

Invited Review

## Big Data in Conservation Genomics: Boosting Skills, Hedging Bets, and Staying Current in the Field

**Rena M. Schweizer<sup>□</sup>, Norah Saarman, Kristina M. Ramstad, Brenna R. Forester, Joanna L. Kelley<sup>□</sup>, Brian K. Hand, Rachel L. Malison, Amanda S. Ackiss<sup>□</sup>, Mrinalini Watsa, Thomas C. Nelson, Albano Beja-Pereira, Robin S. Waples, W. Chris Funk, and Gordon Luikart**

From the Division of Biological Sciences, University of Montana, 32 Campus Drive, Missoula, MT 59812 (Schweizer, Hand, and Luikart); Department of Biology, Utah State University, Logan, UT (Saarman); Department of Biology and Geology, University of South Carolina Aiken, Aiken, SC (Ramstad); Department of Biology, Colorado State University, Fort Collins, CO (Forester); School of Biological Sciences, Washington State University, Pullman, WA (Kelley); Flathead Lake Biological Station, University of Montana, Polson, MT (Hand, Malison, and Luikart); Wisconsin Cooperative Fishery Research Unit, University of Wisconsin Stevens Point, Stevens Point, WI (Ackiss); San Diego Zoo Global, Escondido, CA (Watsa); Embark Veterinary, Inc., Boston, MA (Nelson); Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-UP), InBIO, Universidade do Porto, Vairão, Portugal (Beja-Pereira); DGAOT, Faculty of Sciences, University of Porto, Porto, Portugal (Beja-Pereira); Sustainable Agrifood Production Research Centre (GreenUPorto), Faculty of Sciences, University of Porto, Porto, Portugal (Beja-Pereira); Northwest Fisheries Science Center, NOAA Fisheries, Seattle, WA (Waples); and Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO (Funk).

Address correspondence to R. M. Schweizer at the address above, or e-mail: [rena.schweizer@umontana.edu](mailto:rena.schweizer@umontana.edu).

Received October 16, 2020; First decision November 10, 2020; Accepted April 13, 2021.

Corresponding Editor: Klaus-Peter Koepfli

### Abstract

A current challenge in the fields of evolutionary, ecological, and conservation genomics is balancing production of large-scale datasets with additional training often required to handle such datasets. Thus, there is an increasing need for conservation geneticists to continually learn and train to stay up-to-date through avenues such as symposia, meetings, and workshops. The ConGen meeting is a near-annual workshop that strives to guide participants in understanding population genetics principles, study design, data processing, analysis, interpretation, and applications to real-world conservation issues. Each year of ConGen gathers a diverse set of instructors, students, and resulting lectures, hands-on sessions, and discussions. Here, we summarize key lessons learned from the 2019 meeting and more recent updates to the field with a focus on big data in conservation genomics. First, we highlight classical and contemporary issues in study design that are especially relevant to working with big datasets, including the intricacies of data filtering. We next emphasize the importance of building analytical skills and simulating data, and how these skills have applications within and outside of conservation genetics careers. We also highlight recent technological advances and novel applications to conservation of wild populations. Finally, we provide data and recommendations to support ongoing efforts by ConGen organizers and instructors—and beyond—to increase participation

of underrepresented minorities in conservation and eco-evolutionary sciences. The future success of conservation genetics requires both continual training in handling big data and a diverse group of people and approaches to tackle key issues, including the global biodiversity-loss crisis.

**Subject area:** Conservation Genomics and Biodiversity

**Key words:** population genetics theory, ecological genomics, biodiversity conservation, experimental design, big data filtering, metabarcoding, bioinformatics, training workshops, career planning

When learning classic population genetics theory, we initially consider a single locus with 2 alleles (e.g., Wright 1951). The challenge and exciting promise of the field of conservation genomics is to scale our efforts up to thousands or millions of loci and multiple whole-genome sequences in order to address pressing issues of conservation concern. This escalation requires a set of analytical skills for processing big data that are not as straightforward as those required in decades past (e.g., Tanjo et al. 2020; McLeish et al. 2021). Thus, these big data advances necessitate ongoing learning and training for most conservation geneticists because the field is expanding so dynamically. Online courses, published literature, web forums, workshops, meetings, and seminars are all means to keep up to date.

The ConGen meeting (<https://www.umt.edu/ces/conferences/congen/default.php>) is one way to address the aforementioned challenges and includes training sessions with lectures from experienced instructors, hands-on exercises, and synergistic learning through discussions. From 2–7 September 2019, 36 students and 13 expert instructors gathered at the 11th ConGen meeting in Montana, United States, to consider the latest conceptual and bioinformatic challenges in conservation and population genomic studies. Many of these topics have been presented and summarized in previous reviews of ConGen meetings (Andrews and Luikart 2014; Benestan et al. 2016; Hendricks et al. 2018a; Stahlke et al. 2020), and we refer interested readers to those papers.

Here, we present advances in recent and ongoing issues identified at the 2019 meeting and, beyond that, focus on a primary theme of big data in conservation genetics. We guide readers through 5 topics that include 1) classical and modern considerations of study design, 2) considerations and consequences of data filtering, 3) the value of simulations, computational proficiency, and developing transferable skills, and 4) novel applications of recent technological advancements to conservation. In our fifth topic, we present data collected over several years of ConGen meetings that describe trends of gender representation and country-of-origin at the meeting itself, with goals and actions for further improving the participation of under-represented groups at future meetings and beyond.

## Topic 1: Considering Study Design in the Era of Big Data in Conservation Genetics

Population genetics theory and careful study design are fundamental to conducting informative genomic studies (Allendorf 2017); even the most cutting-edge genomic techniques cannot compensate for a poor study design or deficient understanding of theory. Furthermore, given that sequencing is still relatively expensive and samples in conservation studies may be precious, researchers might only have one opportunity to pursue a study, emphasizing the importance of careful planning. Identifying the type and scale of genomic data to

collect will depend on numerous factors, including your question, the project budget, the size and complexity of your study organism's genome, career goals, and the genomic and bioinformatic resources available for your focal species or a closely related species (Figure 1; Allendorf et al. 2010; Hohenlohe et al. 2018). In this section, we discuss both classic and contemporary issues related to devising a study in conservation genetics, with some special considerations for managing large, complex datasets.

A well-defined study question and hypothesis are critical to choosing among the numerous options of genomic techniques available in light of inherent trade-offs. In other words, given your scientific question, which genomic technique should you use? Are you interested in examining neutral or adaptive processes or some combination of both? Assessing neutral processes such as historical demography, admixture, migration, and/or current population structure might require only tens to hundreds of anonymous genome-wide markers, while detecting processes such as local adaptation, introgression, selective sweeps, and/or adaptive potential may require sequencing candidate adaptive loci, genotyping thousands of markers genome-wide, or novel high-throughput sequencing approaches (HTS; e.g., Schweizer et al. 2016; Hohenlohe et al. 2018; Luikart et al. 2018; Lim et al. 2021; Lovell et al. 2021).

Additional issues to consider include the desired density of markers across the genome (which is influenced by population genetic variation and research question), the number of individuals versus populations available or required to address the question (e.g., minimum sample sizes can vary dramatically by analytical technique), the availability of previously ascertained genomic resources, and access to computational resources, including bioinformatics expertise (Hohenlohe et al. 2018). For example, the size and complexity of the study organism's genome, along with linkage disequilibrium (LD: non-randomly associated loci) along chromosomes, will determine whether whole-genome sequencing (WGS) is required or if reduced representation sequencing will suffice (e.g., RADseq, Andrews et al. 2016). Understanding which factors affect genome complexity (e.g., repetitiveness, proportion coding/intronic) will also help decide which genomic approach is best suited for the target organism.

Many considerations go into the choice of which genotyping technique to use for your study. If your study requires WGS (see Allendorf et al. 2010; Hohenlohe et al. 2018), do you have access to a reference genome for your species or a closely related species? If not, you could produce an annotated assembly, on your own or with a commercial company such as Dovetail Genomics (<https://dovetailgenomics.com/>). Once you have decided how many samples are necessary for your study, consider what depth of sequencing coverage is required. Long-term population monitoring efforts, or other studies which might require consistent sequencing of the same loci across many individuals, are still feasible with reduced representation sequencing through



**Figure 1.** Factors and questions to ask oneself when designing a conservation genomics study. Researchers may have to balance cost, feasibility, genomic information, computational resources, availability of collaborators, sequencing services (e.g., commercial companies), and characteristics of the target species' life history while keeping the main goals of the study a priority. Genome characteristics to consider (if data are available) might include size, complexity, nucleotide diversity, and extent of linkage disequilibrium along chromosomes. Goals of the study might include those relevant to the ecology and evolution of a taxon, or even a goal to gain experience using high-throughput sequence data for more marketable skills.

an approach such as Rapture (Ali et al. 2016) or RADcap (Hoffberg et al. 2016). Finally, do you have access to the bioinformatics resources and expertise needed to process and analyze the resulting data? If not, open-source, web-based platforms such as Galaxy (<https://usegalaxy.org/>) may be useful for some computing. You can learn bioinformatics locally, take an online course, or collaborate with a bioinformatician. There are many options for free online courses, such as those through Coursera (<https://www.coursera.org/courses?query=bioinformatics>), edX (<https://www.edx.org/learn/bioinformatics>), or DataCamp (<https://www.datacamp.com/>).

There are multiple avenues for choice of sequencing technology (Table 1), data analysis, filtering, and more. Even at the early stages of study design, it is important to consider how data will be filtered and which computational methods will be used for analyses, so that factors such as sequencing effort (e.g., read

depth), number of individuals, and expected number of filtered SNPs can be included in cost estimates before project initiation. Aspects such as data filtering are discussed in more detail in the next section. We also recommend more specific references that discuss study design in RAD-seq (Andrews et al. 2016), targeted capture (Jones and Good 2016), RNA-seq (Todd et al. 2016), and WGS (Ekblom and Wolf 2014). Newcomers to the field of conservation genomics may especially appreciate the efforts of these authors to define common terms and jargon that may otherwise cause confusion.

