Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern

Jana Wold[1], Stephanie Galla[1], Anna Santure[2], David Eccles[3], Carolyn J. Hogg[4], Marissa Le Lec[5], Klaus-Peter Koepfli[6], Joseph Guhlin[5], and Tammy Steeves[1]

[1]University of Canterbury
[2]The University of Auckland
[3]Malaghan Institute of Medical Research
[4]The University of Sydney
[5]University of Otago
[6]Smithsonian Conservat Biol Inst

June 2, 2021

Abstract

Structural variants (SVs) are large rearrangements (> 50 bp) within the genome that impact the form and structure of chromosomes. As a result, SVs are a significant source of functional genomic diversity, i.e. variation at genomic regions underpinning phenotype differences, that can have large effects on individual and population fitness. While there are increasing opportunities to investigate functional genomic diversity in threatened species via single nucleotide polymorphism (SNP) datasets, SVs remain understudied despite their potential influence on complex traits of conservation interest. In this future-focused Opinion, we contend that characterizing SVs offers the conservation genomics community an exciting opportunity to complement SNP-based approaches to enhance species recovery. We identify three critical resources to characterize SVs de novo: 1) High-quality, contiguous, annotated reference genome(s); 2) Whole genome resequence data from representative individuals of the target species/populations; and 3) Well-curated metadata including pedigrees. We also leverage the existing literature–predominantly in human health, agriculture and eco-evol biology–to identify pangenomic approaches for readily characterizing SVs and consider how integrating these into the conservation genomics toolbox may transform the way we intensively manage some of the world's most threatened species.

Running Head: Structural variants and conservation genomics

Keywords : Complex traits, Conservation Genetics, Functional Diversity, Pangenomes, Structural Variation, Threatened Species Recovery

Abstract

Structural variants (SVs) are large rearrangements (> 50 bp) within the genome that impact the form and structure of chromosomes. As a result, SVs are a significant source of functional genomic diversity, i.e. variation at genomic regions underpinning phenotype differences, that can have large effects on individual and population fitness. While there are increasing opportunities to investigate functional genomic diversity in threatened species via single nucleotide polymorphism (SNP) datasets, SVs remain understudied despite their

potential influence on complex traits of conservation interest. In this future-focused Opinion, we contend that characterizing SVs offers the conservation genomics community an exciting opportunity to complement SNP-based approaches to enhance species recovery. We identify three critical resources to characterize SVs *de novo* : 1) High-quality, contiguous, well annotated reference genome(s); 2) Whole genome resequence data from representative individuals of the target species/populations; and 3) Well-curated metadata including pedigrees. We also leverage the existing literature–predominantly in human health, agriculture and eco-evol biology–to identify pangenomic approaches for readily characterizing SVs and consider how integrating these into the conservation genomics toolbox may transform the way we intensively manage some of the world's most threatened species.

Structural variants: a new tool in the conservation genomics toolbox

There is a wealth of literature featuring the promise of incorporating functional genomic diversity, defined here as variation at genomic regions underpinning phenotype differences, into conservation management strategies (Harrisson, Pavlova, Telonis-Scott, & Sunnucks, 2014; Kirk & Freeland, 2011). However, over the past 10 years, efforts have largely focused on incorporating genome-wide diversity–which includes both neutral and non-neutral (functional) diversity–as researchers transitioned from low-resolution genetic markers like microsatellites to high-resolution genomic markers like single nucleotide polymorphisms (SNPs) (Allendorf, Hohenlohe, & Luikart, 2010; Mable, 2019). Whether characterized using reduced-representation sequence data or whole-genome sequence (WGS) data, SNP-based estimates of genome-wide diversity are being used to better inform conservation by delineating species (e.g., Quattrini et al., 2019; but see Stanton et al., 2019), detecting hybridisation and introgression (Dufresnes & Dubey, 2020; Peters et al., 2016; but see Forsdick et al., 2021 preprint; Hauser, Athrey, & Leberg, 2021 preprint), identifying conservation units (e.g., Liddell et al., 2020), identifying source populations for conservation translocations (e.g., Dresser, Ogle, & Fitzpatrick, 2017), and guiding conservation breeding programs (e.g., Galla et al., 2020; Wright et al., 2020). Beyond genome-wide diversity, there is a growing consensus around the importance of considering functional genomic diversity in conservation management (Fitzpatrick et al., 2020; Hoelzel, Bruford, & Fleischer, 2019; S. Li et al., 2014; Willi, Van Buskirk, & Hoffmann, 2006). To date, most studies have largely been restricted to SNP-based approaches for targeted immunocompetence gene families like major histocompatibility complex (MHC) and toll-like receptors (TLRs) (Antonides, Ricklefs, & DeWoody, 2017; Elbers, Clostio, & Taylor, 2017; Morrison, Hogg, Gales, Johnson, & Grueber, 2020). Whereas immune-related phenotypes controlled by relatively few, well-characterized gene families are salient targets, the majority of phenotypes of interest to the conservation community remain challenging, in part because most are complex traits that are likely to be highly polygenic (Kardos & Shafer, 2018)(Alonge et al., 2020; Chiang et al., 2017). Despite this, the study of SVs in conservation genomics is limited, in part due to challenges associated with characterizing SVs from SNP data alone (Pang et al., 2010); although a given SV is likely to be linked to a nearby SNP (e.g., Wilder, Palumbi, Conover, & Therkildsen, 2020), they may not be in strong linkage disequilibrium with each other (Pang et al., 2010). Structural variants may also lead to challenges in aligning reads and calling SNPs in a region, such that SNPs are poorly identified and therefore unable to tag a nearby SV (Pang et al., 2010).

