simply a consequence of forcing ANGSD to call genotypes. Rather, the posterior probabilities, themselves, of the genotypes carry extra weight on the homozygote classes, because the uniform prior does not use allele frequency information to help infer the genotypes.

In an increasing manner, recent publications have suggested that probabilistic genotyping obviates the need for high mean depth of coverage (>10 to 20×). For example, Prince et al. (2017) found that PCA analysis applied to their full dataset yielded a first principal component driven largely by variation in read depth (M. Miller, personal communication, February 7, 2018). Randomly subsampling reads from each individual to the same depth eliminated that technical variation, and, though it led them to discard almost 70% of their sequencing reads, with probabilistic genotyping they were still able to recover meaningful population structure. To evaluate

et al., 2014; Le Moan, Gagnaire, & Bonhomme, 2016; Portnoy et al., 2015; Prince et al., 2017; Puritz, Gold, & Portnoy, 2016; Ravinet et al., 2016; Swaegers et al., 2015). While a few had low genotyping error rates (<5%), in others, allelic dropout, low read depth, PCR duplicates, erroneous assembly, and/or poor filtering resulted in much higher estimated error rates, with between 5% and 72% of heterozygotes apparently being miscalled as homozygotes. Although some of these apparent high error rates could reflect true heterozygote deficiencies due to the Wahlund effect or other factors, in all cases the samples were thought to be from a single population. Hence,
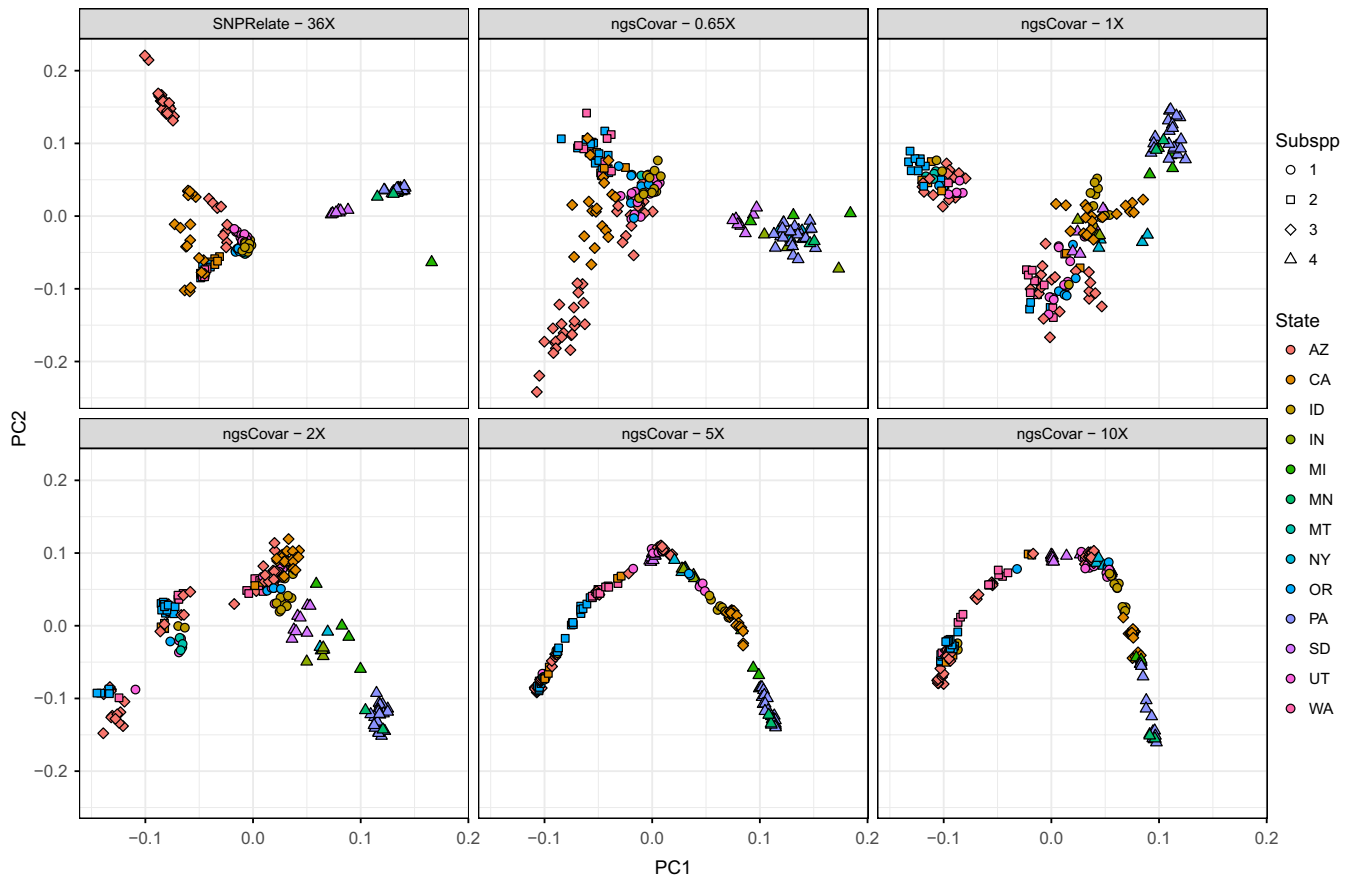
FIGURE 2 Plots of the first two principal components from PCA of unpublished RAD data showing population structure among four subspecies of a North American passerine. Each point is an individual bird. Top left panel shows the result obtained in the original study, which used 105,000 SNPs called with an average read depth of 36× across 175 birds analyzed with *SNPRelate* (Zheng et al., 2012). Remaining panels show results obtained by subsampling the original dataset to depths of 0.65×, 1×, 2×, 5×, and 10×, and analysis with *ANGSD* (Korneliussen et al., 2014) and *ngsCovar* (Fumagalli et al., 2014). Subspecific structure in the 0.65× data is much less distinct than in the full dataset, but is generally concordant with it. However, at higher read depths the clustering is clearly inconsistent with subspecies affiliation

how effectively probabilistic genotype calling can retrieve the same inference with ever-smaller amounts of sequencing, Anderson presented an analysis using subsampled versions of a high-depth RAD dataset. He first performed PCA using *SNPRelate* (Zheng et al., 2012) to resolve population structure of a North American songbird using SNPs called from high-quality, high-read-depth RAD data using a GATK pipeline (mean read depth at 105,000 SNPs across 175 individuals was 36). He then used *ANGSD* and *ngsCovar* (Fumagalli, Vieira, Linderoth, & Nielsen, 2014) a probabilistic genotyping approach to PCA, on the BAM files for the same 175 birds after subsampling so that the mean read depth at each of those 105,000 loci was expected to be 0.65, 1, 2, 5, and 10. *ANGSD* was not restricted to using only the previously discovered 105,000 SNPs, and, in fact called between 29,331 SNPs at 0.65× and 898,320 SNPs at 10×. Figure 2 shows that clusters in the first two principal components from *SNPRelate* on the high-read-depth data resolve subspecies and show structure within subspecies that corresponds to state of origin. Remarkably, at 0.65×, *ngsCovar* identifies roughly similar groupings, albeit with looser clustering. However, at all other read depths, *ngsCovar* identifies clusters that are clearly inconsistent with subspecies

designations and become dominated by Lissajous curves (Novembre & Stephens, 2008).