## Topic 2: Navigating the Perils of Data Filtering

Gone are the days of hand-checking the quality of data; such practices would be impossible across thousands or millions of

**Table 1.** Summary of several approaches to obtaining genotypes, including what each method may measure, when it might be used, some potential drawbacks, and a few references for further study

Genotyping approach	What can best be measured?	Why or when to use it?	Key drawbacks	Ref.
Restriction site-associated DNA sequencing (RAD-seq)	Genetic diversity metrics ( $F_{ST}$ , $H_c$ ), individual inbreeding, relatedness, hybridization & introgression, DNA methylation for epigenetic studies (BsRAD-seq).	1) When first establishing genome resources for a species and/or large genome size, or no or poor genome reference, 2) low budget but need 1000s of loci, 3) to screen 1000s of loci to identify 100–5k informative loci ideal for your question.	Not as useful for measures of linkage disequilibrium (LD), local adaptation (if LD is low), or variation in coding regions. Data filtering greatly influences downstream population genetic inferences.	1–9
RAD capture	Same as above but for targeted markers discovered from a RAD-seq experiment.	For establishing longer term monitoring programs or subsequent research where many individuals will be genotyped (e.g., annually for monitoring).	Expensive initial investment for marker discovery with array design and purchase (but pays off if genotyping thousands of individuals with ~500–50 000 loci).	10–12
Targeted capture	Individual-based genetic diversity metrics, population-level allele frequencies, coding region variants, etc.	For sequencing or re-sequencing candidate genes or other regions, when high coverage for a subset of the genome, or repeated use of markers is needed.	Can be expensive to design and generate probes (but see ExCapSeq and EecSeq); need a reference sequence for probe design.	13–15
Whole-genome sequencing—low depth of coverage (<10X), including Pool-Seq	Population-level allele frequencies, with individuals barcoded or not (Pool-Seq)	When individual genotypes are not important, e.g., measuring population-level variation, genome-wide signatures of selection, identifying runs of homozygosity and inversions.	Expensive when genome size is large (e.g., >1.5 Gb), requires large sample sizes (30–50 at a minimum), Pool-Seq has no individual barcodes or genotypes.	16–21
Whole-genome sequencing—high depth of coverage (>10X)	Individual genotypes with high genome contiguity and fidelity.	Many uses, including building reference genome, individual genotype-level analyses, and characterization of structural variants.	Cost prohibitive when reference genome size is large (e.g., >1.5 Gb) or complicated to sequence (e.g., highly repetitive, high heterozygosity).	22–23

Note that some methods (e.g., RNA-Seq, BsRAD-Seq, Methyl-Seq) are not discussed here. References: 1) Miller et al. 2007; 2) Baird et al. 2008; 3) Hohenlohe et al. 2010; 4) Hoffman et al. 2014; 5) Andrews et al. 2016; 6) Kovach et al. 2016; 7) McKinney et al. 2017; 8) Shafer et al. 2017; 9) Marconi et al. 2019; 10) Ali et al. 2016; 11) Hoffberg et al. 2016; 12) Kelson et al. 2020; 13) Jones & Good 2016; 14) Hendricks, et al. 2018b; 15) Puritz & Lotterhos 2018; 16) Ekblom & Wolf 2014; 17) Therikildsen & Palumbi 2017; 18) Kofler et al. 2011; 19) Schlötterer, et al. 2014; 20) Kardos et al. 2015; 21) Micheletti & Narum 2018; 22) Koepfli et al. 2019; and 23) Wright et al. 2020.

SNPs. Thus, a major challenge of working with big data sets for conservation genetics is deciding on adequate but not too stringent filtering of SNPs. In this section, we discuss several aspects of filtering data that should be carefully considered while planning an experiment, as well as during quality control and at the analytical stages.

### Why and How to Filter SNP Data?

Commonly, SNP data are filtered to achieve the following goals: 1) to improve reliability of genotype data, and 2) to reduce correlations of information content (lack of independence) across loci. There are technical solutions that help address some of these goals (Ali et al. 2016; O'Leary et al. 2018). For example, using Unique Molecular Identifiers (UMI) to remove PCR duplicates can help to improve accuracy of heterozygote calls (Aird et al. 2011; Krehenwinkel et al. 2017; Euclide et al. 2020), randomizing samples among libraries and including technical replicates can minimize the effects of sequencing errors (O'Leary et al. 2018), and use of target enrichment (i.e., probes or biotinylated adaptors) can improve sequence quality (Souza et al. 2017; Rochette et al. 2019). However, technical solutions are constantly being updated and discussed in the literature (Aird et al. 2011; Krehenwinkel et al. 2017; Euclide et al. 2020), leaving an inevitably important role for data filtering in any population genomics project. Understanding the goals, pros,

and cons of each filtering step can help one make up-to-date choices of appropriate filters.

### Reliability

A major goal of data filtering is to avoid using data that do not reflect true genotypes. SNP genotypes, especially when well-validated, can be more repeatable and more easily standardized than microsatellites or other methods (Morin et al. 2004). However, SNP calls are not infallible, notably when called from high-throughput sequencing data. There are at least 3 main sources of unreliable genotypes. First, the genetic locus might not follow Mendelian segregation or Hardy-Weinberg (HW) proportions if, for example, it is a restriction or amplified fragment length polymorphism (e.g., where some individuals lack the restriction site or have allelic dropout), or if the locus represents one of multiple pseudogenes or repetitive elements (Vuylsteke et al. 2007). Waples (2015) covers the history of filtering to address this problem in his review of filtering for HW proportions, and others have reviewed the applications and pitfalls of applying such a filter to large genomics-scale datasets (O'Leary et al. 2018; Meisner and Albrechtsen 2019).

Second, preparation of DNA for high-throughput sequencing for SNP genotyping (e.g., RADseq, targeted DNA capture) introduces sources of error such as inconsistent sequencing of loci, variance in coverage, null alleles, and PCR artifacts. The effects and their

severity may vary with each library preparation (Korneliussen et al. 2014; O'Leary et al. 2018).

Third, bioinformatics pipelines and filtering choices can introduce biases and errors of their own, such as alignment clustering errors that can cause artefactual contigs, which in turn influence variant detection and genotyping (Shafer et al. 2017). For example, bioinformatic approaches that do not identify and remove duplicated sequences (i.e., reads from paralogs that are treated as the same locus) can produce artefactual contigs, introducing error by artificially increasing heterozygosity (McKinney et al. 2018). Such filtering errors can lead to faulty conclusions. For example, Larson et al. (2021) reanalyzed published data and found that incomplete bioinformatic filtering could cause erroneous conclusions that the harvesting of fish populations drove a rapid reduction in body size of walleye.

Across software programs, there could be differences in variant detection from the same sequence data due to variation in the underlying methods that account for and correct sequencing errors (Baes et al. 2014; Hwang et al. 2015; Mielczarek and Szyda 2016; Wright et al. 2019), and to a lesser degree, differences in genotype calls (Bresadola et al. 2020). In some cases, it is possible to assign probabilities to SNP calls, and these can be used as grounds for filtering based on confidence in the true underlying genotype. Alternatively, one can deal directly with the probabilities rather than called genotypes, which allows one to incorporate and account for statistical uncertainty in downstream analysis (e.g., ANGSD, Korneliussen et al. 2014). This approach has merit, but can preclude the use of software designed to handle “hard” genotype calls, so is only useful when downstream analysis software is built to work with genotype likelihoods.

Finally, single-occurrence alleles (singletons) are common and can be the majority of the genotyped loci, but can also represent a combination of genotyping and sequencing errors (Hotaling et al. 2018). Obviously, it is desirable to remove singletons that are a result of technical errors. However, true rare alleles can also provide useful information on fine-scale gene flow, inference of demographic history, and local adaptation (Gravel et al. 2011; O'Leary et al. 2018). Because it can be difficult to differentiate error from truth, scientists often implement a minor allele frequency (MAF) or minor allele count (MAC) filter. It is advised to test out multiple filtering thresholds before settling on one or two for subsequent data analysis, depending on the specific study (Shafer et al. 2017; Hendricks et al. 2018a).

### Independence

Many genetic analyses assume that loci and individuals represent independent samples of genetic information. However, statistical independence can be violated both by sampling of loci that are in LD and by non-representative sampling of closely related individuals. Loci that are physically close together on the same chromosome are likely non-randomly associated and do not provide independent information. Closely related individuals also may not provide accurate and independent information about population-level genetic processes such as demography, gene flow, or selection. A number of tests and models assume that loci in close physical linkage or genotypic disequilibrium are removed from the dataset. For example, the program LDNe assumes independent loci for calculation of effective population size (Waples and Do 2008), and GENECLASS assumes independence of individuals for detection of first-generation migrants (Piry et al. 2004). Characterizing linkage among loci can

also help identify genomic regions undergoing positive selection, so whether the independence of loci and individuals becomes a problem will depend on the goals of a particular study. A recent study from Waples and colleagues (Waples et al. 2020, bioRxiv) quantified pseudoreplication caused by LD in genomic-scale datasets. They showed that the marginal benefits to precision of adding more loci decline very quickly for estimating  $N_e$  via the LDNe method, and decline more slowly for estimating  $F_{ST}$ . In both cases, the true confidence intervals for large datasets are often much wider than is computed using current methods, which assume all loci (or pairs of loci) are independent. Studies such as those of Waples et al. (2020, bioRxiv) are useful for planning how many loci are needed when designing a genomics project.

### Striking a Balance Between Over- and Under-Filtering Genomic Data

Understanding the goals for filtering and optimizing filtering approaches for a specific dataset or question are continually evolving challenges of working with big data in conservation and eco-evolutionary genomics. Once armed with an understanding of those challenges, one attempts to maximize the reliability of the dataset, that is, remove all erroneous data and retain all authentic data. However, there are no infallible ways of distinguishing between the two. For example, HW screening can be a useful way of detecting data errors, but comes with the risk of removing a true biological signal—including selection signals at outlier loci (Waples et al. 2015; Meisner and Albrechtsen 2019). Similarly, the stringency of LD filtering (e.g., the magnitude of correlation and/or window size used when removing loci to determine which loci remain in downstream analysis) can be somewhat subjective, and the method will always accept some level of data loss in favor of removing redundant or erroneous data. MAF or MAC filters represent a continuous spectrum of possible screening stringency, and appropriate criteria will be case-specific (Hotaling et al. 2018; Linck and Battey 2019).

When measuring relatedness, it is not sufficient to identify all the possible relationship categories; one also wants to know whether close relatives appear in the sample more often than they would by chance. Even for random samples, removing close relatives reduces sample size and hence precision, so this has to be balanced against potential reductions in bias from removing relatives. For example, genetic indices of allele frequency, population differentiation, and effective population size are less precise when siblings are removed (Waples and Anderson 2017). Furthermore, filtering too conservatively can reduce sample sizes to the point where they no longer answer your questions because of low statistical power. This problem can be assessed and mitigated by step-wise filtering, for example, of missing data (Hotaling et al. 2018; O'Leary et al. 2018).