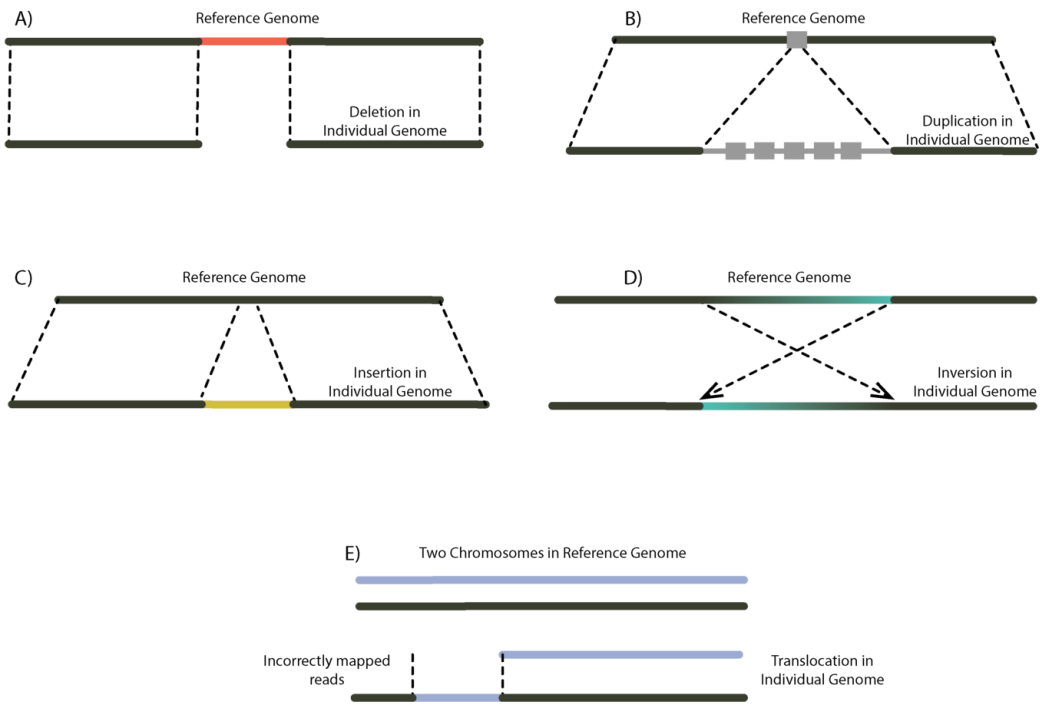

Figure 1: Structural Variant (SV) types and how they differ from a reference genome: A) deletion, where a segment of DNA is not present in an individual, but present in the reference; B) duplication, a rearrangement where there is more than one copy of a particular region of the genome, often back to back; C) insertion, a DNA sequence present in an individual sample, but not present in the reference; D) inversion, a segment of the chromosome that has had a double-stranded break at an upstream and downstream location and become reversed in order; E) translocation, a rearrangement where a portion of a chromosome has been transposed onto another.

Variation in chromosome form and structure is one of the earliest studied types of genetic diversity (e.g., McClung, 1905; Sturtevant, 1921), yet the introduction of Sanger sequencing in the 1970's prompted a departure from the study of variation at the scale of chromosomes to nucleotides (e.g., Sanger & Coulson, 1975; Sanger, Coulson, Barrell, Smith, & Roe, 1980; Woese, Sogin, Stahl, Lewis, & Bonen, 1976). Recent advancements in WGS sequencing technologies coupled with increased computational capacities, primarily applied in the human health sector, have renewed interest in the role of SVs in determining complex traits (Chiang et al., 2017; Pang et al., 2010; Sadowski et al., 2019; Yi & Ju, 2018). However, these advancements are increasingly being applied outside of human health. For example, characterizing SVs in agricultural species has led to the identification of variants associated with economically significant traits in crops like tomato (Alonge et al., 2020), soybean (Liu et al., 2020) and maize (Yang et al., 2019), and identified specific genes and gene regions associated with domestication in vertebrates (Bertolotti et al., 2020; Cagan & Blass, 2016; vonHoldt et al., 2017). In addition, assessments of the impacts of SVs on model organisms indicate that SVs are a likely source of complex trait variation (e.g., SVs are present in ~50% of *Drosophila melanogaster* quantitative trait loci; Chakraborty et al., 2019). Outside of these sectors, SVs have been found to have significant evolutionary, ecological and behavioral impacts (Buchanan & Scherer, 2008; Huynh, Maney, & Thomas, 2011; K.-W. Kim et al., 2017; Knief et al., 2017; Küpper et al., 2016; Wellenreuther & Bernatchez, 2018). For example, SVs may suppress recombination and subsequently result in the formation of a 'supergene', which is a cluster of tightly linked co-adapted alleles. Complex traits determined by supergenes include sperm morphology in zebra finch (K.-W. Kim et al., 2017; Knief et al., 2017), morphology and mating strategy in the Eurasian ruff (Küpper et al., 2016) and migration timing in rainbow trout (Pearse et al.,

2019). However, in some cases where a particular SV is associated with a trait, it may not be the causal variant, but increase the likelihood of *de novo* causal variants. This is the case for a relatively common 1.3Mb inversion on the human Y-chromosome where microdeletions accrue at inversion breakpoints, which can result in profound impacts on male fertility (Hallast et al., 2021).