Overall, the results suggest that some probabilistic methods developed for low-coverage data might behave unpredictably when provided with high-quality, high-read-depth RAD data. However, new methods based on probabilistic genotyping are continually emerging. For example, the ANGSD methods PCAngsd and PCA_MDS are both reported to outperform ngsCovar with variable sequencing depth (see http://www.popgen.dk/angsd/index.php/PCA). Probabilistic inference from next-generation sequencing data is an important advance; however, one should not assume that it will automatically overcome shortcomings in sequence data caused by unsatisfactory sample quality, poor library preparation, or insufficient sequencing. As with many approaches for next-generation sequencing, user-specified settings of models, priors, and filtering can have strong effects on the results.

3.1.3 | Relatedness

Many researchers have concluded that it is important to remove putative siblings from population genetics datasets before conducting

downstream analyses (Corlett, 2017; Johnson et al., 2016), but there are several good reasons why this can create more problems than it solves (Waples & Anderson, 2017). First, siblings occur naturally in all natural populations, at frequencies that are inversely related to effective population size; therefore, removing siblings erases signals characteristic of small populations and makes the populations appear to be larger. Second, removing individuals reduces sample size and decreases statistical power, perhaps greatly, so any benefits must be large to offset this cost. Third, methods for sibling inference are not infallible, so it is important to consider the consequences of imperfect pedigree reconstruction. At last, sibling removal cannot be used to convert a nonrandom sample into a random sample, unless one has independent information about the degree to which the proportion of siblings in the sample exceeds the random expectation.

An alternative to removing individuals is to use a best linear unbiased estimator approach (BLUE; McPeek, Wu, & Ober, 2004), which gives each individual a weight that reflects its degree of relatedness to others in the sample. As shown by Waples and Anderson (2017), however, performance of the BLUE also depends on having accurate pedigree information. When sample identification is not reliable, the use of the full dataset outperforms BLUE. Because of these potential adverse effects, researchers should be cautious about adjusting their datasets for putative siblings unless they have a good reason to believe that doing so will not actually make things worse.

3.2 | Effects of filtering on downstream analyses (Paul Hohenlohe and Tiago Antao)

Methods for producing reduced representation libraries, such as RAD-seq, are rapidly evolving, and more than 15 methods exist with variations in data quality, genotyping errors, cost, and the number of loci discovered (reviewed in Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). Furthermore, filtering choices (see figure 2 in Benestan et al., 2016) can greatly influence downstream summary statistics. A recent study testing the impact of data processing on population genetic inferences using RAD-seq data observed large differences between reference-based and de novo approaches in population genetic summary statistics, particularly those based on the site frequency spectrum (Shafer et al., 2016). In addition, the recent debate over the effectiveness of RAD-seq for discovering loci under selection (Catchen et al., 2017; Lowry et al., 2016; McKinney, Larson, Seeb, & Seeb, 2017) has highlighted the importance of testing the extent of linkage disequilibrium (LD) over the genome, whenever possible, in order to assess the power of genome scans to detect selected loci (e.g., Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016).

To further explore the impacts of filtering on downstream analyses, students at ConGen used various minor allele frequencies (MAF; 0.01, 0.05, 0.1, and 0.2) to filter a RAD-seq dataset and computed F_{ST} using the populations function in *Stacks* (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Participants detected a general trend of increasing estimates of genomewide mean F_{ST} with higher MAF thresholds. This may be the result of the relationship

between expected heterozygosity and maximum possible F_{ST} at SNP loci; given the variation in F_{ST} across loci, a subsample of loci with lower MAF may be expected to have a lower maximum and therefore lower mean F_{ST} (Roesti, Salzburger, & Berner, 2012). Thus, some filtering by MAF can be used to remove sequencing errors and avoid bias in genome scans (Roesti et al., 2012) and may also remove rare alleles that are less informative for estimating F_{ST}. On the other hand, imposing MAF filters that are too strict (e.g., above 0.05 or 0.1) could skew metrics based on the site frequency spectrum or inadvertently remove loci under selection or with functional significance. As others have recommended, testing the effects of a range of analytical (filtering) parameters is critical to produce robust population genetic and demographic inferences (Mastretta-Yanes et al., 2014; Paris, Stevens, & Catchen, 2017; Shafer et al., 2016).

3.2.1 | Stringent filtering

The *Anopheles gambiae* 1,000 Genomes Project (Ag1000G) is a large-scale project to sequence the main vector of malaria, mosquitoes (*Anopheles gambiae*; The Anopheles gambiae Genomes Consortium, 2017), and it has conducted extensive empirical verification of error rates and filtering rules. Parents from different mosquito colonies were mated and produced ~19 offspring for each of four crosses. WGS of all individuals produced a minimum mean coverage of at least 14×. The error rate of SNP variant calling (inferred from parent–offspring inheritance) without filtering was between 13.0% and 21.7%. After filtering, the Mendelian error rate fell to 0.3%–0.9%. The filtering rules devised from this empirical dataset were then applied to the WGS analysis of 765 mosquitoes sampled across Africa. Not using any filtering with GATK would have produced 95,335,499 SNPs, but with optimized filtering rules the number of SNPs fell to 52,525,957 (see Supplementary material of The Anopheles gambiae Genomes Consortium (2017) for filtering parameters). Filtering parameters are dataset dependent and should be modified based on multiple criteria (e.g., depth of coverage, mapping quality, and strand bias) to reduce the number of false discoveries (see GATK forms on applying hard filters for detailed information).

3.3 | Retaining haplotypes in amplicon and RAD datasets (Eric Anderson)

Common approaches for dealing with multiple SNPs across an amplicon or RAD locus can result in low power or incorrect inference in subsequent analyses. When multiple SNPs are detected, these SNPs are handled as either unlinked (likely untrue) or only one of the SNPs is used in downstream analyses. However, retaining each haplotypic combination as an allele can increase power for relationship inference and pedigree reconstruction (Baetscher, Clemento, Ng, Anderson, & Garza, 2017). Further, haplotype calling allows for the retention of low-frequency variants, which may be useful for population structure assessment in recently diverged populations. Rare alleles (or haplotypes) reveal recombination events that generated alternative sequences of ancestry and thereby identify fine-scale

structure that would be missed when using independent marker approaches (Lawson, Hellenthal, Myers, & Falush, 2012).

The software microhaplot (https://github.com/ngthomas/microhaplot) takes a variant file and designates nucleotides that occur together on the same read as "microhaplotypes" and allows for the visualization, filtering, and exporting of the data. The *Stacks* software package (Catchen et al., 2013) can also export multi-SNP haplotypes from RAD-seq data. Unlike single SNP assays, the microhaplotype data collection method uses assays designed with multi-allelic loci and can yield useful data for nontarget species phylogenies and for genealogical inference (Sunnucks, 2000).