Striking a satisfactory balance between over- and under-filtering can be so “freaking” difficult that filtering has been called the “F-word” (Andrews and Luikart 2014; J. Seeb pers. comm.). Prior to initiating a new HTS-based project, we recommend consideration of potential filtering strategies since some steps of data filtering are influenced by library preparation, number of samples, and sampling design, while others can be avoided with technical modifications prior to sequencing (see beginning of Topic 2). For example, increasing the amount of starting DNA or reducing the number of PCR cycles may diminish the risk of sequencing (then having to remove) PCR duplicates. Likewise, any prior knowledge of relatedness amongst individuals could be used to pick those that are least related, if appropriate for the project goals. In sum, we advise designing your

data collection efforts to minimize downstream biases and maximize the potential to solve outstanding issues in conservation.

### Topic 3: Building Your Skills and Hedging Your Bets

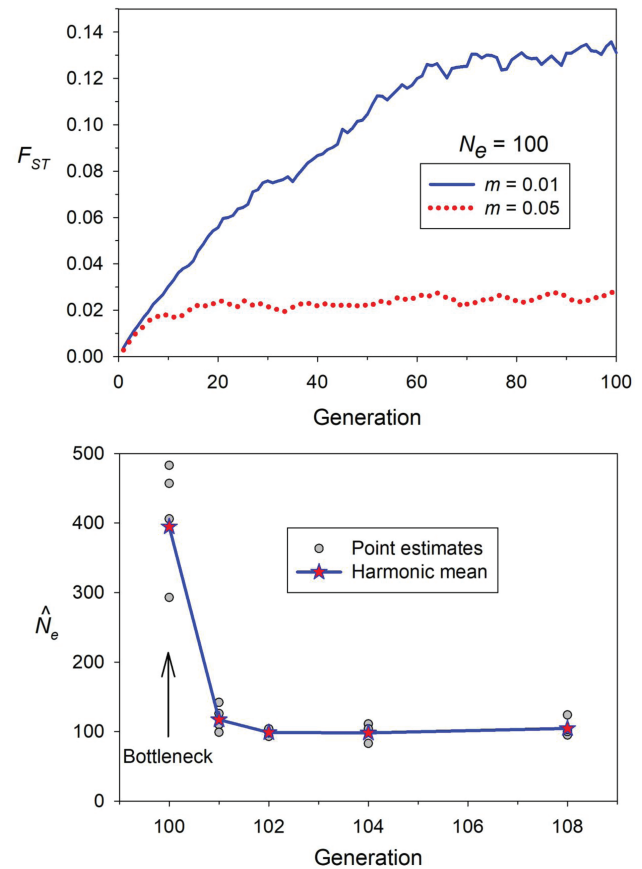
Modern conservation genomics and big data analysis are aided by competence in a variety of fields, including bioinformatics, population genetics theory, and molecular biology. As we describe in this section, building the skills to manage, simulate, and analyze large data sets will be useful for applications in academia, applied management, and beyond.

### The Value of Simulation Modeling in Conservation Genetics

Among the most popular activities at each ConGen is Robin Waples' simulation mini-project (Andrews and Luikart 2014). In small groups, participants use the program EasyPop (Balloux 2001) to simulate data and investigate the consequences of population sizes, migration rates, bottlenecks, mating systems, and divergence time on genetic diversity within populations and genetic differentiation among populations. Research questions addressed with simulation might include: How do precision and bias differ for microsatellites and SNPs when using a given estimator or software (e.g., STRUCTURE, LDNe, BayesAss)? What are the relative benefits of sampling more individuals versus sampling more SNPs? How long does it take before a change in population size (e.g., bottleneck) can be detected with single-sample and temporal (2-sample) estimators of effective population size?

Small groups simulate data, analyze them, and prepare presentations on their results, all in less than 24 hours (Figure 2). These intense, hands-on efforts not only allow participants to investigate the complex effects of population demography on estimating genetic diversity, but to also explore the power of relatively simple simulations to address consequential questions in population and conservation genetics. Many groups learn how their questions can grow into large, factorial simulation study designs that quickly expand beyond the allotted time. This is an important primary lesson of simulation modeling: choose clearly defined questions and specific parameters of interest, because you will almost always have more questions that you will want to address once you get started.

Authors Ackiss and Watsa, participants at ConGen 2019, stress the value of both the hands-on analyses and the accompanying thought exercise. Designing simulations requires careful consideration of the parameters that can be manipulated (e.g., sex ratios, mutation rate, migration model), which can be daunting when attempting to model complex population dynamics. Although to some degree simulations require oversimplifications of the natural systems being modeled, the process of designing and interpreting simulations also encourages deeper consideration of a study species than often encountered in standard population genetics analyses. For example, predicting the time it will take to see measurable effects from a disturbance (e.g., population fragmentation, bottleneck) requires an estimate of generation length and a clear understanding of reproductive strategy, often gleaned from life history studies of the target species in a natural setting. Even an oversimplified model using discrete generations can provide an informative comparison to empirical data from populations exhibiting overlapping generations when the effects of these differences are considered (Waples et al. 2014).



**Figure 2.** Results generated by ConGen 2019 participants in the simulation exercises using EasyPop. Top (from the “Leopard” student group): Changes in  $F_{ST}$  for 4 populations, each of  $N_e = 100$ , that begin with identical allele frequencies at generation 0 and then diverge with island-model migration rates of 1% or 5% per generation. Notable results: 1) Equilibrium  $F_{ST}$  is reached much faster at the higher migration rate. 2) Even when data are averaged over 1000 diallelic (SNP) loci assayed for all individuals, demographic stochasticity leads to considerable generation-to-generation variance in  $F_{ST}$ . Bottom (from the “Sparrow” group): Sensitivity of estimates of  $N_e$  (LDNe method) to detect a population bottleneck. At generation 100, a panmictic population of  $N_e = 400$  is fragmented into 4 isolated subpopulations of  $N_e = 100$ . In generations 101, 102, 104, 108,  $N_e$  is estimated for each subpopulation using data for 100 diallelic (SNP) loci assayed for all 100 individuals. A single generation after the bottleneck, harmonic mean  $N_e$  (117) is much closer to the reduced bottleneck size than the pre-bottleneck  $N_e$ .

Robin Waples' ConGen exercise illustrates why simulation modeling is such a valuable skill to learn as a conservation or population geneticist. However, diving into simulations for the first time can be intimidating. Developing a familiarity with the available program options is a good first step (see Hoban 2014, and the frequently updated Genetic Simulation Resources catalog provided by the National Institute of Health (<https://popmodels.cancercontrol.cancer.gov/gsr/>)). There is no one ideal simulation program, there is only the program(s) best suited to address your question. Some programs are simple and easy to learn, while others provide more sophisticated functionality, but have a steeper learning curve. Reading papers, talking with colleagues, and honing your questions and hypotheses are all great ways to narrow down the options. The payoff for investing time in becoming proficient with one or more simulation programs is the ability to address myriad questions of consequence in conservation genetics.

For example, simulations can be used to test new and existing population genetic methods, providing a valuable service to the community by evaluating best practices, settings, recommended uses, and comparing different approaches (e.g., [Evanno et al. 2005](#); [Lotterhos and Whitlock 2014, 2015](#); [Waples et al. 2016](#); [Zeller et al. 2016](#); [Forester et al. 2018](#); [Linck and Battey 2019](#); [Battey et al. 2020](#); [Waples et al. 2020](#), bioRxiv; [Allendorf et al. 2022](#)). Simulations are also useful for designing appropriate sampling schemes in conservation genetics research or monitoring and optimizing the value of limited conservation research funds (e.g., [Hoban et al. 2013](#); [Smith and Wang 2014](#); [Blower et al. 2019](#); [Selmoni et al. 2020](#); [Luikart et al. 2021](#)). For example, Waples and colleagues ([Waples et al. 2020](#), bioRxiv) used simulations to systematically quantify the effects of non-independence amongst loci on overall information content. They found that if you have  $X$  total loci, after accounting for linkage you have the same information content as you would with  $Y$  completely independent loci (with the value of  $Y$  depending on covariates such as genome size, true effective size, and number of individuals sampled). So, if you simulate  $Y$  unlinked loci, it should approximate the precision you can expect with  $X$  total loci, and provide insight into how to design a useful sampling scheme.

Simulations can also be a powerful way to corroborate empirical results and inform downstream management actions ([Landguth et al. 2017](#); [Thatte et al. 2018](#); [Grueber et al. 2019](#); [Hoban 2019](#); [Rougemont et al. 2019](#), [Ackiss et al. 2020](#); [Antão et al. 2020](#)). In addition, simulation data can be generated, analyzed, and written up from the office or home, providing a great backup for field or laboratory-based research that may be put on hold (e.g., due to the COVID-19 pandemic). Finally, simulations can greatly advance our understanding of a parameter's behavior in certain biologically-relevant scenarios, thereby allowing many biologists to improve their work ([Kardos and Luikart 2021](#)).

### Other Computational Skills to Increase Efficiency and Transferability to Other Careers

Ever-increasing amounts of data, whether genome-scale sequencing data or simulated data, necessitate a corresponding computational skillset. Computational skills, such as familiarity with shell ([https://linuxcommand.org/lc3\\_learning\\_the\\_shell.php](https://linuxcommand.org/lc3_learning_the_shell.php)), R (<https://www.r-project.org/>), Python ([www.python.org](http://www.python.org)) or another scripting language, and the ability to move seamlessly between a Unix environment and Windows or Mac environment, facilitate the ease with which data can be managed, parsed, and ultimately analyzed. Indeed, it has been argued that “all biology is computational biology” ([Markowitz 2017](#)) and as datasets grow, this is increasingly true. Moreover, as remote work becomes more common and we make sense of the new normal that has occurred since the start of the global COVID-19 pandemic, investing time in developing computational skills will improve the speed, reproducibility, and utility of scientific pursuits ([Carey and Papin 2018](#)). For example, tools such as Rmarkdown (see below), which integrate across code blocks and formatted annotations, can be used to improve the reproducibility of your research by providing details of each analytical step, from the raw data to figures in the paper. Furthermore, most scientific software is now written in one of a few languages, most of which are introduced at ConGen, so understanding and navigating these languages will help within conservation genetics and beyond.