Structural variants have the potential to have large effects on individual and population fitness as they have been found to affect more overall genome content, intersect with genes more often and as a result are more likely to directly affect gene expression than SNPs (Catanach et al., 2019; Chakraborty et al., 2019; Chiang et al., 2017; Frayling, 2014; Pang et al., 2010). Significant progress in understanding the relationships between regulatory elements and genes has been made in the genomics of human health, and these findings may provide insights into how SVs may impede or enhance threatened species recovery. For example, mouse studies helped identify the mechanisms by which copy number variants (deletions or duplications) modify gene dosage resulting in velocardiofacial syndrome (Lindsay et al., 2001; Merscher et al., 2001). In Mendelian diseases, such as velocardiofacial syndrome, SVs may result in haploinsufficiency, where one copy of a gene is inactivated or deleted and the remaining functional gene copy/copies are unable to compensate for the loss (Lupski et al., 1992). Further, in human cancer research, chromatin conformation studies have revealed that SVs can change the physical distance between regulatory elements and genes in 3D space (Shanta et al., 2020). That is, a regulatory element may appear to be distant along the length of the genome but, once folded within the nucleus, regulatory elements and the genes they control may be physically close, and SVs have been shown to disrupt these interactions (Sadowski et al., 2019; Shanta et al., 2020). There is also growing evidence that SVs can further impact the gene regulatory landscape by altering the formation of topologically associating domains (TADs), which are genomic regions that physically interacts with themselves more frequently than with genomic regions elsewhere (Sadowski et al., 2019; Shanta et al., 2020).

Here, we leverage our cross-sector expertise in conservation (AWS, CJH, JG, JRW, KPK, MLL, SJG, TES), human health (DE), agriculture (JG), eco-evol biology (AWS) and bioinformatics (DE, JG, JRW, MLL) to explore SVs in a conservation context. We identify three critical resources for characterizing SVs *de novo* : 1) High-quality, contiguous, well annotated reference genome(s); 2) Whole genome resequence data from representative individuals of the target species/populations; and 3) Well-curated metadata, including pedigrees. We also highlight how the incorporation of genome assemblies for multiple reference individuals– that is, a pangenomic approach–can be used to improve SV discovery and genotyping, even for the most complex SVs (Alonge et al., 2020; Eizenga et al., 2020; Gao et al., 2019; Golicz et al., 2016; Tettelin et al., 2005). We then discuss the opportunities and challenges for relating SVs to complex traits and explore how eco-evol modelling can be leveraged to mitigate the impacts of managing maladaptive complex trait(s) on genome-wide diversity. Together, these approaches have strong potential to improve conservation outcomes, particularly for intensively managed threatened species.

Structural variant discovery and genotyping

A high-quality reference genome is an invaluable investment

Increased accessibility to high-quality reference genomes through partnerships with global genome consortia are providing high-quality, contiguous, well annotated genome assemblies that include both short and long read data and scaffolding approaches (e.g., VGP, Bat1K, Bird10K; Rhie et al., 2021; Teeling et al., 2018; Zhang, 2015), both for intensively managed threatened species and their close relatives (Whibley, Kelley, & Narum, 2020). As costs of generating and analyzing WGS data continue to drop, a growing number of conservation genomicists working on species beyond global genome consortia are investing in the assembly and annotation of high-quality reference genomes. We readily recognize that not all conservation programs are able to access the resources or sample quality to generate a high-quality, contiguous, well annotated reference genome, nor do we recommend this action for all threatened species. However, as described below, this investment is a near necessity for the accurate characterization of *de novo* SVs, particularly if the goal

is to characterize a broad range of SV types. Further, as described below, if the goal is to characterize SVs–including complex SVs–across the genome, a pangenome is invaluable (Figure 2).