3.4 | Draft genomes to improve data analyses (Ben Koop)

Some molecular biologists have claimed that we are in the post-genomic era (Wu, 2001); however, only a very small proportion of reference genomes are assembled to the chromosomal level. Despite this, having even a draft genome (in 1000s of scaffolds) can help improve data analyses in many ways including the following: (a) reliable discovery of SNPs (e.g., avoiding duplicated loci), (b) reducing genotyping error rates (Hand et al., 2015; Shafer et al., 2016), (c) detecting loci under selection by allowing sliding-window approaches along scaffolds (Hohenlohe, Phillips, & Cresko, 2010), and (d) finding the underlying genes associated with phenotype or adaptation (facilitated by mapping scaffolds to related species with well-annotated genomes; e.g., Ekblom & Wolf, 2014; Kohn, Murphy, Ostrander, & Wayne, 2006; McKinney et al., 2016). In addition, it is possible with this information to estimate effective population size (N_e; e.g., Li & Durbin, 2011) or effective number of breeders (N_b) using LD-based methods, as comparisons can be restricted to pairs of loci on different scaffolds, which should reduce or eliminate LD due to physical linkage. Depending on the genome size and complexity, an investment of $10k to $20k could achieve a useful reference genome with an N50 of ~100 kbp, which can be sufficient to improve data analysis as mentioned above (see Goodwin, McPherson, & McCombie, 2016) for costs per Gb for various sequencing platforms). Furthermore, the reference assembly could likely be provided by a commercial company (e.g., DoveTail, https://dovetailgenomics.com/) for this price, as long as the genome is not too large ($\gg$3 GB) or complex (e.g., duplicated, numerous repeats), and if the initial DNA is of high molecular weight (many fragments >15–20 kb).

There are a growing number of approaches for genome assembly using "single molecule real-time" sequencing (SMRT-seq) or "synthetic long-read" sequencing (SLR-seq) technology (Fuentes-Pardo & Ruzzante, 2017; Goodwin et al., 2016). The SMRT-seq technology offered by PacBio (http://www.pacb.com/) produces read lengths of ~10 kbp (some >60 kbp). Oxford Nanopore (https://nanoporetech.com/) and minION also use a single molecule approach to nucleotide identification that passes an ionic charge through a nanoscale hole and measures the changes in current as each molecule passes through (see Michael et al. (2017) for assembly comparison). SLR-seq

technologies, such as 10× Genomics (https://www.10xgenomics.com/) or Dovetail Genomics (https://dovetailgenomics.com/), still rely on short read technology and, using statistical phasing algorithms, have the capacity to assemble continuous haplotypes and scaffolds that can span whole chromosomes with high accuracy.

While the per sample cost of WGS is still relatively high, the per locus cost is low compared to reduced representation library costs (see table 1 in Oyler-McCance, Oh, Langin, & Aldridge, 2016). A greater proportion of positions within the genome are covered with WGS, which lowers the per base sequencing costs, but increases the costs per individual. With sequencing prices still falling, it is becoming more likely that most ecologists and evolutionary biologists will have access to genome assemblies for their study species (or sister taxa) in the near future (Ellegren, 2014).

3.5 | Experimental design: which method to choose (Paul Hohenlohe)

The diversity of options for experimental design of population genomic studies continues to expand as sequencing costs continue to drop and new technologies emerge. As discussed in previous ConGen workshops (Benestan et al., 2016), a general guideline is to consider carefully the biological question, and the downstream analyses and statistical power that will be required to most efficiently address it. This should guide all aspects of experimental design, including the genomic approach, type of genetic markers, number of markers, sequencing depth, number of individuals and populations sampled, spatial distribution of individuals, and tissue type (for transcriptome sequencing). For all of these factors, there is a wide range of options for most population genomic studies, as well as trade-offs among methods and sampling approaches that are important to consider (Andrews et al., 2016; Benestan et al., 2016).

Focusing on the choice of sequencing method, a particular point of discussion at the ConGen 2017 workshop was the recent set of papers addressing the limitations of RADseq to illuminate the genetic basis of adaptation (Catchen et al., 2017; Lowry et al., 2016; McKinney et al., 2017). The primary criticism raised by Lowry et al. (2016) is that RAD loci, depending on the choice of restriction enzyme(s) and the specific protocol used (Andrews et al., 2016), may be sparsely distributed across the genome, so that selected loci may lie some distance away from the nearest genotyped RAD marker. By definition, all reduced representation approaches face this issue, although RADseq approaches are more limited than other techniques (such as sequence capture) in their ability to specifically target previously identified candidate loci. In a RADseq study (and most other marker-based population genomic studies), the key factor is linkage disequilibrium (LD), which determines the extent to which genotypes at a genetic marker are correlated with those of a functionally important locus, and therefore, the signal of selection that can be detected from marker data.

If the scale of LD is larger than the distance between markers, a RAD-seq study has a high probability of identifying functionally important loci across the genome. The extent of LD can be directly

estimated if a reference genome is available (Catchen et al., 2017), and it is recommended that LD should be estimated whenever possible in population genomic studies. Moreover, many conservation and population genomic questions can be answered without exhaustive sampling of the genome or detection of all functionally important loci, and alternative techniques such as WGS may impose substantial costs and other trade-offs (Catchen et al., 2017). In particular, increasing the density of markers may necessitate reducing the number of individuals or populations sampled, and choosing methods that target candidate loci can bias against detecting selection at previously unknown loci. Overall, there is no universally applicable genomic method, and the biological question and details of the study system should drive the choice of technique.

4 | IMPROVING DOWNSTREAM COMPUTATIONAL ANALYSES

4.1 | Genomic analysis of inbreeding and demographic history (Marty Kardos)

In a traditional manner, individual inbreeding has been measured with the pedigree inbreeding coefficient (F_p) via path analysis (Pemberton, 2008). More recently, large numbers of genetic markers (Berenos, Ellis, Pilkington, & Pemberton, 2016; Hoffman et al., 2014; Huisman, Kruuk, Ellis, Clutton-Brock, & Pemberton, 2016) and whole-genome sequences (Kardos et al., 2018; Palkopoulou et al., 2015; Xue et al., 2015) have been used to estimate individual inbreeding directly from the genome by analyzing parameters like multiple-locus heterozygosity, genomic relatedness matrices, and Runs Of Homozygosity (ROH; Kardos et al., 2016). Genomic approaches capture variation in realized inbreeding that is missed by pedigree analysis due to the stochastic effects of linkage and unknown common ancestors of parents (Franklin, 1977; Thompson, 2013). Thus, while deep and accurate pedigrees can often precisely measure individual inbreeding in species with many chromosomes and/or high recombination rates (Kardos et al., 2018; Knief, Kempenaers, & Forstmeier, 2017; Nietlisbach et al., 2017), genomic approaches are expected to more reliably measure inbreeding and inbreeding depression (Kardos, Luikart, & Allendorf, 2015a; Kardos et al., 2018; Keller, Visscher, & Goddard, 2011; Wang, 2016). Given that many studies have used only shallow pedigrees or few DNA markers, it is possible that power to detect inbreeding depression has been low; therefore, inbreeding depression could be more common, widespread, and severe than previously thought.