Another advantage of developing computational and reproducible research skills is that many careers outside of traditional science paths recommend or require them. There are lucrative data scientist

positions in research and industry that biologists are well suited for because we are trained to analyze and make sense of complex datasets. Organizational skills that are the bedrock of reproducible computation, such as maintaining an organized workspace and keeping a corresponding computational notebook, demonstrate important know-how to future employers. Computational notebooks often make use of a language called “Markdown,” and there are well-maintained options for Python (Project Jupyter: <https://daringfireball.net/projects/markdown/>; Google Collab: <https://colab.research.google.com>) and R (RMarkdown: <https://rmarkdown.rstudio.com/index.html>). One great resource for organizing computational biology projects is [Noble \(2009\)](#). Even if you are not yet in a position to execute analyses yourself, knowledge of how and why specific computational approaches are taken is essential for communication among a team ([Carey et al. 2019](#)). Specific computational skills that may be beneficial and transferable to other careers include: 1) data management and processing, 2) data analysis, 3) knowledge of a scripting language (e.g., Python), 4) version control (e.g., with github, <https://github.com>), 5) statistical computing (e.g., R, MatLab, or SAS), 6) data visualization, and 7) communication ([Hampton et al. 2017](#)).

### Consideration of Backup Plans, Both for Data Analysis and Your Career

Backup plans for obtaining samples and analyzing data are critical parts of any initial study design. Even the best laid study plans can be derailed by unforeseen circumstances, such as years of low abundance for your study populations, an inability to get sampling permissions or international import/export permits for tissues, or a sudden loss of access to computing power, study populations or bioinformatic expertise. Alternative sources of samples and data include accessing museum specimens or tissue biobanks ([Buerki and Baker 2016](#)). Another source of data is the mining of publicly available genetic sequences from curated sources such as NCBI's Short Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>), the Genomes OnLine Database (<https://gold.jgi.doe.gov/>), or EvolMarkers ([Li et al. 2012](#); <http://bioinformatics.unl.edu/cli/evolmarkers/index.html>). As we detailed at the beginning of this section, conducting a computer simulation study also provides an alternative source of data for a dissertation chapter and influential publication to advance your career and your discipline.

With regard to data analyses, there are numerous secure networks for high-performance computing available that allow movement and processing of terabyte-scale data ([Langmead and Nellore 2018](#)) as well as commercial consultants (e.g., Duke Center for Genomic and Computational Biology (<https://genome.duke.edu/cores-and-services/genomic-analysis-and-bioinformatics/>), Taxa Genomics (<https://www.taxagenomics.com/>), and Bioinformatics Consultants (<https://www.bioinformaticsconsultants.com/>) that can provide expert assistance. In addition to potentially saving a lost field season or study, using these approaches and resources can enhance your efficiency, skill set, and even your scientific and professional networks.

Just as study designs need backup options, career plans do as well; the curriculum at ConGen is designed for students to practice transferrable skills and provide them with knowledge and resources to be successful outside of their current focal research. Developing teaching and scientific communication skills will increase your work's visibility and your employment options. For example, faculty that can effectively teach introductory genomics and bioinformatics

are in high demand (Campion et al. 2019; Goodman et al. 2020). Adequately communicating the value, potential, and limitations of genomics to resource managers and the public is also important. You can gain these experiences by offering your skills to regional natural resource managers and by volunteering to present your work to local community groups, museums, zoos, or schools. You may learn that you enjoy and have a talent for teaching and outreach, opening up the possibility of working in conservation management or education.

In this section, we have suggested skills to learn and strengthen, such as the ability to simulate data, work efficiently with large data sets, and plan for backup research projects, data sources, and careers. Students at ConGen practice these skills throughout the course, and leave more suited to obtain their research and career potential.

#### Topic 4: Emerging Technologies and Applications to Conservation Genetics

Certainly, a key driver of the challenges of working with big data in conservation is the continual release of newer, more powerful, and often cheaper technologies. In this section, we detail some recent technological advances, with a focus on how those technologies have been or could be used in conservation-specific applications.

##### Reference Genomes and Detection of Structural Variants

Long-read sequencing platforms like Pacific Biosciences (PacBio) and Oxford Nanopore Technology, along with long insert library preps (Chicago and Dovetail Hi-C), are quickly improving the quality and quantity of reference genomes available in conservation and population genomics. Over about 10 years, PacBio sequencing platforms have changed 3 times from the original PacBio RS I in 2011, to the RS II, to the Sequel 1 in 2015, and Sequel II in 2019 (<https://www.pacb.com/products-and-services/sequel-system/>). The Sequel II has decreased the costs of sequencing approximately 10-fold over the RS II platform while dramatically increasing accuracy, opening the way for conservation genomics research, which often lags behind in funding support. Meanwhile, Oxford Nanopore sequencers have increased the ability to sequence long strands of DNA outside of a lab through development of pocket-size sequencing platforms like the MinION in-the-field (reviewed in Krehenwinkel et al. 2019; <https://nanoporetech.com/products/minion>). To take advantage of long-read technologies, recent genome assembly algorithms (e.g., *Flye*: Kolmogorov et al. 2019; *Redbean*: Ruan and Li 2019) have been designed specifically for these types of data and also focus on rapid assembly compared to the most commonly used long-read assembly programs (e.g., *FALCON*: Chin et al. 2016; *CANU*: Koren et al. 2017).

Genome assembly has also received a further boost from commercial-based genome assembly services that can perform all steps—including sample preparation, sequencing, assembly, and gene annotation—often producing highly contiguous assemblies using long-read technologies (Armstrong et al. 2020; Nong et al. 2020). Cost for a chromosome-level assembly (including all bioinformatics) is \$15 000 to \$20 000 for many bird or mammal species with genome sizes of 1 to 3 Gb. Draft genomes can be produced for less than \$5000, or even less if one opts to do the assembly oneself after purchasing a sequencer.

Although improvements in HTS technologies have many benefits to conservation genomics, one major advantage of a higher quality genome assembly is the potential to detect structural

variants that occur in the form of insertion/deletions, copy number variants, and inversions (Hohenlohe et al. 2018). Studies that use mainly low depth of coverage WGS (Table 1) can still benefit from identifying structural variants that are likely to also be under selection (Wellenreuther et al. 2019), or infer haplotype information for historical demography and selection (Leitwein et al. 2020). However, accurate identification of structural variants is still a relatively new practice in many species and requires careful consideration and possibly the application of multiple tools (Kuzniar et al. 2020). Despite being a relatively new practice, many cutting-edge examples exist for reference (e.g., Special Issue, *Molecular Ecology*, 2019; Tigano et al. 2020, bioRxiv).

##### Genomics for Informative Marker Sets

The use of low-cost genotyping methods such as GT-seq and Rapture is continuing to revolutionize the field of conservation genomics (Meek and Larson 2019). This is due to declining sequencing expenses and creative, economical methods of preparing libraries and targeting loci. As a result, conservation biologists can target thousands of loci even without a reference genome and can use these genotyping panels to assess a variety of fundamental and applied questions (see Topic 1 and Allendorf et al. 2010).

Targeted amplicon panels such as GT-seq panels can incorporate both previously-established genetic resources, including microsatellite panels (Bradbury et al. 2018; Gruenthal and Larson 2021) and TaqMan qPCR assays (McKinney et al. 2020), and novel genomic data from RAD-seq, RNA-seq or low-coverage WGS (e.g., Bootsma et al. 2020; Schmidt et al. 2020). This asset allows conservation geneticists to continue long-term monitoring initiatives while leveraging the capabilities of recent genomic advances to target the most informative loci for resource management. These panels provide a cost-effective means to survey taxa at the individual and population level (Campbell et al. 2015; Meek and Larson 2019), are effective on low-quality DNA from non-invasive samples such as hair and feces (Natesh et al. 2019; Eriksson et al. 2020), and can be designed to supply the most pertinent information for the system of interest, including the presence of adaptive differences, sex determination, and stock or ecotype identification. As new genomic resources are developed, it is relatively easy to incorporate new amplicon loci into existing panels either directly or via pooling prior to sequencing. Amplicon panels also offer the added benefit of microhaplotypes (multiple SNPs at a locus treated as a single haplotype), which can substantially increase the power of genetic stock identification (McKinney et al. 2017; McKinney et al. 2020) and relationship inference (Baetscher et al. 2018).

The utility of microhaplotypes was briefly highlighted in the ConGen 2017 summary (Hendricks et al. 2018a), and since then an increasing number of studies have incorporated the analysis of multi-SNP loci. For example, Baetscher et al. (2019) used parentage analysis with 96 microhaplotype markers to examine the dispersal of rockfish offspring within and around marine reserves and conservation areas. Additionally, Reid et al. (2020) used a panel of 114 microhaplotype loci to assign ecotype ancestry and examine hybridization in a previously landlocked population of alewife after new fish passages restored access to the ocean after 300 years of isolation. Finally, Morin et al. (2021) illustrated the value of small numbers of microhaplotypes derived from degraded tissue samples in identifying population and stock structure in the North Pacific harbor porpoise, a nearshore species of conservation concern that is



difficult to sample in the wild (see also [Batz et al. 2020](#); [McKinney et al. 2020](#); [Bootsma et al. 2021](#)).

Pool-Seq, a cost-efficient approach to low-coverage WGS ([Table 1](#)), is also a tool of increasing use by the conservation genomics community and supported by community-built bioinformatic tools tailored to this type of data (e.g., *PoolParty*, [Micheletti and Narum 2018](#); *poolfstat*, [Hivert et al. 2018](#)). Careful consideration of sampling design is important when undertaking a pooled-sequencing project because, once pooled, individual-level genotype data are lost. It is essential to use replicate pools and sufficient sample sizes per pooled population (at least 30–50 individuals) for accurate estimation of allele frequencies ([Futschik and Schlötterer 2010](#); [Gautier et al. 2013](#); [Lynch et al. 2014](#); [Schlötterer et al. 2014](#)), in addition to following standard recommended sampling protocols for low-coverage WGS data such as maintaining equal sex ratios within populations ([Benestan et al. 2017](#)). When these factors are accounted for, this approach is a cost-efficient method to identify selective sweeps across the genome for multiple populations and to identify the genetic basis of important phenotypes and life-history traits ([Narum et al. 2018](#); [Chen and Narum 2021](#); [Horn et al. 2020](#)).

### Conservation Epigenetics

Epigenetics, and particularly DNA methylation studies, are a relatively underexplored aspect of conservation biology, yet may serve as a direct measure of an organism's response to its environment ([Rey et al. 2019](#)). Epigenetic markers can provide information on past and present stress caused by the environment, including current physiological condition ([Rey et al. 2019](#)). Additionally, epigenetic mechanisms can translate environmental selection pressures into heritable changes in phenotype ([Mukherjee et al. 2019](#)). However, assessing DNA methylation has previously required a high-quality genome and/or only surveying a subset of an organism's methylation profile via CG methylation ([Marconi et al. 2019](#)).