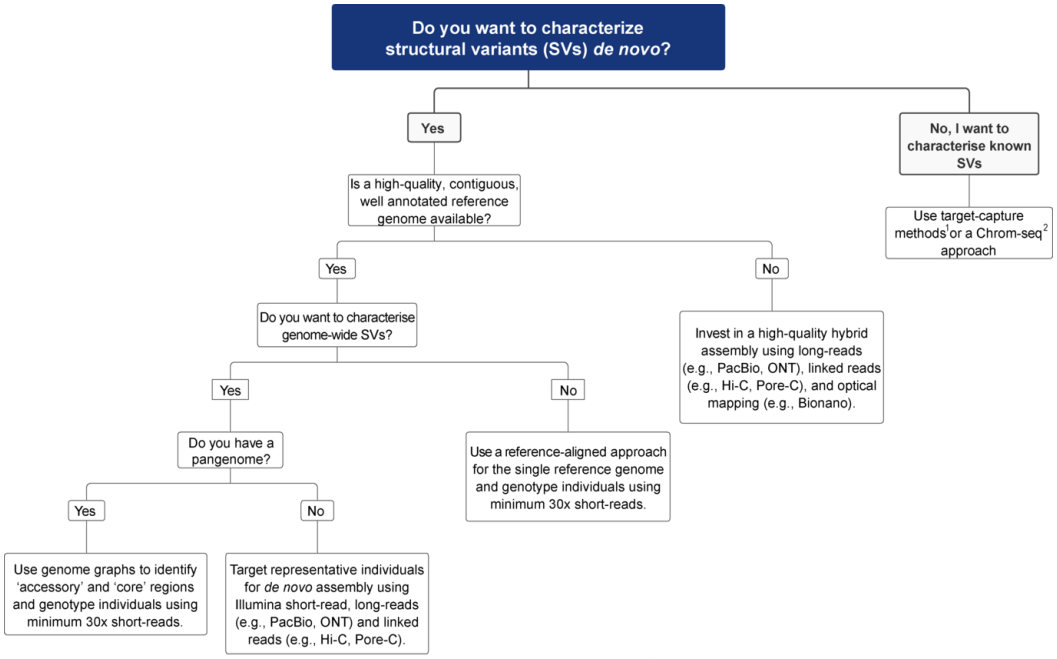

Figure 2: Decision tree for characterizing structural variants (SVs). See text for details. PacBio = Pacific Biosciences; ONT = Oxford Nanopore Technologies. [1]Choi et al., (2009); [2]Iannucci et al., (2021).

Alignment-based SV discovery methods

Alignment-based SV discovery methods rely on identifying patterns in read mapping between a sample and the reference genome. Common algorithms include read-pair, read depth, split-reads and *de novo*or local assembly (Hajirasouliha et al., 2010; Korbel et al., 2007; Yoon, Xuan, Makarov, Ye, & Sebat, 2009). No single algorithm is well suited to detecting all SV types. For example, read pair-based algorithms, where the orientation and distance between paired ends are assessed, are suitable for detecting deletions, duplications, and inversions. In the case of read depth approaches, deletions and duplications are identified through variation in mapping depth; whereas split-reads identify regions where alignments map across a breakpoint, this is best suited for discerning the ends of an inversion or translocation (Figure 3). Finally, *de novo* and local assembly algorithms are best suited to identifying variants not present in the reference (e.g., insertions) as reads aligned to contigs may be reassembled alongside unmapped reads for pairwise comparison to a reference (Mahmoud et al., 2019). Early implementation of SV discovery programs typically used a single approach (e.g., BreakDancer, Pindel; Chen et al., 2009; Ye, Schulz, Long, Apweiler, & Ning, 2009). Although these programs are less computationally intensive, they are limited to calling a few SV classes and tend to underperform against generalist software that incorporate at least three algorithms (e.g., Lumpy, Manta, SvABA; X. Chen et al., 2016; Layer, Chiang, Quinlan, & Hall, 2014; Wala et al., 2018).

Because of the challenges of identifying and classifying SVs, many SV discovery approaches have a high (and systematic) false positive rate (e.g., Cameron et al., 2019). In order to address this error, an ensemble approach for SV characterization has been extensively applied in human health (Ho, Urban, & Mills, 2020 for review). With an ensemble approach, multiple SV callers are integrated into a single pipeline as a means to create more certainty around SV detection, with only variants that intersect multiple SV callers retained (Becker et al., 2018; Mohiyuddin et al., 2015; Zarate et al., 2018 preprint). The basic principle

being that multiple lines of evidence supporting the same event should increase accuracy and precision of SV detection. While this approach has improved upon SV characterization in human genome consortium projects with established 'true' data sets, it has also been shown to negatively impact the true positive rate (the proportion of 'true' SVs that are detected) and positive predictive value (the proportion of classified SVs that are 'true' SVs). In fact, a recent benchmarking study found that distinguishing between true variants and false positives in an ensemble discovery method is difficult due to the significant overlap in false positives (Cameron et al., 2019). Although there are a few ensemble pipelines that have attempted to standardize the process of variant discovery, merging and quality assessments of consensus calls (e.g., FusorSV, MetaSV, Parliament2; Becker et al., 2018; Mohiyuddin et al., 2015; Zarate et al., 2018), many projects develop ensemble methods independently (Ho et al., 2020). This is significant given that choice of discovery pipelines, filtering, and merging methods heavily impact the performance of an ensemble approach (Ho, Urban, & Mills, 2020 for review). Further, standardized methods for benchmarking variant calls are under active development (Ho et al., 2020; Parikh et al., 2016; Zook et al., 2020). Access to standardized benchmark data representative of a given non-model organism is largely improbable at this time and will be challenging as conservation genomicists move forward in this space. This is in part because intensively managed threatened species generally do not have the luxury of extensive genomic and/or financial resources to establish robust reference databases of SVs verified with multiple sequencing platforms as is common in human genomics (Ho et al., 2020; discussed below in section on*SV characterization with low coverage short-read sequence data*).