Analyses of ROH can also be used to understand the genetic basis of inbreeding depression. Candidate regions for loci contributing to inbreeding depression can be identified as chromosome segments containing fewer ROH in a sample of individuals than expected by chance (Kardos et al., 2018; Pemberton et al., 2012). Homozygosity mapping (Charlier et al., 2008) and association analyses based on the correlation of phenotype with the presence/absence of ROH in particular genome regions (Keller et al., 2012; Pryce, Haile-Mariam, Goddard, & Hayes, 2014) can be used to identify loci affecting inbreeding depression. Genomic approaches have the potential to greatly advance our understanding of the strength and genetic basis of inbreeding depression in natural populations.

Analyses of identity-by-descent (IBD) can also be used to infer historical effective population size (N_e). Differences in historical N_e among populations can be qualitatively inferred by analyzing the abundance of ROH. The abundance of very short ROH is informative of N_e in distant history, while long ROH is informative of more recent N_e (Kardos, Qvarnström, & Ellegren, 2017; Kirin et al., 2010; Pemberton et al., 2012). A limitation of this approach is that it is only qualitative and requires data on multiple populations to be informative.

A particularly exciting new approach for studies of recent demographic history in natural populations is to explicitly estimate a time series of recent N_e using inference of IBD. The program IBDSeq (Browning & Browning, 2013) searches the genomes of all pairs of individuals to identify chromosome segments of shared ancestry between individuals. The program IBDNe (Browning & Browning, 2015) then uses the inferred pairwise IBD segments to find the most likely recent time series of N_e given the IBD data. A limitation of this approach for most natural populations is that it requires a minimum of approximately 100 individuals and the genetic mapping locations (i.e., on a linkage map) of at least several hundred thousand SNPs (Browning & Browning, 2015). However, the approach has great potential to infer recent demographic history (i.e., to test for and quantify recent population bottlenecks and expansions) in natural populations where it would be difficult or impossible to evaluate recent N_e otherwise (Kardos et al., 2017).

4.2 | Genomewide association studies (Marty Kardos)

Genomewide association studies (GWAS) have recently identified loci with large effects on several ecologically important phenotypic traits. For example, single loci have explained a large fraction of the variance in age of maturation in Atlantic salmon (Barson et al., 2015) and horn development in free-ranging Soay sheep (Johnston et al., 2011, 2013). In an intelligible manner, some traits are governed largely by variation at individual loci, but these are likely rare among all traits of interest to evolutionary biologists. Many adaptive traits are likely driven by a large number of loci with small effect sizes, low minor allele frequency, and/or epistatic interactions (Visscher et al., 2017). GWAS of complex traits will therefore often fail to identify enough genotype–phenotype associations to explain a useful fraction of the heritability of traits of interest. This is particularly true of studies on populations with very large N_e or high recombination rates where strong linkage disequilibrium (LD) extends only very short distances from the genotyped loci, or where relatively few loci are analyzed, thus resulting in low power to detect loci even with relatively large phenotypic effects (Kardos et al., 2015b). However, encouraging for studies in small or fragmented populations, the power to detect large effect quantitative trait loci (QTL) is expected to be higher in populations with small N_e because strong LD extends over longer chromosomal distances in such populations. Therefore, the design and interpretation of GWAS are greatly improved by

evaluating the extent of strong LD and the power to detect large effect QTL.

By good fortune, GWAS failing to explain a large fraction of the heritability in loci with statistically significant genotype–phenotype associations are still highly useful. It is arguably more important in ecological and conservation genetics to understand the heritability of a trait than to identify some of the loci responsible for heritable variation in the trait, as it is the heritability of a trait that determines the magnitude of the expected response to selection. The additive genetic variance and heritability can readily be estimated using linear mixed effects models (Rönnegård et al., 2016; Santure et al., 2013; Yang, Lee, Goddard, & Visscher, 2011) in GWAS, even in cases where no individual loci pass the stringent thresholds of statistical significance. In addition, heritability can be partitioned among chromosomes to determine whether the trait of interest is likely to be polygenic (i.e., affected by a very large number of loci), in which case chromosome-specific heritability is expected to increase with the number of genes on a chromosome (Santure et al., 2013).

Participants at ConGen used the R package, RepeatABEL (Rönnegård et al., 2016), to test for loci associated with clutch size using previously published data from a long-term study of collared flycatchers (*Ficedula albicollis*; Husby et al., 2015). This helped to familiarize students with data structures, available software, and interpretation of results from GWAS. In addition, analyzing the collared flycatcher data allowed students to consider the importance of accounting for repeated phenotypic measurements when conducting a GWAS. Students were encouraged to critically evaluate effect size estimates from GWAS in light of the Beavis effect (Beavis, 1998), and the "winner's curse" (Kraft, 2008), which state that the effect sizes of loci passing a stringent statistical significance thresholds in QTL mapping or GWAS analyses are often upwardly biased, particularly in studies with low statistical power.

4.3 | Landscape genomics (Brenna Forester)

Landscape genomics is an emerging analytical framework that investigates how environmental and spatial processes structure the amount and distribution of neutral and adaptive genetic variation among populations (Balkenhol et al., 2017). Landscape genomics is sometimes conflated with genotype–environment association (GEA) analysis, which includes a wide variety of statistical approaches for identifying candidate adaptive loci that covary with environmental predictors (Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015). However, landscape genomics includes many other techniques for identifying and analyzing spatially structured, selection-driven variation, including GWAS across multiple environments, simulation studies, experimental approaches such as environmentally stratified common gardens, epigenetic and transcriptomic studies, and innovative approaches that combine analytical techniques (Berg & Coop, 2014; Lasky, Forester, & Reimherr, 2018; Storfer, Antolin, Manel, Epperson, & Scribner, 2015).

Most importantly, landscape genomics is not just the application of these statistical techniques to identify candidate adaptive variation, but is an approach with a developing theoretical framework linking

genomic variation, spatial complexity, environmental heterogeneity, and evolutionary processes (Balkenhol, Cushman, Waits, & Storfer, 2015). The wide range of ecological and evolutionary questions and management issues that can be addressed through this framework was highlighted with recent published examples (Brauer, Hammer, & Beheregaray, 2016; Creech et al., 2017; Lasky et al., 2015; Manthey & Moyle, 2015; Razgour et al., 2017; Swaegers et al., 2015).