An exciting new approach is that of MCSed (Methylation Content Sensitive Enzyme ddRAD), which does not require a reference genome but surveys whole-genome methylation patterns in a cost-effective manner ([Marconi et al. 2019](#)). This type of approach could, for example, enhance previous studies examining the role of epigenetic mechanisms in rapid adaptation to new environments in species of conservation concern, such as Chinook salmon ([Venney et al. 2020](#)) and Darwin's finches ([McNew et al. 2017](#)).

There is also evidence that epigenetic mechanisms may be important in rapid evolutionary changes such as those involved in host-parasite coevolution ([Mukherjee et al. 2019](#)), and could provide solutions to managing wildlife diseases such as transmissible cancer in marsupials ([Ingles and Deakin 2015](#)). Finally, reduced sets of epigenetic markers are being developed to determine biological age clocks (reviewed in [Horvath and Raj 2018](#)). These tools may be invaluable in long-term monitoring of mammals for whom age cannot be easily determined.

### High-Throughput Approaches to Assess Wildlife Health

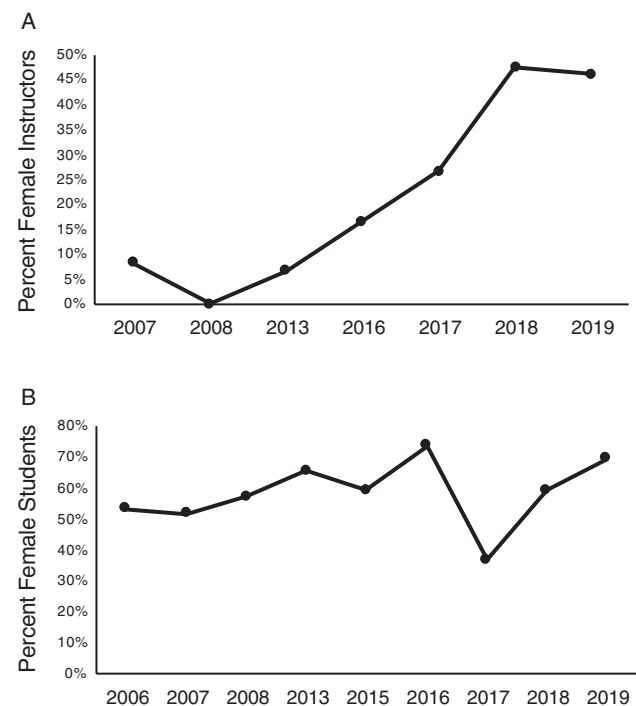
Wildlife health intersects with human health in many ways, brought into startling focus by the COVID-19 pandemic. Coronavirus, like many human pathogens, is thought to have emerged via a zoonotic spillover event ([Ye et al. 2020](#)). Yet, broad genomic screening for multiple wildlife diseases occurs less frequently ([Watsa et al. 2020](#)) than targeted approaches that track specific pathogens (e.g.,

*Batrachochytrium* pathogen in amphibians; [Farrer et al. 2017](#)). Metagenomic approaches have advantages of screening for multiple pathogens in less-studied or newly-identified systems. For instance, viral community diversity in vampire bats across the Americas varied not with colony size or inter-colony distance, but instead with elevational gradient and availability of anthropogenic food resources ([Bergner et al. 2020](#)). High-throughput sequencing can identify human-wildlife interfaces with increased contact, identify hotspots for pathogen transmission, and finally, assist in vector surveillance via DNA derived from invertebrate parasites (iDNA; [Kocher et al. 2017](#)) to screen diseases in hosts ([Titcomb et al. 2019](#)). Another relatively new tool for conservation biologists is VirScan. This system combines microarray-based immunoprecipitation with high-throughput sequencing to screen for large numbers of antibodies in very small quantities of blood ([Burbelo et al. 2019](#)), and is highly customizable for specific pathogen or host groups.

Metabarcoding and metagenomic approaches also provide exciting, powerful approaches to wildlife health. For example, metagenomics can be used to detect effects of stress, malnutrition, or starvation using noninvasively-collected fecal samples from wildlife ([Moustafa et al. 2021](#); [Yan et al. 2021](#)). Combined with creative uses of technology, such as unmanned aerial vehicles, metagenomic sampling can even be used to sample respiratory microbiomes ([Centelleghes et al. 2020](#)). We expect the continued application of cutting-edge approaches of technology, HTS, and big data to revolutionize studies of wildlife health.

### Staying Up to Date

Sequencing technologies, assembly algorithms, and genotyping software continue to change at a rapid pace, and it can be difficult



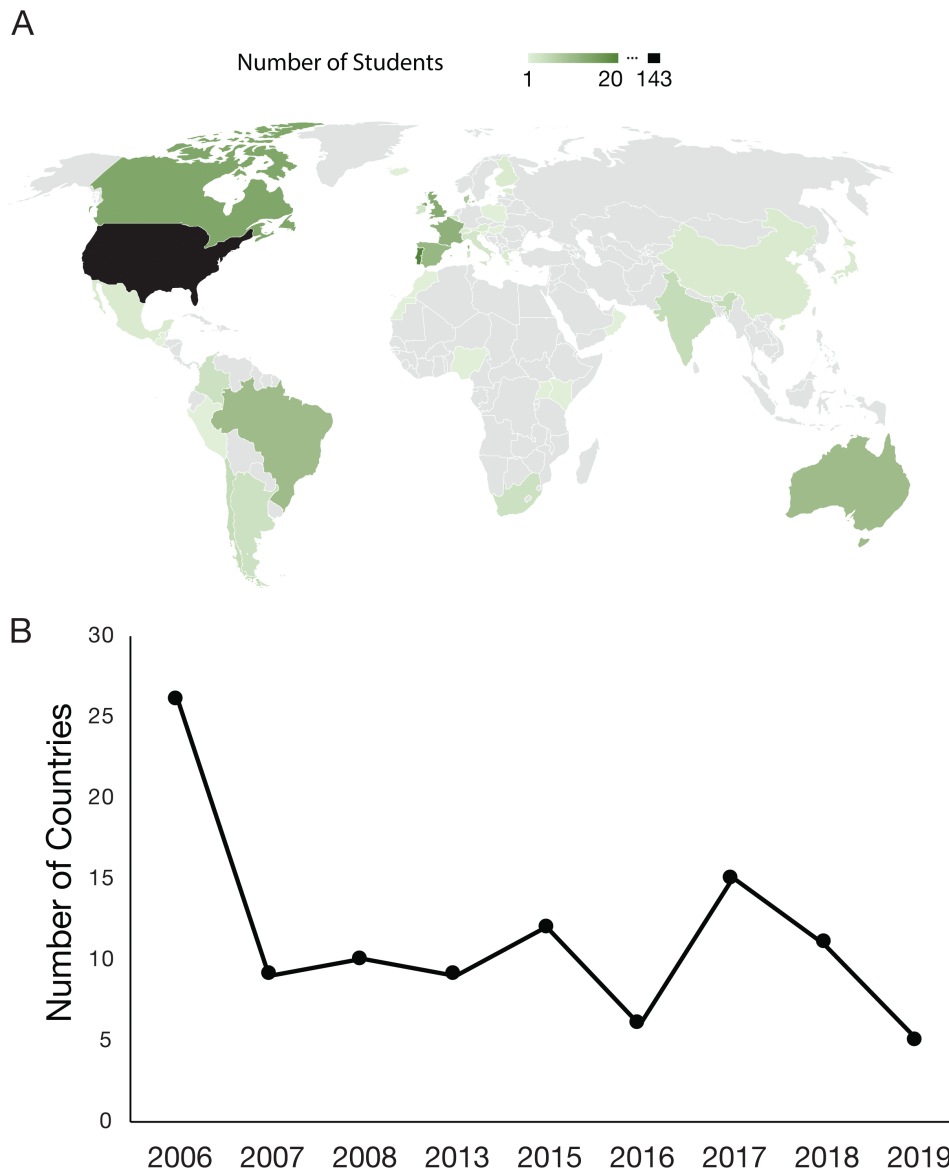
**Figure 3.** Data on participation of self-identifying females at 9 years of ConGen workshops. **A)** Percent female instructors over time. **B)** Percent female student participants over time. ConGen was not held in 2010, 2012, or 2014. Data are not available for some years when ConGen was held (A—2006, 2009, 2011 and 2015, B—2009, 2011).

to keep abreast of new developments. Luckily, there are several peer-reviewed resources for learning about these topics, such as *Annual Review of Genomics and Human Genetics*, *Nature Reviews Genetics*, *Biotechnology Advances*, and other journals (e.g., Hess et al. 2020; Schloss et al. 2020). Non-peer-reviewed options include development platforms such as GitHub (<https://github.com/>), forums such as SEQanswers (<http://seqanswers.com/>), and genomics-specific news outlets such as GenomeWeb (<https://www.genomeweb.com/sequencing>). Additionally, many annual conferences have sessions or booths that feature representatives from sequencing companies that may provide information on upcoming technological developments, such as the Plant and Animal Genome conference or ConGen

workshop. Equipped with these resources, scientists will continue to tackle pressing questions of conservation with cutting-edge big data approaches.

### Topic 5: Trends Towards Increasing Participant Diversity in Conservation Genomics

Scientists from different backgrounds offer an array of experiences, opinions, and insights, all of which result in increased performance (e.g., Hong and Page 2004) and increased representation of role models in the sciences (Jimenez et al. 2019). However, the fields of ecology and evolution continue to suffer



**Figure 4.** International representation of participants from 9 years of ConGen workshops. **A)** Countries shaded according to total number of participants. **B)** Total number of countries represented by student participants over time. ConGen was not held in 2010, 2012, or 2014. Data are not available for ConGen 2011. Note that in 2006, ConGen was held in Porto, Portugal, and participants' country of origin, not place of current employment, was recorded, and so had a larger number of countries represented than in other years when ConGen was held in Montana, United States. Country of student-participant origin is the country of current residence/employment. Many students residing in the United States are citizens of other countries, which may further diversify the number of countries beyond what is represented here.

from gender, ethnic, racial, and socioeconomic inequalities in academic and federal government positions (Arismendi and Penaluna 2016; Jimenez et al. 2019). Hendricks et al. (2018a) reviewed gender bias in the conservation genetics and population genomics fields, focusing on overcoming systematic biases against women, and others have highlighted key strategies to support Black, Indigenous, and people of color in the field (Tseng et al. 2020). Here we present data from ConGen workshops to investigate trends in gender identity of instructors and students, as well as the diversity of participants in career level and representation of students worldwide.