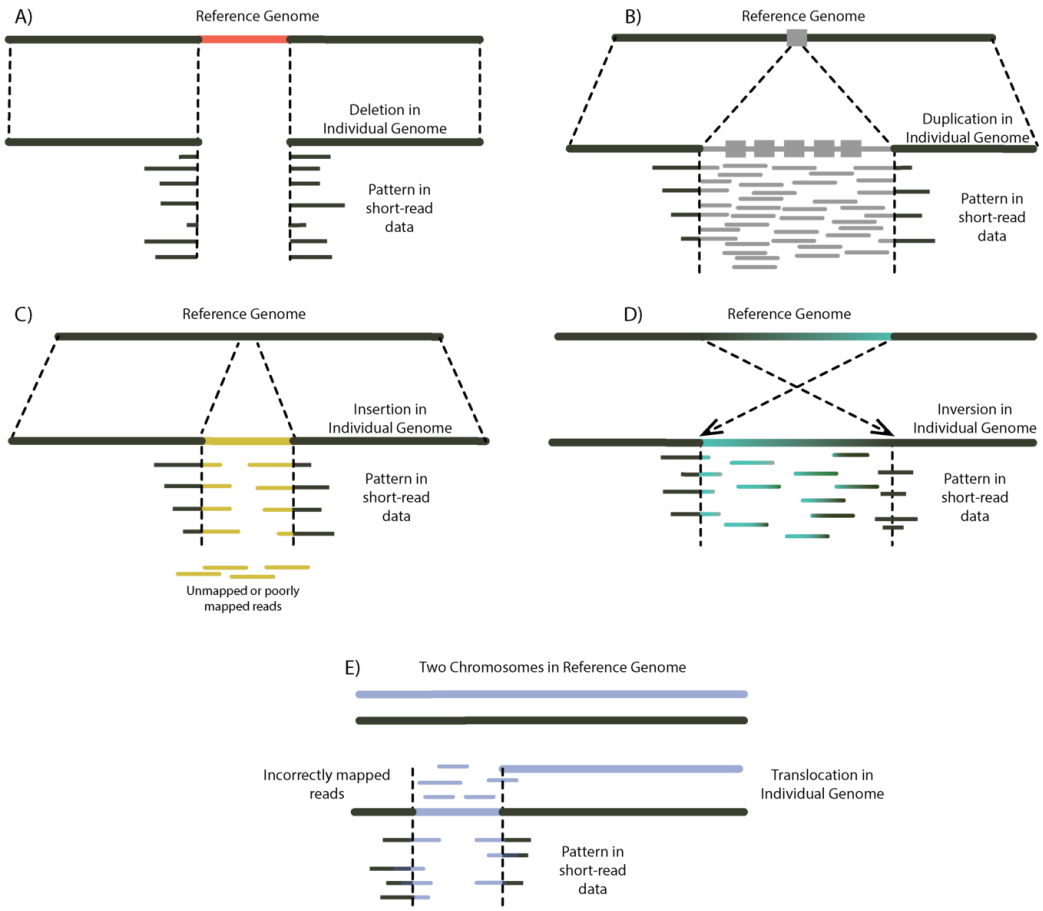

Figure 3: Structural Variant (SV) types and common problems associated with short-read sequence data: A) deletion, called when reads do not map to, and/or are split across, a given region on the reference genome. Deletions are generally the most straightforward SV to detect with short-read data, but complex rearrangements may preclude mapping and result in a false call; B) duplication, typically identified by an increase in read depth, however reference error may result in a false call, preclude assessments of copy number variation, or miss sequence variation due to unmapped reads; C) insertion, short-reads may be mapped if the majority of each read aligns to the reference genome, but reads composed mostly of the insertion sequence may remain unmapped; D) inversion, breakpoints (i.e., exact positions of double-stranded DNA breaks) are difficult to resolve as they typically occur in highly repetitive regions; E) translocation, breakpoints difficult to resolve as they commonly occur in highly repetitive regions, reads may also incorrectly map to the 'original' chromosome from which the translocation arose. Further, it is common for more than one SV to occur in close proximity. Resolving these complex SVs is particularly challenging with short-read sequence data as multiple mapping signals may contradict one another (e.g., multiple deletions in close proximity, deletions that overlap with inversion/translocation breakpoints).

In light of these challenges, rather than an ensemble approach for SV discovery for intensively managed threatened species, it may be best to use a generalist SV discovery software package that combines a range of SV discovery algorithms to target a broad range of SVs at once. In the absence of a validated SV call set for error correction, the use of a generalist SV discovery software package may help mitigate the uncertainty in best practice to merge SVs and assess the quality of calls from multiple programs in ensemble approaches. For example, ensemble pipelines may be outperformed by generalist programs that implement assembly-based

methods alongside read depth, read pairs and/or split-read approaches (i.e., Manta or GRIDSS; Cameron et al., 2019; Kosugi et al. 2019). These programs not only perform well on their own, but have the added benefit of performing well over a range of SV types (Cameron et al., 2019; Kosugi et al., 2019). However, the computational resources required to characterize SVs at the population scale using generalist SV discovery software are substantial (e.g., for a diploid 1.15Gb genome; 72 physical cores, 460 Gb RAM, >3 Tb storage for 170 individuals sequenced to 30X coverage).

Pangenomes improve genome-wide structural variant discovery and genotyping

A pangenome is the aggregate characterization of genomic diversity present in a group of interest, including species and populations (e.g., diversity between strains of tomato, Alonge et al., 2020). Pangenomes offer a straightforward solution to address the challenges associated with SV discovery and genotyping with short-read sequence data. Although originally developed to characterize variation in bacteria (Tettelin et al., 2005), they are commonly used in studies of complex trait diversity in humans (Pang et al., 2010) and agriculturally significant species (e.g., cattle, goats, soybean and maize; Bickhart et al., 2020; Golicz, Batley, & Edwards, 2016; H. Li, Feng, & Chu, 2020; Y. Liu et al., 2020; Low et al., 2020; McHale et al., 2012; Yang et al., 2019). There are two components of a pangenome: the 'core' genomic regions that do not vary among individuals, and 'accessory' genomic regions that vary among individuals (Bayer, Golicz, Scheben, Batley, & Edwards, 2020; Golicz, Batley, & Edwards, 2016; Hurgobin & Edwards, 2017; Figure 4). In a pangenomic approach, genomes of multiple individuals are assembled *de novo* using multiple platforms (e.g., long-reads, Hi-C, Optical mapping; Soto et al., 2020; Weissensteiner et al., 2020), followed by pairwise comparisons of whole-genome alignments for SNP and SV discovery (e.g., Cortex, MUMmer, Minimap2; Delcher et al., 1999; Iqbal, Caccamo, Turner, Flicek, & McVean, 2012; H. Li, 2018). Once variant discovery is complete, genome graphs representing the variation in the pangenome may be constructed to efficiently represent 'core' and 'accessory' regions (Eizenga et al., 2020; H. Li, 2018; Tettelin et al., 2005). Genome graphs are a powerful method for population-level genotyping and consistently outperform alignment-based genotyping (e.g., Ebler et al., 2020 preprint; Eggertsson, 2017; Iqbal et al., 2012; D. Kim, Paggi, Park, Bennett, & Salzberg, 2019; H. Li, 2018). As a result, pangenomic approaches can capture complex variants including SVs across the genome, and offer an exciting opportunity to explore the genomic basis of complex traits (e.g., Gao et al., 2019).