With this introduction to landscape genomics, ConGen participants worked on applications of GEA analysis, currently the most widely used landscape genomic technique (Balkenhol et al., 2017). The reasons for the popularity of GEA analyses are practical: They require no phenotypic data or prior genomic resources, do not require experimental approaches (such as reciprocal transplants) to demonstrate local adaptation, and are often more powerful than differentiation-based outlier detection methods (De Mita et al., 2013; de Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014; Forester, Lasky, Wagner, & Urban, 2018; Lotterhos & Whitlock, 2015). In particular, participants considered how and why detection rates differed between univariate and multivariate GEAs, exploring the use of latent factor mixed models (Frichot, Schoville, Bouchard, & Francois, 2013) and redundancy analysis (Forester, Jones, Joost, Landguth, & Lasky, 2016; Lasky et al., 2012), respectively. Recent work has shown that RDA is an effective means of detecting adaptive processes that result in weak, multilocus molecular signatures (Forester et al., 2018), providing a powerful tool for investigating the genetic basis of local adaptation and informing management actions to conserve evolutionary potential (Flanagan et al., 2017; Harrisson et al., 2014; Hoffmann et al., 2015). Finally, participants were encouraged to move beyond simply documenting candidate adaptive loci in their datasets, and instead focus on the ecological, evolutionary, and management-relevant questions that can be addressed by more fully integrating a landscape genomic analytical framework.

4.4 | Ancestral demography with migration (Arun Sethuraman)

Estimation of ancestral demography, particularly under an Isolation with Migration (IM) model (Nielsen, 2001), is useful for many molecular ecologists and conservation geneticists. A prominent set of tools for this analysis includes IM, IMa, IMa2, and IMa2p (Hey, 2010; Hey, Chung, & Sethuraman, 2015; Hey & Nielsen, 2007; Sethuraman & Hey, 2015). In general, these methods utilize a Bayesian Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) method to estimate effective population sizes, migration rates, and divergence times under the IM model from haplotypic data. In its latest edition, IMa2p offers parallelized estimation under this framework, providing almost linear improvement in computational time by increasing the number of processors utilized. This in turn allows the analyses of a large number of genomic loci to estimate demographic history, a task that was previously intractable owing to computational overhead. These tools assume that genomic loci are independent, freely recombining between loci, nonrecombining within loci, and putatively neutral (summarized in Strasburg & Rieseberg, 2010). When datasets

fit these assumptions, the methods give robust results (summarized in Sousa & Hey, 2013). As of late, Hey et al. (2015) simulated data where the number of loci sampled was small and exhibited very low divergence between populations to detect an excess of false positive for the presence of migration (also described in Cruickshank & Hahn, 2014). This study points to high false-positive rates for detecting migration using likelihood ratio tests while using the IM suite of tools on data that show low divergence (i.e., very low F_{ST}), and while using a small number of loci. In addition, much like other MCMC methods (e.g., STRUCTURE; Pritchard, Stephens, & Donnelly, 2000), the length of a "run" of the IM tool used is paramount (longer and many duplicate runs preferred) in ensuring mixing, convergence, and adequate sampling of genealogies.

Recent studies that have estimated demography under the IM model as applied to conservation include McKelvy and Burbrink (2017) that tests 24 nested models of evolution and species delineation across the North American range of the yellow-bellied kingsnake (*Lampropeltis calligaster*), and Vázquez-Miranda et al. (2017) study of Le Conte's thrashers (*Toxostoma lecontei*) in estimating negligible migration among subspecies to recommend conservation status across their Western North American range. Other tools to test complex demographic models using genomic data include coalescent simulation-based methods (e.g., FASTSIMCOAL; Excoffier & Foll, 2011; Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013), Approximate Bayesian Computation (ABC; Beaumont, Zhang, & Balding, 2002; Robinson, Bunnefeld, Hearn, Stone, & Hickerson, 2014)-based methods that compare summary statistic distributions in simulated versus observed populations, and diffusion approximations to the joint allele frequency spectrum for demographic inference (e.g., *∂a∂i*, Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). In general, model-based estimation of evolutionary demographic history (both ancient and recent) when applied in combination with summary population genetic statistics as described above (including F_{ST}, inbreeding coefficients, and homozygosity), and non-model-based methods (including STRUCTURE and ADMIXTURE; Alexander, Novembre, & Lange, 2009; Pritchard et al., 2000) can prove to be useful means to bridge genomics and conservation in particular.

5 | BROAD RECOMMENDATIONS AND CONCLUSIONS

Common advice among instructors was to gain extensive experience in computer programming. Students were encouraged to seek out online resources and to work in interdisciplinary teams, where through mentorship and close collaboration they can learn the basics in an applied setting. A key theme was the importance of continuing to develop and teach programming at all levels (e.g., elementary through graduate), with a specific focus on better integrating bioinformatics instruction into undergraduate life sciences education.

The advent of "big data" presents a critical challenge in the fields of population and conservation genomics. Interdisciplinary collaboration is a key as it becomes more difficult for researchers to be experts in both data production (e.g., field work, biological sampling) and bioinformatics or mathematical modeling. Koop acknowledges that he fills his team with bioinformaticians as well as biologists, but "when you find the rare individual who understands both the population genomics and the bioinformatics, you do everything you can to hold onto them." Furthermore, the "Ten Simple Rules for a Successful Cross Disciplinary Collaboration" by Knapp et al. (2015) is a useful resource for gaining skills for a successful, synergistic collaboration.

In conclusion, the genomic era presents both new data analysis challenges and opportunities to visualize, understand, and apply population genomic data to conservation in novel ways. Here, we emphasize producing and visualizing erroneous datasets, possible effects of filtering on downstream analyses, and how to improve downstream computational analyses to prevent drawing erroneous conclusions. The experts at ConGen instructed students to understand and use reliable biological models and to develop clear questions and hypotheses rooted in evolutionary and ecological theory. In summary, ConGen and this article present problems and solutions with the goal of improving the use of genomics in the fields of population genomics, molecular ecology, and conservation biology.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

None declared.

ORCID

Sarah Hendricks http://orcid.org/0000-0001-9571-9461

Eric C. Anderson http://orcid.org/0000-0003-1326-0840

Martin Kardos http://orcid.org/0000-0002-7587-3004

Robin S. Waples http://orcid.org/0000-0003-3362-7590

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2017). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *22*, 4733. https://doi.org/10.1111/1755-0998.12737

Balkenhol, N., Cushman, S. A., Waits, L. P., & Storfer, A. (2015). Current status, future opportunities, and remaining challenges in landscape genetics. In N. Balkenhol, S. A. Cushman, A. T. Storfer & L. P. Waits (Eds.), *Landscape genetics*, Vol. *19* (pp. 247–256). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118525258.ch14

Balkenhol, N., Dudaniec, R. Y., Krutovsky, K. V., Johnson, J. S., Cairns, D. M., Segelbacher, G., ... Joost, S. (2017). Landscape genomics: Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In O. P. Rajora (Ed.), *Population genomics concepts*, *strategies and approaches*. Berlin, Germany: Springer International Publishing. https://doi.org/10.1007/13836_2017_2

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., ... Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in Atlantic salmon. *Nature*, *528*, 405–408. https://doi.org/doi:10.1038/nature16062

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.2307/1390807

Beavis, W. D. (1998). QTL analyses: Power, precision, and accuracy. In A. H. Paterson (Ed.), *Molecular dissection of complex traits* (pp. 145–162). New York: CRC Press.

Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., ... Luikart, G. (2016). Conservation genomics of natural and managed populations: Building a conceptual and practical framework. *Molecular Ecology*, *25*(13), 2967–2977. https://doi.org/10.1111/mec.13647

Berenos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2016). Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. *Molecular Ecology*, *25*(13), 3152–3168. https://doi.org/10.1111/mec.13681

Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, *10*(8), e1004412. https://doi.org/10.1371/journal.pgen.1004412

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519–2556. https://doi.org/10.1111/jfb.13145

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., ... Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, *32*(9), 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Blackburn, H. (2017). The status of women in STEM in higher education: A review of the literature 2007–2017. *Science & Technology Libraries*, *36*(3), 235–273. https://doi.org/10.1080/0194262X.2017.1371658

Brauer, C. J., Hammer, M. P., & Beheregaray, L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, *25*(20), 5093–5113. https://doi.org/10.1111/mec.13830

Browning, B. L., & Browning, S. R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *The American Journal of Human Genetics*, *93*(5), 840–851. https://doi.org/10.1016/j.ajhg.2013.09.014

Browning, S. R., & Browning, B. L. (2015). Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *The American Journal of Human Genetics*, *97*(3), 404–418. https://doi.org/10.1016/j.ajhg.2015.07.012

Calisi, R. M., & A Working Group of Mothers in Science. (2018). Opinion: How to tackle the childcare–conference conundrum. *Proceedings of the National Academy of Sciences*, *115*(12), 201803153-2849. https://doi.org/10.1073/pnas.1803153115

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

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365. https://doi.org/10.1111/1755-0998.12669

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., ... Georges, M. (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics*, *40*(4), 449–454. https://doi.org/10.1038/ng.96

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35. https://doi.org/10.1037/bul0000052

Corlett, R. T. (2017). A bigger toolbox: Biotechnology in biodiversity conservation. *Trends in Biotechnology*, *35*(1), 55–65. https://doi.org/10.1016/j.tibtech.2016.06.009

Creech, T. G., Epps, C. W., Landguth, E. L., Wehausen, J. D., Crowhurst, R. S., Holton, B., & Monello, R. J. (2017). Simulating the spread of selection-driven genotypes using landscape resistance models for desert bighorn sheep. *PLoS ONE*, *12*(5), e0176960. https://doi.org/10.1371/journal.pone.0176960

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*(5), 1383–1399. https://doi.org/10.1111/mec.12182

de Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, *23*(8), 2006–2019. https://doi.org/10.1111/mec.12705

Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S., & Block, C. J. (2016). Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience*, *9*(11), 805–808. https://doi.org/10.1038/ngeo2819

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042. https://doi.org/10.1111/eva.12178

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905–e1003917. https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334. https://doi.org/10.1093/bioinformatics/btr124

Fernández, R., Schubert, M., Vargas-Velázquez, A. M., Brownlow, A., Víkingsson, G. A., Siebert, U., ... Orlando, L. (2016). A genomewide catalogue of single nucleotide polymorphisms in white-beaked and Atlantic white-sided dolphins. *Molecular Ecology Resources*, 16, 266–276. https://doi.org/10.1111/1755-0998.12427

Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2017). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, 119(1), 1–18. https://doi.org/10.1111/eva.12569

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, 25(1), 104–120. https://doi.org/10.1111/mec.13476

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Using genotype-environment associations to identify multilocus local adaptation. *Molecular Ecology*, 27(9), 2215–2233. https://doi.org/doi:10.1111/mec.14584

Franklin, I. R. (1977). The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theoretical Population Biology*, 11(1), 60–80. https://doi.org/10.1016/0040-5809(77)90007-7

Frichot, E., Schoville, S. D., Bouchard, G., & Francois, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 500(6), 207–238. https://doi.org/10.1111/mec.14264

Fumagalli, M., Vieira, F. G., Linderoth, T., & Nielsen, R. (2014). ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, 30(10), 1486–1487. https://doi.org/10.1093/bioinformatics/btu041

Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., ... Luikart, G. (2016). Genomics in conservation: Case studies and bridging the gap between data and application. *Trends in Ecology & Evolution*, 31(2), 81–83. https://doi.org/10.1016/j.tree.2015.10.009

Gasser, C. E., & Shaffer, K. S. (2014). Career development of women in academia: Traversing the leaky pipeline. *The Professional Counselor*, 4(4), 332–352. https://doi.org/10.15241/ceg.4.4.332

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Goulden, M., Mason, M. A., & Frasch, K. (2011). Keeping women in the science pipeline: The ANNALS of the American academy of political and social. *Science*, 638(1), 141–162. https://doi.org/10.1177/0002716211416925

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hand, B. K., Hether, T. D., Kovach, R. P., Muhlfeld, C. C., Amish, S. J., Boyer, M. C., ... Hohenlohe, P. A. (2015). Genomics and introgression: Discovery and mapping of thousands of species-diagnostic SNPs using RAD sequencing. *Current Zoology*, 61(1), 146–154. https://doi.org/10.1093/czoolo/61.1.146

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. https://doi.org/10.1111/eva.12149

Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology*, 24, 5573–5595. https://doi.org/10.1111/mec.13409

Hey, J. (2010). Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, 27(4), 905–920. https://doi.org/10.1093/molbev/msp296

Hey, J., Chung, Y., & Sethuraman, A. (2015). On the occurrence of false positives in tests of migration under an isolation-with-migration model. *Molecular Ecology*, 24(20), 5078–5083. https://doi.org/10.1111/mec.13381

Hey, J., & Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, 104(8), 2785–2790. https://doi.org/10.1073/pnas.0611164104

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburg, G., Lowry, D. B., ... Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379–397. https://doi.org/10.1086/688018

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A., ... Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, 111(10), 3775–3780. https://doi.org/10.1073/pnas.1318945111

Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., ... Sgrò, C. (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, 2(1), 1. https://doi.org/10.1186/s40665-014-0009-x

Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *International Journal of Plant Sciences*, 171(9), 1059–1071. https://doi.org/10.1086/656306

Holmes, M. A., OConnell, S., & Dutt, K. (Eds.). (2015). *Women in the geosciences: Practical, positive practices toward parity*. Hoboken, NJ: John Wiley & Sons. https://doi.org/doi: 10.1002/9781119067573

Huisman, J., Kruuk, L. E. B., Ellis, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), 3585–3590. https://doi.org/10.1073/pnas.1518046113

Husby, A., Kawakami, T., Ronnegard, L., Smeds, L., Ellegren, H., & Qvarnstrom, A. (2015). Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806), 1–7. https://doi.org/10.1098/rspb.2015.0156