We gathered available information on participants and instructors from 9 years of ConGen courses held from 2006 to 2019. The participation of female instructors has increased during the 13 years of ConGen courses, rising from less than 10% representation in 2007 to almost 50% in 2018 and 2019 (Figure 3A). In contrast, self-identifying male and female students have been represented in roughly equal numbers since the beginning of ConGen (37–74% female, Figure 3B), and the overall mean is approximately 55% female. Students from 38 different countries have attended ConGen courses since 2006, with 5 to 26 countries being represented in any given year (from the 9 years with available data, Figures 3C and 4). Career level data for 2013–2019 show that the majority of ConGen participants are graduate students and postdocs (Figure 5). A recent increase in participants from state and federal agencies and non-profit organizations may reflect ConGen's emphasis on genomics methods and bioinformatics, which draws in career professionals who learned population genetics before the genomics revolution.

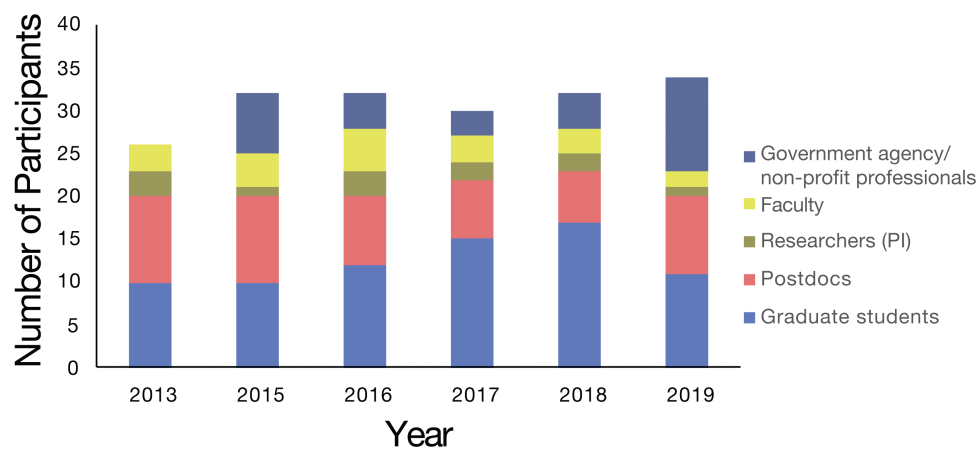
The faculty, researchers, and government agency employees who participate further increase the reach and impact of ConGen by sharing what they have learned with their students and colleagues. As one example, after teaching at ConGen in 2019, author Ramstad used the ConGen course as a model for a new undergraduate course in Population Genomics for the Biology and Geology department at University of South Carolina Aiken. USC Aiken has a highly diverse student body (30% underrepresented minorities, 65% female), with a large number of first-generation college students (at least 21%) and a high proportion of students (at least 24%) from low-income families. This new course at USC Aiken is a tangible outcome of how

instructors of ConGen may over time increase underrepresented minorities in conservation genomics.

While the trends for instructors and students at ConGen have improved, there is still room for improvement. For example, the ability of ConGen organizers to assess progress is limited by the data collected; there is still a lack of data on ethnicity, age, socioeconomic status, and disability status. Other areas of improvement include more targeted advertisement, outreach, and scholarships, which could diversify participation in future ConGen workshops, and other related courses.

## Conclusions and Outlook

The challenge in conservation genetics of scaling efforts from a few loci to millions of loci can require additional training in computer sciences, statistics, and population genetics theory. Symposia, meetings, and workshops are educational resources that may help fill this training gap. Here, we have summarized key updates to the field of conservation genomics and the necessary tasks for dealing with big data. We highlighted classical and contemporary issues in how study design should be carefully considered to align with research questions, and how to strike a balance between over- and under-filtering high-throughput sequencing data. Furthermore, we presented several key suggestions for building one's skills as a conservation genomicist, including learning how to simulate data, how to improve efficiency in multiple stages of the workflow using computational skills, and how to continually consider backup plans for data analysis and career. We presented updates on sequencing technologies, highlighting their applications to conservation biology, as well as how to stay informed on technological and methodological changes. Finally, we presented data collected during 9 years of ConGen courses, which show an increase in female participation at the instructor level that may reflect and reinforce the retention of women in conservation genetics. Looking forward, we hope that the ConGen workshop and others will continue to strive for excellence in training of the next generation of conservation biologists in conceptual and practical aspects of data analysis, while also ensuring those participants and their instructors represent a diverse array of backgrounds and perspectives which is increasingly needed to help curb the global conservation crisis.



**Figure 5.** Participants by career level, including graduate students, postdocs, researchers (principal investigators), faculty, and government agency/non-profit professionals. ConGen was not held in 2010, 2012, or 2014, and career level data are not available prior to 2013.

## Funding

Funding for ConGen was generously provided by a National Science Foundation Dimensions of Biodiversity Grant (#DOB-1639014) and the National Aeronautics and Space Administration (Grant #80NSSC19K0185), with additional support from the American Genetic Association. Funding support was provided to RMS by the National Science Foundation (#IOS-NSF 1755411).

## Acknowledgments

The authors are grateful to the other instructors, participants, and organizers of ConGen 2019 for enabling such a productive, insightful, and fun week. The authors thank William Murphy and an anonymous reviewer for helpful comments on earlier versions of the manuscript.

## Data Availability

Data for past ConGen participants and instructors are available upon request from the corresponding author.

## References

- Ackiss AS, Larson WA, Stott W. 2020. Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes. *Evol Appl.* 13:1037–1054.
- Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR. 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics.* 202:389–400.
- Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet.* 11:697–709.
- Allendorf FW. 2017. Genetics and the conservation of natural populations: allozymes to genomes. *Mol Ecol.* 26:420–430.
- Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G. 2022. *Conservation and the genomics of populations*. New York: Oxford University Press.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet.* 17:81–92.
- Andrews KR, Luikart G. 2014. Recent novel approaches for population genomics data analysis. *Mol Ecol.* 23:1661–1667.
- Antão T, Cosart T, Trethewey B, Waples RS, Ackerman MW, Luikart G, Hand BK. 2020. AgeStrucNb: software for simulating and detecting changes in the effective Number of Breeders (Nb). *J Hered.* 111:491–497.
- Arismendi I, Penaluna BE. 2016. Examining diversity inequities in fisheries science: a call to action. *BioScience* 66:584–591.
- Armstrong EE, Taylor RW, Miller DE, Kaelin CB, Barsh GS, Hadly EA, Petrov D. 2020. Long live the king: chromosome-level assembly of the lion (*Panthera leo*) using linked-read, Hi-C, and long-read data. *BMC Biol.* 18:3.
- Baes CF, Dolezal MA, Koltjes JE, Bapst B, Fritz-Waters E, Jansen S, Flury C, Signer-Hasler H, Stricker C, Fernando R, et al. 2014. Evaluation of variant identification methods for whole genome sequencing data in dairy cattle. *BMC Genomics.* 15:948.
- Baetscher DS, Anderson EC, Gilbert-Horvath EA, Malone DP, Saarman ET, Carr MH, Garza JC. 2019. Dispersal of a nearshore marine fish connects marine reserves and adjacent fished areas along an open coast. *Mol Ecol.* 28:1611–1623.
- Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC. 2018. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Mol Ecol Resour.* 18:296–305.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* 3:e3376.
- Balloux F. 2001. EASYPOP (version 1.7): a computer program for population genetics simulations. *J Hered.* 92:301–302.
- Batley CJ, Ralph PL, Kern AD. 2020. Space is the place: effects of continuous spatial structure on analysis of population genetic data. *Genetics.* 215:193–214.
- Batz ZA, Clemento AJ, Fritzenwanker J, Ring TJ, Garza JC, Armbruster PA. 2020. Rapid adaptive evolution of the diapause program during range expansion of an invasive mosquito. *Evolution.* 74:1451–1465.
- Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJ, Baums IB, Schwartz MK, Kelley JL, Luikart G. 2016. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol Ecol.* 25:2967–2977.
- Benestan L, Moore JS, Sutherland BJG, Le Luyer J, Maaroufi H, Rougeux C, Normandeau E, Rycroft N, Atema J, Harris LN, et al. 2017. Sex matters in massive parallel sequencing: evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Mol Ecol.* 26:6767–6783.
- Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, Biek R, Streicker DG. 2020. Demographic and environmental drivers of metagenomic viral diversity in vampire bats. *Mol Ecol.* 29:26–39.
- Blower DC, Riginos C, Ovenden JR. 2019. neogen: a tool to predict genetic effective population size (Ne) for species with generational overlap and to assist empirical Ne study design. *Mol Ecol Resour.* 19:260–271.
- Bootsma ML, Gruenthal KM, McKinney GJ, Simmons L, Miller L, Sass GG, Larson WA. 2020. A GT-seq panel for walleye (*Sander vitreus*) provides important insights for efficient development and implementation of amplicon panels in non-model organisms. *Mol Ecol Resour.* 20:1706–1722.
- Bootsma ML, Miller L, Sass GG, Euclide PT, Larson WA. 2021. The ghosts of propagation past: haplotype information clarifies the relative influence of stocking history and phylogeographic processes on contemporary population structure of walleye (*Sander vitreus*). *Evol Appl.* doi:10.1111/evo.13186.
- Bradbury IR, Wringe BF, Watson B, Paterson I, Horne J, Beiko R, Lehnert SJ, Clément M, Anderson EC, Jeffery NW, et al. 2018. Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. *Evol Appl.* 11:918–930.
- Bresadola L, Link V, Buerkle CA, Lexer C, Wegmann D. 2020. Estimating and accounting for genotyping errors in RAD-seq experiments. *Mol Ecol Resour.* 20:856–870.
- Buerki S, Baker WJ. 2016. Collections-based research in the genomic era. *Biol J Linn Soc.* 117:5–10.
- Burbelo PD, Iadarola MJ, Chaturvedi A. 2019. Emerging technologies for the detection of viral infections. *Future Virol.* 14:39–49.
- Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour.* 15:855–867.
- Campion M, Goldgar C, Hopkin RJ, Prows CA, Dasgupta S. 2019. Genomic education for the next generation of health-care providers. *Genet Med.* 21:2422–2430.
- Carey MA, Papin JA. 2018. Ten simple rules for biologists learning to program. *PLoS Comput Biol.* 14:e1005871.
- Carey CC, Ward NK, Farrell KJ, Lofton ME, Krinos AI, McClure RP, Kensworth C, Subratie KC, Figueiredo RJ, Doubek JP, et al. 2019. Enhancing collaboration between ecologists and computer scientists: lessons learned and recommendations forward. *Ecosphere.* 10:e02753.
- Centelleghè C, Carraro L, Gonzalvo J, Rosso M, Esposti E, Gili C, Bonato M, Pedrotti D, Cardazzo B, Povinelli M, et al. 2020. The use of Unmanned Aerial Vehicles (UAVs) to sample the blow microbiome of small cetaceans. *PLoS One.* 15:e0235537.
- Chen Z, Narum SR. 2021. Whole genome resequencing reveals genomic regions associated with thermal adaptation in redband trout. *Mol Ecol.* 30:1–13.
- Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* 13:1050–1054.
- Daniel A, Michael R, Wei-Sheng C, Maxwell D, Timothy F, Carsten R, Jaffe DB, Nusbaum C, Andreas G. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12:1–14.