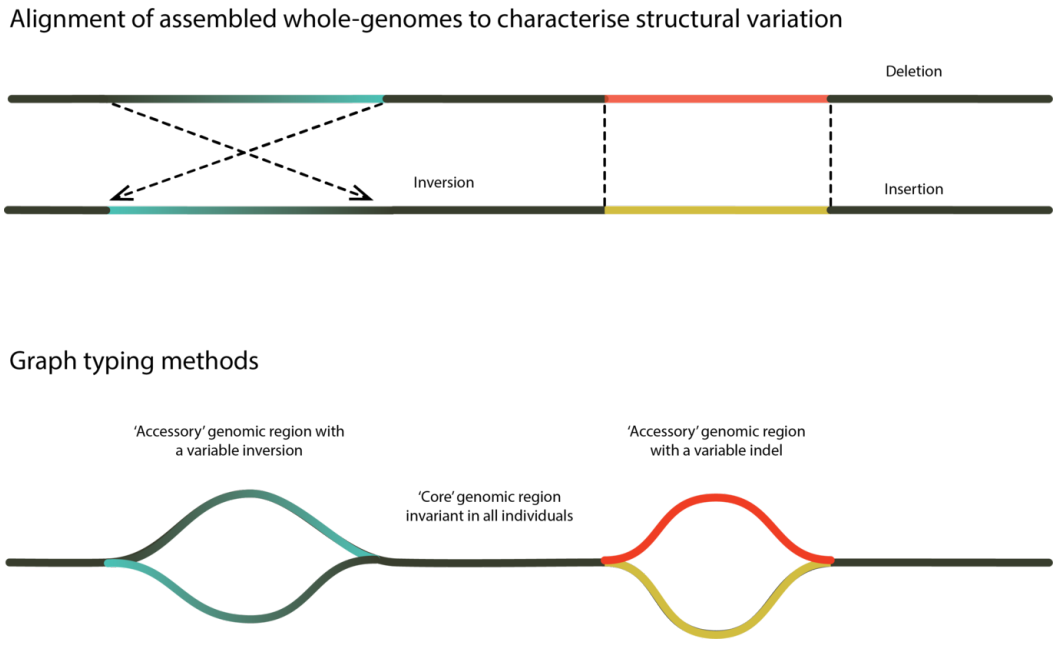

Figure 4: Characterization of SVs using assembled whole-genomes and genome graphs using a pangenomic approach: (a) Schematic of structural variants (SVs) discovery using alignment of assembled whole genomes. (b) Schematic of genome graphs, which can then be used to characterize a pangenome and facilitate rapid genotyping of individuals, as 'core' regions (genomic regions that do not vary among individuals) are readily distinguished from 'accessory' regions (genomic regions that do vary among individuals). Genome graphs can also facilitate accurate genotyping as more than one allele may be considered at once. Pangenomic approaches remove reference bias and are better able to fully capture insertions, inversion haplotypes and complex SVs (i.e., genomic regions where multiple SVs occur in close proximity).

For intensively managed threatened species, individuals selected for reference genomes may not be the most biologically representative in that many are chosen to showcase a charismatic individual or are the result of opportunistic sampling. Regardless, in most cases, it is unlikely that a single individual accurately represents the diversity of the species as a whole, even in highly inbred species (Gao et al., 2019; McHale et al., 2012). The bias that comes with using a single reference genome presents a significant barrier to capturing all genomic diversity–including large and complex SVs–in species of interest, the impacts of which have been observed in soy (Liu et al., 2020), wheat (Montenegro et al., 2017), and *Brassica* spp. (Golicz et al., 2016). The primary advantage of a pangenomic approach is that multiple alternative alleles are assessed while read mapping, effectively removing reference bias (Paten, Novak, Eizenga, & Garrison, 2017). In other words, sequence reads are not precluded from mapping if they are not represented in the primary reference (e.g. insertions, Figure 3). The promise of pangenomes is exemplified in agriculturally significant species such as soy where many more SVs are resolved using a pangenome compared to using an alignment-based approach with a single reference genome (Liu et al., 2020). However, the ability of conservation programs to establish pangenomes largely depends on the availability of existing genomic resources and the ability to secure substantive funding, likely limiting this approach to intensively managed threatened species.