Jensen, J. D., Foll, M., & Bernatchez, L. (2016). The past, present and future of genomic scans for selection. *Molecular Ecology*, 25(1), 1–4. https://doi.org/10.1111/mec.13493

Johnson, J. A., Altwegg, R., Evans, D. M., Ewen, J. G., Gordon, I. J., Pettorelli, N., & Young, J. K. (2016). Is there a future for genome-editing technologies in conservation? *Animal Conservation*, 19(2), 97–101. https://doi.org/10.1111/acv.12273

Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502(7469), 93–95. https://doi.org/10.1038/nature12489

Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., ... Slate, J. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, 20(12), 2555–2566. https://doi.org/10.1111/j.1365-294X.2011.05076.x

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., ... Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution*, 2, 124–131. https://doi.org/10.1038/s41559-017-0375-4

Kardos, M., Luikart, G., & Allendorf, F. W. (2015a). Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity*, 115, 63–72. https://doi.org/10.1038/hdy.2015.17

Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W., McWilliam, S., … Kijas, J. (2015b). Whole genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology*, 24, 5616–5632. https://doi.org/10.1111/mec.13415

Kardos, M., Qvarnström, A., & Ellegren, H. (2017). Inferring individual inbreeding and demographic history from segments of identity by descent in Ficedula flycatcher genome sequences. *Genetics*, 205(3), 1319–1334. https://doi.org/10.1534/genetics.116.198861

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, 9(10), 1205–1218. https://doi.org/10.1111/eva.12414

Keller, M. C., Simonson, M. A., Ripke, S., Neale, B. M., Gejman, P. V., Howrigan, D. P., … Schizophrenia Psychiatric Genome-Wide Association Study Consortium. (2012). Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genetics*, 8(4), e1002656-11. https://doi.org/10.1371/journal.pgen.1002656

Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, 189(1), 237–249. https://doi.org/10.1534/genetics.111.130922

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*, 5(11), e13996–e13997. https://doi.org/10.1371/journal.pone.0013996

Knapp, B., Bardenet, R., Bernabeu, M. O., Bordas, R., Bruna, M., Calderhead, B., … Deane, C. M. (2015). Ten simple rules for a successful cross-disciplinary collaboration. *PLoS Computational Biology*, 11(4), e1004214–e1004217. https://doi.org/10.1371/journal.pcbi.1004214

Knief, U., Kempenaers, B., & Forstmeier, W. (2017). Meiotic recombination shapes precision of pedigree- and marker-based estimates of inbreeding. *Heredity*, 118(3), 239–248. https://doi.org/10.1038/hdy.2016.95

Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in Ecology & Evolution*, 21(11), 629–637. https://doi.org/10.1016/j.tree.2006.08.001

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kraft, P. (2008). Curses—Winner's and otherwise—in genetic epidemiology. *Epidemiology*, 19(5), 649–651. https://doi.org/10.1097/EDE.0b013e318181b865

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzent, M., Budzinski, H., … Bernatchez, L. (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, 25, 219–237. https://doi.org/10.1111/mec.13466

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, 7, 355–369. https://doi.org/10.1111/eva.12128

Lasky, J. R., Forester, B. R., & Reimherr, M. (2018). Coherent synthesis of genomic associations with phenotypes and home environments. *Molecular Ecology Resources*, 18(1), 91–106. https://doi.org/10.1111/1755-0998.12714

Lasky, J. R., Marais des, D. L., McKay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Molecular Ecology*, 21(22), 5512–5529. https://doi.org/10.1111/j.1365-294x.2012.05709.x

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., … Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, 1(6), e1400218. https://doi.org/10.1126/sciadv.1400218

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, 8(1), e1002453. https://doi.org/10.1371/journal.pgen.1002453

Le Luyer, J., Laporte, M., Beacham, T. D., Kaukinen, K. H., Withler, R. E., Leong, J. S., … Bernatchez, L. (2017). Parallel epigenetic modifications induced by hatchery rearing in a Pacific Salmon. bioRxiv, 148577. https://doi.org/10.1101/148577

Le Moan, A., Gagnaire, P. A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, 25, 3187–3202. https://doi.org/10.1111/mec.13627

Lerback, J., & Hanson, B. (2017). Journals invite too few women to referee. *Nature*, 541, 455–457.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. https://doi.org/10.1038/nature10231

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031–1046. https://doi.org/10.1111/mec.13100

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016). Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 1–40, https://doi.org/10.1111/1755-0998.12596

Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, 94(6), 1591–1599. https://doi.org/10.1037/a0016539

Manthey, J. D., & Moyle, R. G. (2015). Isolation by environment in White-breasted Nuthatches (Sitta carolinensis) of the Madrean Archipelago sky islands: A landscape genomics approach. *Molecular Ecology*, 24(14), 3628–3638. https://doi.org/10.1111/mec.13258

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2014). Restriction site-associated DNA sequencing, genotyping error estimation and de novoassembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. https://doi.org/10.1111/1755-0998.12291

McKelvy, A. D., & Burbrink, F. T. (2017). Ecological divergence in the yellow-bellied kingsnake (Lampropeltis calligaster) at two North American biodiversity hotspots. *Molecular Phylogenetics and Evolution*, 106, 61–72. https://doi.org/10.1016/j.ympev.2016.09.006

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: Comment on Breaking RAD by Lowry et al. (2016). *Molecular Ecology Resources*, 17(3), 356–361. https://doi.org/10.1111/1755-0998.12649

McKinney, G. J., Seeb, L. W., Larson, W. A., Gomez Uchida, D., Limborg, M. T., Brieuc, M. S. O., … Seeb, J. E. (2016). An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook salmon (Oncorhynchus tshawytscha). *Molecular Ecology Resources*, 16(3), 769–783. https://doi.org/10.1111/1755-0998.12479

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, 7(9), 999–1007. https://doi.org/10.1111/eva.12193

McPeek, M. S., Wu, X., & Ober, C. (2004). Best linear unbiased allele-frequency estimation in complex pedigrees. *Biometrics*, 60(2), 359–367. https://doi.org/10.1111/j.0006-341X.2004.00180.x

Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., … Ecker, J. R. (2017). High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. *bioRxiv*, *000*, 1–18. https://doi.org/10.1101/149997

NCSES. (2016). Women, minorities, and persons with disabilities in science and engineering: 2017 Digest, pp. 1–28.

Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. *Heredity*, *86*(6), 641–647. https://doi.org/10.1046/j.1365-2540.2001.00895.x

Nielsen, M. W., Alegria, S., Börjeson, L., Etzkowitz, H., Falk-Krzesinski, H. J., Joshi, A., … Schiebinger, L. (2017). Opinion: Gender diversity leads to better science. *Proceedings of the National Academy of Sciences*, *114*(8), 1740–1742. https://doi.org/10.1073/pnas.1700616114

Nietlisbach, P., Keller, L. F., Camenisch, G., Guillaume, F., Arcese, P., Reid, J. M., & Postma, E. (2017). Pedigree-based inbreeding coefficient explains more variation in fitness than heterozygosity at 160 microsatellites in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), https://doi.org/10.1098/rspb.2016.2763

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*(5), 646–649. https://doi.org/10.1038/ng.139

Oyler-McCance, S. J., Oh, K. P., Langin, K. M., & Aldridge, C. L. (2016). A field ornithologist's guide to genomics: Practical considerations for ecology and conservation. *The Auk*, *133*(4), 626–648. https://doi.org/10.1642/AUK-16-49.1

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., … Dalén, L. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, *25*(10), 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, *188*, 799–814. https://doi.org/10.1111/2041-210X.12775

Pemberton, J. M. (2008). Wild pedigrees: The way forward. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), 613–621. https://doi.org/10.1098/rspb.2007.1531

Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., & Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics*, *91*(2), 275–292. https://doi.org/10.1016/j.ajhg.2012.06.014

Portnoy, D. S., Puritz, J. B., Hollenbeck, C. M., Gelsleichter, J., Chapman, D., & Gold, J. R. (2015). Selection and sex-biased dispersal in a coastal shark: The influence of philopatry on adaptive variation. *Molecular Ecology*, *24*, 5877–5885. https://doi.org/10.1111/mec.13441

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., … Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, *3*, https://doi.org/10.1126/sciadv.1603198

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Pryce, J. E., Haile-Mariam, M., Goddard, M. E., & Hayes, B. J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution*, *46*(1), 71. https://doi.org/10.1186/s12711-014-0071-7

Puritz, J. B., Gold, J. R., & Portnoy, D. S. (2016). Fine-scale partitioning of genomic variation among recruits in an exploited fishery: Causes and consequences. *Scientific Reports*, *6*, 36095. https://doi.org/doi:10.1038/srep36095

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of Littorina saxatilis at a local scale. *Molecular Ecology*, *25*, 287–305. https://doi.org/10.1111/mec.13332

Razgour, O., Taggart, J. B., Manel, S., Juste, J., Ibáñez, C., Rebelo, H., … Park, K. (2017). An integrated framework to identify wildlife populations under threat from climate change. *Molecular Ecology Resources*, *91*(5), 2437. https://doi.org/10.1111/1755-0998.12694

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Robinson, J. D., Bunnefeld, L., Hearn, J., Stone, G. N., & Hickerson, M. J. (2014). ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology*, *23*(18), 4458–4471. https://doi.org/10.1111/mec.12881

Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, *12*(1), 94. https://doi.org/10.1186/1471-2148-12-94

Rönnegård, L., McFarlane, S. E., Husby, A., Kawakami, T., Ellegren, H., & Qvarnström, A. (2016). Increasing the power of genome wide association studies in natural populations using repeated measures—evaluation and implementation. *Methods in Ecology and Evolution*, *7*(7), 792–799. https://doi.org/10.1111/2041-210X.12535

Santure, A. W., De Cauwer, I., Robinson, M. R., Poissant, J., Sheldon, B. C., & Slate, J. (2013). Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. *Molecular Ecology*, *22*(15), 3949–3962. https://doi.org/10.1111/mec.12376

Sethuraman, A., & Hey, J. (2015). IMa2p—parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. *Molecular Ecology Resources*, *16*(1), 206–215. https://doi.org/10.1111/1755-0998.12437

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2016). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. https://doi.org/10.1111/2041-210X.12700

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., … Zieliński, P. (2015a). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*(2), 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Colling, G., Dalén, L., … Zieliński, P. (2015b). Reply to Garner et al. *Trends in Ecology & Evolution*, *31*(2), 1–2. https://doi.org/10.1016/j.tree.2015.11.010

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404–414. https://doi.org/10.1038/nrg3446

Storfer, A., Antolin, M. F., Manel, S., Epperson, B. K., & Scribner, K. T. (2015). Genomic approaches in landscape genetics. In N. Balkenhol, S. A. Cushman, A. T. Storfer & L. P. Waits (Eds.), *Landscape genetics*, Vol. *1* (pp. 149–164). Chichester, UK: John Wiley & Sons Ltd. https://doi.org/10.1002/9781118525258.ch09

Strasburg, J. L., & Rieseberg, L. H. (2010). How robust are "isolation with migration" analyses to violations of the im model? A simulation study. *Molecular Biology and Evolution*, *27*(2), 297–310. https://doi.org/10.1093/molbev/msp233

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, *15*(5), 199–203. https://doi.org/10.1016/S0169-5347(00)01825-5

Swaegers, J., Mergeay, J., Van Geystelen, A., Therry, L., Larmuseau, M. H. D., & Stoks, R. (2015). Neutral and adaptive genomic signatures of rapid poleward range expansion. *Molecular Ecology*, *24*, 6163–6176. https://doi.org/10.1111/mec.13462

The Anopheles gambiae Genomes Consortium. (2017). Genetic diversity of the African malaria vector Anopheles gambiae. *Nature Reviews Genetics*, *552*(7683), 96–100. https://doi.org/10.1038/nature24995

Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, *194*(2), 301–326. https://doi.org/10.1534/genetics.112.148825

Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society*, *14*(2), 191–220. https://doi.org/10.1177/0957926503014002277

Vázquez-Miranda, H., Griffin, J. A., Sheppard, J. M., Herman, J. M., Rojas-Soto, O., & Zink, R. M. (2017). Morphological and molecular evolution and their consequences for conservation and taxonomy in the Le Conte's thrasher Toxostoma lecontei. *Journal of Avian Biology*, *48*(7), 941–954. https://doi.org/10.1111/jav.01057

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS discovery: Biology, function, and translation. *The American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wade, A. A., Hand, B. K., Kovach, R. P., Luikart, G., Whited, D. C., & Muhlfeld, C. C. (2016). Accounting for adaptive capacity and uncertainty in assessments of species' climate-change vulnerability. *Conservation Biology*, *31*(1), 136–149. https://doi.org/10.1111/cobi.12764

Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology*, *107*, 4–13. https://doi.org/10.1016/j.tpb.2015.08.006

Waples, R. S., & Anderson, E. C. (2017). Purging putative siblings from population genetic data sets: A cautionary view. *Molecular Ecology*, *26*(5), 1211–1224. https://doi.org/10.1111/mec.14022

Wellenreuther, M., & Otto, S. (2015). Women in evolution—highlighting the changing face of evolutionary biology. *Evolutionary Applications*, *9*(1), 3–16. https://doi.org/10.1111/eva.12343

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences*, *112*(17), 5360–5365. https://doi.org/10.1073/pnas.1418878112

Wu, T. D. (2001). Bioinformatics in the post-genomic era. *Trends in Biotechnology*, *19*(12), 479–480. https://doi.org/10.1016/S0167-7799(01)01807-8

Xue, Y., Prado-Martinez, J., Sudmant, P. H., Narasimhan, V., Ayub, Q., Szpak, M., … Scally, A. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*, *348*(6231), 242–245. https://doi.org/10.1126/science.aaa3952

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606