- Eklblom R, Wolf JB. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl.* 7:1026–1042.
- Eriksson CE, Ruprecht J, Levi T. 2020. More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. *Mol Ecol Resour.* 20:1505–1516.
- Euclides PT, McKinney GJ, Bootsma M, Tarsa C, Meek MH, Larson WA. 2020. Attack of the PCR clones: rates of clonality have little effect on RAD-seq genotype calls. *Mol Ecol Resour.* 20:66–78.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14:2611–2620.
- Farrer RA, Martel A, Verbrugge E, Abouelleil A, Ducatelle R, Longcore JE, James TY, Pasmans F, Fisher MC, Cuomo CA. 2017. Genomic innovations linked to infection strategies across emerging pathogenic chytrid fungi. *Nat Commun.* 8:14742.
- Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol Ecol.* 27:2215–2233.
- Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics.* 186:207–218.
- Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A. 2013. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol.* 22:3766–3779.
- Goodman AL, Van Stry M, Lopatto D, Leung W, Reed LK, Community GF. 2020. Growing and sustaining a nationwide CURE: genomics education partnership enhances research opportunities for students and faculty at diverse institutions. *FASEB J.* 34:1.
- Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD; 1000 Genomes Project. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A.* 108:11983–11988.
- Grueber CE, Fox S, McLennan EA, Gooley RM, Pemberton D, Hogg CJ, Belov K. 2019. Complex problems need detailed solutions: harnessing multiple data types to inform genetic management in the wild. *Evol Appl.* 12:280–291.
- Gruenthal KM, Larson WA. 2021. Efficient genotyping with backwards compatibility: converting a legacy microsatellite panel for muskellunge (*Esox masquinongy*) to genotyping-by-sequencing chemistry. *Conservation Genet Resour.* doi:10.1007/s12686-020-01185-1
- Hampton SE, Jones MB, Wasser LA, Schildhauer MP, Supp SR, Brun J, Hernandez RR, Boettiger C, Collins SL, Gross LJ, et al. 2017. Skills and knowledge for data-intensive environmental research. *Bioscience.* 67:546–557.
- Hendricks S, Anderson EC, Antão T, Bernatchez L, Forester BR, Garner B, Hand BK, Hohenlohe PA, Kardos M, Koop B, et al. 2018a. Recent advances in conservation and population genomics data analysis. *Evol Appl.* 11:1197–1211.
- Hendricks SA, Schweizer RM, Harrigan RJ, Pollinger JP, Paquet PC, Darimont CT, Adams JR, Waits LP, vonHoldt BM, Hohenlohe PA, et al. 2018b. Natural re-colonization and admixture of wolves (*Canis lupus*) in the US Pacific Northwest: challenges for the protection and management of rare and endangered taxa. *Heredity* 26:1–17.
- Hess JF, Kohl TA, Kotrová M, Rönsch K, Paprotka T, Mohr V, Hutzenlaub T, Brüggemann M, Zengerle R, Niemann S, et al. 2020. Library preparation for next generation sequencing: a review of automation strategies. *Biotechnol Adv.* 41:107537.
- Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. 2018. Measuring genetic differentiation from Pool-seq data. *Genetics.* 210:315–330.
- Hoban S. 2014. An overview of the utility of population simulation software in molecular ecology. *Mol Ecol.* 23:2383–2401.
- Hoban S. 2019. New guidance for ex situ gene conservation: sampling realistic population systems and accounting for collection attrition. *Biol Conserv.* 235:199–208.
- Hoban S, Gaggiotti O, Bertorelle G. 2013. Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): a software for choosing the appropriate number of markers and samples. *Methods Ecol Evol.* 4:299–303.
- Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, Glenn TC. 2016. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Mol Ecol Resour.* 16:1264–1278.
- Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MA, Lacy RC, Dasmahapatra KK. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci U S A.* 111:3775–3780.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6:e1000862.
- Hohenlohe PA, Hand BK, Andrews KR, Luikart G. 2018. Population genomics provides key insights in ecology and evolution. In: Rajora O, editors. *Population genomics.* Cham, Switzerland: Springer.
- Hong L, Page SE. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc Natl Acad Sci U S A.* 101:16385–16389.
- Horn RL, Kamphaus C, Murdoch K, Narum SR. 2020. Detecting genomic variation underlying phenotypic characteristics of reintroduced Coho salmon (*Oncorhynchus kisutch*). *Conserv Genet.* 21:1011–1021.
- Horvath S, Raj K. 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet.* 19:371–384.
- Hotaling S, Muhlfeld CC, Giersch JJ, Ali OA, Jordan S, Miller MR, Luikart G, Weisrock DW. 2018. Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post-Pleistocene landscape. *J Biogeogr.* 45:304–317.
- Hwang S, Kim E, Lee I, Marcotte EM. 2015. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep.* 5:1–8.
- Ingles ED, Deakin JE. 2015. Global DNA Methylation patterns on marsupial and devil facial tumour chromosomes. *Mol Cytogenet.* 8:74.
- Jimenez MF, Lavery TM, Bombaci SP, Wilkins K, Bennett DE, Pejchar L. 2019. Underrepresented faculty play a disproportionate role in advancing diversity and inclusion. *Nat Ecol Evol.* 3:1030–1033.
- Jones MR, Good JM. 2016. Targeted capture in evolutionary and ecological genomics. *Mol Ecol.* 25:185–202.
- Kardos M, Luikart G, Bunch R, Dewey S, Edwards W, McWilliam S, Stephenson J, Allendorf FW, Hogg JT, Kijas J. 2015. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Mol Ecol.* 24:5616–5632.
- Kardos M, Luikart G. 2021. The genetic architecture of fitness drives population viability during rapid environmental change. *Am Nat* 197:511–525.
- Kelson SJ, Miller MR, Thompson TQ, O'Rourke SM, Carlson SM. 2020. Temporal dynamics of migration-linked genetic variation are driven by stream-flows and riverscape permeability. *Mol Ecol.* 29:870–885.
- Kocher A, de Thoisy B, Catzeflis F, Valière S, Bañals AL, Murielle J. 2017. iDNA screening: disease vectors as vertebrate samplers. *Mol Ecol.* 26:6478–6486.
- Koepfli K-P, Tamazian G, Wildt D, Dobrynin P, Kim C, Frandsen PB, Godinho R, Yurchenko AA, Komissarov A, Krashenninnikova K, et al. 2019. Whole genome sequencing and re-sequencing of the Sable Antelope (*Hippotragus niger*): a resource for monitoring diversity in ex situ and in situ populations. *Genes|Genomes|Genetics.* Doi:10.1534/g3.119.400084
- Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 27:3435–3436.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 37:540–546.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722–736.
- Korneliusson TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* 15:356.
- Kovach RP, Hand BK, Hohenlohe PA, Cosart TF, Boyer MC, Neville HH, Muhlfeld CC, Amish SJ, Carim K, Narum SR, et al. 2016. Vive la