SV discovery and genotyping with short-read sequence data

Low coverage (e.g., ~10x) short-read sequencing currently dominates the field of conservation genomics (Cam et al., 2020; Lado et al., 2020; Lew et al., 2015; Lujan, Weir, Noonan, Lovejoy, & Mandrak, 2020; Oyler-McCance, Cornman, Jones, & Fike, 2015; Robinson et al., 2016). These low coverage datasets are

generally cost effective and appropriate for characterizing SNPs using a single reference genome, whereas a minimum of 30x coverage is necessary *de novo* SV discovery and genotyping (Ahn et al., 2009; Kosugi et al., 2019; Sims, Sudbery, Ilott, Heger, & Ponting, 2014; Wang et al., 2008). This is because at low coverage it is challenging to determine whether a detected variant is an artefact of sequencing/mapping error or a 'true' variant (Figure 3). For example, mapping errors may occur when the relative size of a SV spans a large portion or the entire length of a read (Sedlazeck, Lee, Darby, & Schatz, 2018), or in the case of a complex rearrangement, prevents mapping altogether (Sedlazeck et al., 2018; Yi & Ju, 2018). In addition, read insert size (400 - 600 bp) may bias the discovery of insertions and deletions in some pipelines (Kosugi et al., 2019). By increasing read depth to 30x on average, the likelihood of distinguishing between sequencing error, read mapping errors and true genomic variation increases as well. The question then becomes how best to fully utilize these short-read datasets to investigate SVs. For example, if a validated variant call set is not available and the goal is to quantify and compare patterns of SV diversity across populations, investing in a high-quality, contiguous, well annotated reference genome and 30x coverage short-read resequencing data may be sufficient for SV discovery and genotyping (Kosugi et al., 2019).

However, it is important to note that short-read SV discovery pipelines underperform when genotyping individuals (Chander, Gibbs, & Sedlazeck, 2019). If accurate individual genotypes are essential for downstream analyses, it is not recommended to use genotype outputs from a short-read SV discovery pipeline, but rather validate SVs through targeted and/or long-read sequencing approaches, then use a software package specific for genotyping with short-read data (Chander et al., 2019). In instances where an established call set of SVs is available (e.g., SVs characterized with a pangenomic approach), this enables population-level analyses since genotyping samples is less computationally intensive than *de novo* SV discovery as the pattern of discordant reads may be assessed at specific points throughout the genome (i.e., SV breakpoints). There are some caveats to consider, like many SV discovery pipelines, genotyping software packages have variable performance across a range of SV sizes and types. As such, many software packages may not be able to provide a genotype for the whole range of SVs called in the discovery pipeline. Another challenge is that some SV genotyping programs are specific to a particular *de novo* SV discovery pipeline (e.g., SVTyper and Lumpy), and may require formatting to conform to strict data input requirements. A final consideration is that genotyping programs are still an area of active development. Although they may alleviate some of the false discovery rates in some instances, they are not a definitive solution to high false discovery rates prevalent in SV discovery programs (Chander et al., 2019 for review). As a result, putative SVs identified using short-read data alone should be treated as preliminary. One method that may aid in removing false variants is a trio-binning approach, whereby SVs that fail to abide by Mendelian inheritance patterns (either due to challenges with accurate genotyping or incorrect variant calls) are removed from the SV call set. Alternatively, if sufficient resources are available, the sensitivity of long-read sequence data may be leveraged in a modified pangenomic approach for SV discovery and may facilitate the use of lower coverage short-read resequencing data for SV genotyping, but this largely remains untested across species. Target-capture methods for genes or regions of interest can provide a more affordable alternative to 30x WGS for species with large genomes and projects with tight budgets (Andermann et al., 2020). For example, genomic studies in dogs indicated that a region on chromosome 6 was under positive selection in domestic dog breeds (vonHoldt et al., 2017). A targeted resequencing approach significantly reduced sequence coverage requirements and facilitated SV characterization for this region of the genome (vonHoldt et al., 2017). It is notable that this approach is commonly followed up with PCR validation of relevant SVs (vonHoldt et al., 2017), and that the larger genome-wide context of these variants may be lost as a consequence of targeting specific regions.

SV discovery and genotyping beyond short-read sequence data

Short read data is a useful starting point for identifying some SVs. However, there are biases in the type and size that can be easily detected due to the read length of short read data, this leaves many larger and/or complex SVs undiscovered (Figure 3; also see Ho et al., 2020). When characterizing genomic features, there are many sequencing platforms and approaches to choose from, and although they may perform well when addressing specific challenges, each has its own caveats (Table 1).

Two providers prominently feature in long-read sequencing: Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). Since the launch of these technologies in 2011 and 2014, the long-read sequencing space has been characterized by fast-paced progress and innovation as demonstrated by the first telomere to telomere assembly of the human X chromosome achieved with ultra-long-read sequencing (Miga et al., 2020). The precise error rates between these two technologies remain somewhat contentious (Dohm, Peters, Stralis-Pavese, & Himmelbauer, 2020; Lang et al., 2020 preprint), but as a general rule, ONT currently provides longer average read lengths than PacBio overall (Logsdon, Vollger, & Eichler, 2020) at the cost of higher sequence error rates. Despite these challenges, the power of long-read sequencing technologies to span a significant portion, if not the entire length, of complex regions of the genome in a single read provides a powerful tool for SV discovery and population-level genotyping. When used in conjunction with a high-quality, contiguous, well annotated reference genome, this improves confidence in read mapping genome-wide (Amarasinghe et al., 2020 for review), and substantially increases precision (the proportion of variant calls that are 'true') and recall (the proportion of 'true' SVs detected) rates for both SNPs and SVs (Wenger et al., 2019). In addition, platforms that directly sequence native DNA remove the amplification bias common in many short-read sequencing approaches (Depledge et al., 2019). Furthermore, there are emerging 'adaptive' sequencing approaches that have the potential to selectively sequence specific regions of the genome (Payne et al., 2020 preprint). It remains to be seen however if this technology is ready for wide use beyond human clinical applications.