- résistance: genome-wide selection against introduced alleles in invasive hybrid zones. *Proc R Soc Lond B Biol Sci.* 283:20161380–9.
- Krehenwinkel H, Pomerantz A, Prost S. 2019. Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. *Genes* 10:858.
- Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci Rep.* 7:17668.
- Kuzniar A, Maassen J, Verhoeven S, Santuari L, Shneider C, Kloosterman WP, de Ridder J. 2020. sv-callers: a highly portable parallel workflow for structural variant detection in whole-genome sequence data. *PeerJ.* 8:e8214.
- Landguth EL, Holden ZA, Mahalovich MF, Cushman SA. 2017. Using landscape genetics simulations for planting blister rust resistant whitebark pine in the US Northern Rocky Mountains. *Front Genet.* 8:9.
- Langmead B, Nellore A. 2018. Cloud computing for genomic data analysis and collaboration. *Nat Rev Genet.* 19:208–219.
- Larson WA, Isermann DA, Feiner ZS. 2021. Incomplete bioinformatic filtering and inadequate age and growth analysis lead to an incorrect inference of harvested-induced changes. *Evol Appl.* 14:278–289.
- Leitwein M, Duranton M, Rougemont Q, Gagnaire PA, Bernatchez L. 2020. Using haplotype information for conservation genomics. *Trends Ecol Evol.* 35:245–258.
- Li C, Riethoven JJ, Naylor GJ. 2012. EvolMarkers: a database for mining exon and intron markers for evolution, ecology and conservation studies. *Mol Ecol Resour.* 12:967–971.
- Lim M, Bi K, Witt CC, Graham CH, Dávalos LM. 2021. Pervasive genomic signatures of local adaptation to altitude across highland specialist Andean hummingbird populations. *J Hered.* 112:229–240.
- Linck E, Batten CJ. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol Ecol Resour.* 19:639–647.
- Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol.* 23:2178–2192.
- Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol.* 24:1031–1046.
- Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, et al. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature.* 590:438–444.
- Luikart G, Antao T, Hand BK, Muhlfeld CC, Boyer MC, Cosart T, Trethewey B, Al-Chokhachy R, Waples RS. 2021. Detecting population declines via monitoring the effective number of breeders (Nb). *Mol Ecol Resour.* 21:379–393.
- Luikart G, Hand BK, Kardos M, Rajora OP, Aitkin S, Hohenlohe P. 2018. Population genomics: advancing understanding of nature. In: Rajora Om P, editor. *Population genomics concepts, approaches and applications.* Switzerland: Springer Nature.
- Lynch M, Bost D, Wilson S, Maruki T, Harrison S. 2014. Population-genetic inference from pooled-sequencing data. *Genome Biol Evol.* 6:1210–1218.
- Marconi G, Capomaccio S, Comino C, Acquadro A, Portis E, Porceddu A, Albertini E. 2019. Methylation content sensitive enzyme ddRAD (MCSeEd): a reference-free, whole genome profiling system to address cytosine/adenine methylation changes. *Sci Rep.* 9:14864.
- Markowitz F. 2017. All biology is computational biology. *PLoS Biol.* 15:e2002050.
- McKinney GJ, Larson WA, Seeb LW, Seeb JE. 2017. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). *Mol Ecol Resour.* 17:356–361.
- McKinney GJ, Pascal CE, Templin WD, Gilk-Baumer SE, Dann TH, Seeb LW, Seeb JE. 2020. Dense SNP panels resolve closely related Chinook salmon populations. *Can J Fish Aquat Sci.* 77:451–461.
- McKinney GJ, Seeb JE, Seeb LW. 2017. Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. *Can J Fish Aquat Sci.* 74:429–434.
- McKinney GJ, Waples RK, Pascal CE, Seeb LW, Seeb JE. 2018. Resolving allele dosage in duplicated loci using genotyping-by-sequencing data: a path forward for population genetic analysis. *Mol Ecol Resour.* 18:570–579.
- McLeish MJ, Fraile A, García-Arenal F. 2021. Population genomics of plant viruses: the ecology and evolution of virus emergence. *Phytopathology.* 111:32–39.
- McNew SM, Beck D, Sadler-Riggelman I, Knutie SA, Koop JAH, Clayton DH, Skinner MK. 2017. Epigenetic variation between urban and rural populations of Darwin's finches. *BMC Evol Biol.* 17:183.
- Meek MH, Larson WA. 2019. The future is now: amplicon sequencing and sequence capture usher in the conservation genomics era. *Mol Ecol Resour.* 19:795–803.
- Meisner J, Albrechtsen A. 2019. Testing for Hardy-Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. *Mol Ecol Resour.* 19:1144–1152.
- Micheletti SJ, Narum SR. 2018. Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol Ecol Resour.* 18:825–837.
- Mielczarek M, Syzda J. 2016. Review of alignment and SNP calling algorithms for next-generation sequencing data. *J Appl Genet.* 57:71–79.
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17:240–248.
- Morin PA, Forester BR, Forney KA, Crossman CA, Hancock-Hanser BL, Robertson KM, Barrett-Lennard LG, Baird RW, Calambokidis J, Gearin P, et al. 2021. Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor-quality samples. *Mol Ecol.* 30:1457–1476.
- Morin PA, Luikart G, Wayne RK, the SNP Workshop Group. 2004. SNPs in ecology, evolution and conservation. *Trends Ecol Evol.* 19:208–216.
- Moustafa MAM, Chel HM, Thu MJ, Bawm S, Htun LL, Win MM, Oo ZM, Ohsawa N, Lahdenperä M, Mohamed WMA, et al. 2021. Anthropogenic interferences lead to gut microbiome dysbiosis in Asian elephants and may alter adaptation processes to surrounding environments. *Sci Rep.* 11:741.
- Mukherjee K, Dubovskiy I, Grizanova E, Lehmann R, Vilcinskis A. 2019. Epigenetic mechanisms mediate the experimental evolution of resistance against parasitic fungi in the greater wax moth *Galleria mellonella*. *Sci Rep.* 9:1626.
- Narum SR, Di Genova A, Micheletti SJ, Maass A. 2018. Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proc Biol Sci.* 285:20180935.
- Natesh M, Taylor RW, Truelove NK, Hadly EA, Palumbi SR, Petrov DA, Ramakrishnan U. 2019. Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods Ecol Evol.* 10:853–859.
- Noble WS. 2009. A quick guide to organizing computational biology projects. *PLoS Comput Biol.* 5:e1000424.
- Nong W, Law STS, Wong AYP, Baril T, Swale T, Chu LM, Hayward A, Lau DTW, Hui JHL. 2020. Chromosomal-level reference genome of the incense tree *Aquilaria sinensis*. *Mol Ecol Resour.* 20:971–979.
- O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018. These aren't the loci you're looking for: principles of effective SNP filtering for molecular ecologists. *Mol Ecol.* 27:3193–3206.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered.* 95:536–539.
- Puritz JB, Lotterhos KE. 2018. Expressed exome capture sequencing: a method for cost-effective exome sequencing for all organisms. *Mol Ecol Resour.* 18:1209–1222.
- Reid K, Carlos Garza J, Gephard SR, Caccone A, Post DM, Palkovacs EP. 2020. Restoration-mediated secondary contact leads to introgression of alewife ecotypes separated by a colonial-era dam. *Evol Appl.* 13:652–664.
- Rey O, Eizaguirre C, Angers B, Baltazar Soares M, Sagonas K, Prunier JG, Blanchet S. 2019. Linking epigenetics and biological conservation: towards a conservation epigenetics perspective. *Funct Ecol.* 34:414–427.

- Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol*. 28:4737–4754.
- Rougemont Q, Carrier A, Le Luyer J, Ferchaud AL, Farrell JM, Hatin D, Brodeur P, Bernatchez L. 2019. Combining population genomics and forward simulations to investigate stocking impacts: a case study of Muskellunge (*Esox masquinongy*) from the St. Lawrence River basin. *Evol Appl*. 12:902–922.
- Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. 17:155–158.
- Schloss JA, Gibbs RA, Makhijani VB, Marziali A. 2020. Cultivating DNA sequencing technology after the Human Genome Project. *Annu Rev Genomics Hum Genet*. 21:117–138.
- Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat Rev Genet*. 15:749–763.
- Schmidt DA, Campbell NR, Govindarajulu P, Larsen KW, Russello MA. 2020. Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Mol Ecol Resour*. 20:114–124.
- Schweizer RM, Robinson J, Harrigan R, Silva P, Galverni M, Musiani M, Green RE, Novembre J, Wayne RK. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Mol Ecol*. 25:357–379.
- Selmoni O, Vajana E, Guillaume A, Rochat E, Joost S. 2020. Sampling strategy optimization to increase statistical power in landscape genomics: a simulation-based approach. *Mol Ecol Resour*. 20:154–169.
- Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol Evol*. 8:907–917.
- Smith O, Wang J. 2014. When can noninvasive samples provide sufficient information in conservation genetics studies? *Mol Ecol Resour*. 14:1011–1023.
- Souza CA, Murphy N, Villacorta-Rath C, Woodings LN, Ilyushkina I, Hernandez CE, Green BS, Bell JJ, Struggnell JM. 2017. Efficiency of ddRAD target enriched sequencing across spiny rock lobster species (Palinuridae: Jasus). *Sci Rep*. 7:6781.
- Stahlke A, Bell D, Dhendup T, Kern B, Pannoni S, Robinson Z, Strait J, Smith S, Hand BK, Hohenlohe PA, et al. 2020. Population genomics training for the next generation of conservation geneticists: ConGen 2018 Workshop. *J Hered*. 111:227–236.
- Tanjo T, Kawai Y, Tokunaga K, Ogasawara O, Nagasaki M. 2020. Practical guide for managing large-scale human genome data in research. *J Hum Genet*. 66:39–52.
- Thatte P, Joshi A, Vaidyanathan S, Landguth E, Ramakrishnan U. 2018. Maintaining tiger connectivity and minimizing extinction into the next century: insights from landscape genetics and spatially-explicit simulations. *Biol Conserv*. 218:181–191.
- Therkildsen NO, Palumbi SR. 2017. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Mol Ecol Resour*. 17:194–208.
- Tigano A, Jacobs A, Wilder AP, Nand A, Zhan Y, Dekker J, Therkildsen NO. 2020. Chromosome-level assembly of the Atlantic silverside genome reveals extreme levels of sequence diversity and structural genetic variation. *bioRxiv*. doi:10.1101/2020.10.27.357293
- Titcomb GC, Jerde CL, Young HS. 2019. High-throughput sequencing for understanding the ecology of emerging infectious diseases at the wildlife-human interface. *Front Ecol Evol*. 7:126.
- Todd EV, Black MA, Gemmill NJ. 2016. The power and promise of RNA-seq in ecology and evolution. *Mol Ecol*. 25:1224–1241.
- Tseng M, El-Sabaawi RW, Kantar MB, Pantel JH, Srivastava DS, Ware JL. 2020. Strategies and support for Black, Indigenous, and people of colour in ecology and evolutionary biology. *Nat Ecol Evol*. 4:1288–1290.
- Venney CJ, Wellband KW, Heath DD. 2020. Rearing environment affects the genetic architecture and plasticity of DNA methylation in Chinook salmon. *Heredity*. 126:38–49.
- Vuylsteke M, Peleman JD, van Eijk MJ. 2007. AFLP technology for DNA fingerprinting. *Nat Protoc*. 2:1387–1398.
- Waples RS. 2015. Testing for Hardy-Weinberg proportions: have we lost the plot? *J Hered*. 106:1–19.
- Waples RS, Anderson EC. 2017. Purging putative siblings from population genetic data sets: a cautionary view. *Mol Ecol*. 26:1211–1224.
- Waples RS, Antao T, Luikart G. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*. 197:769–780.
- Waples RS, Do C. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol Resour*. 8:753–756.
- Waples RK, Larson WA, Waples RS. 2016. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*. 117:233–240.
- Waples RS, Waples RK, Ward EJ. 2020. Pseudoreplication in genomics-scale datasets. *bioRxiv*. doi: 10.1101/2020.11.12.380410.
- Watsa M; Wildlife Disease Surveillance Focus Group. 2020. Rigorous wildlife disease surveillance. *Science*. 369:145–147.
- Wellenreuther M, Mérot C, Berdan E, Bernatchez L. 2019. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Mol Ecol*. 28:1203–1209.
- Wright S. 1951. The genetical structure of populations. *Ann Eugen*. 15:323–354.
- Wright B, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. 2019. From reference genomes to population genomics: comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics*. 20:453.
- Wright BR, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. 2020. A demonstration of conservation genomics for threatened species management. *Mol Ecol Resour*. 20:1526–1541.
- Yan D, Hu D, Li K, Li B, Zeng X, Chen J, Li Y, Wronski T. 2021. Effects of chronic stress on the fecal microbiome of Malayan Pangolins (*Manis javanica*) rescued from the illegal wildlife trade. *Curr Microbiol*. 78:1017–1025.
- Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. 2020. Zoonotic origins of human coronaviruses. *Int J Biol Sci*. 16:1686–1697.
- Zeller KA, Creech TG, Millette KL, Crowhurst RS, Long RA, Wagner HH, Balkenhol N, Landguth EL. 2016. Using simulations to evaluate Mantel-based methods for assessing landscape resistance to gene flow. *Ecol Evol*. 6:4115–4128.