Structural variants significantly alter genome topology and impact the gene regulatory landscape (Sadowski et al., 2019; Shanta et al., 2020). In light of these impacts, the hierarchical organisation of DNA within the nucleus is of particular interest when investigating the relationship of transcriptional regulation mechanisms. Chromatin conformation capture (3C) based sequencing approaches enable the investigation of the organisation of chromatin genome-wide (Kong & Zhang, 2019 for review) and identified the chromatin signature in gene expression (Lieberman-Aiden et al., 2009; Lupiáñez et al., 2015; Shanta et al., 2020). In addition, there are emerging advancements in Nanopore sequencing methods to integrate chromatin conformation capture with long-read sequencing (i.e., Pore-C; Ulahannan et al., 2019 preprint). Rather than the amplification bias introduced by preparing a short-read library, long-read sequencing provides data on chromatin at a range of distances along the linear genome and enables contacts to be sequenced without amplification. However, long-read sequence data alone cannot consistently resolve whole chromosomes (Belser et al., 2018). Optical mapping approaches are a useful complement to long-read sequencing approaches, and have enhanced genome assembly outcomes by providing insights into the 'big picture' of large-scale genomic variants (as per Weissensteiner et al., 2020). Optical mapping utilises a technique based on light-microscopy to identify specific sequence motifs (such as restriction enzyme cut sites), which are then used to generate images of fluorescently-labeled DNA molecules (Schwartz et al., 1993), enabling the characterization of large, complex rearrangements missed by long-reads alone (Yuan, Chung, & Chan, 2020). On average, optical maps span ~225 kb, providing information on the physical distance and relationship among genomic features. Besides being used to improve the scaffolding of genome assemblies (Howe & Wood, 2015; Zhang, 2015), including those of endangered species (S. Li et al., 2014), optical mapping methods directly enable the identification of both intraspecific and interspecific SVs (Levy-Sakin et al., 2019; Zhihai et al., 2016). The primary current commercial provider of optical mapping technology is Bionano Genomics and their Saphyr instrument, which uses a nano-channel microfluidic chip to linearise and capture images of fluorescently-labeled ultra-long DNA fragments to generate optical maps at a resolution of 500bp (Yuan et al., 2020). While optical maps provide information on the physical topology of chromosomes, they do not provide sequence information on an allele. Because long-reads and optical maps complement each other, the ideal data set for SV discovery would include both data types (e.g., Soto et al., 2020; Weissensteiner et al., 2020).

Feature	Short-reads	Long-reads	Pangenomes	Optical mapping
Type of SV that can be assessed	Deletions, Insertions, Inversions, Translocations	Deletions, Insertions, Inversions, Translocations, Duplications, Complex SVs	Deletions, Insertions, Inversions, Translocations, Duplications, Complex SVs	Deletions, Insertions, Inversions, Translocations, Duplications
High-quality reference genome	Required	Required	Required (multiple)	Optional
Technology used to generate data for SV discovery and genotyping	Illumina HiSeq, NovaSeq, NextSeq	Pacific Biosciences (PacBio) Sequel; Oxford Nanopore Technologies (ONT)	1. Pangenome generated using a combination of short-, long-, linked-reads (e.g., Hi-C, Pore-C) for representative individuals 2. Short- and/or long-reads generated for population-level genotyping	Bionano Genomics Saphyr
DNA quality	Moderate to high molecular weight	High to ultra-high molecular weght[a]	High to ultra-high molecular weight[a]	Ultra-high molecular weight[a]
Length of SVs detected	50 bp to < 1 Mb	50 bp to > 1 Mb	50 bp to > 1 Mb	500 bp to > 1 Mb
Sequence coverage required	> 30x	> 10x	1. > 50x to generate pangenome 2. > 30x for population-level genotyping	> 30x
Method of SV discovery	Alignment-based, including read-pair, read depth, splits-reads and local assembly	Alignment-based with local assembly	1. Alignment of multiple *de novo* assembled whole genomes; 2. Genome graphs	Alignment-based, either to reference genome or to optical maps from different samples
SV genotype evidence	Multiple sources of evidence required depending on SV type (read pair, read depth, split read)	Single source of evidence (long continuous reads) usually sufficient	Two sources of evidence required (pangenome and population-level sampling)	Order, position, and orientation of fluorescently-labeled sequence motifs
Example algorithms/programs	BreakDancer; LUMPY; Manta	SMRT-SV (PacBio); NanoVar (ONT)	Minimap2; MuMmer; Graphtyper2	Bionano SVCaller
Proportion of SVs discovered across the genome	Moderate	High	High	High

Relating structural variants to complex traits of conservation interest

For intensively managed threatened species, complex traits of conservation interest include, but are not limited to, disease susceptibility, reduced fertility and developmental abnormalities (e.g., Moran et al., 2021; Murchison et al., 2012; Roelke, Martenson, & O'Brien, 1993; Savage, Crane, Team, & Hemmings, 2020). Identifying the genomic underpinnings of maladaptive compex traits like these, many of which are likely to be polygenic, is challenging, particularly using SNPs (Kardos & Shafer, 2018). However, SVs are a promising complement to SNPs given that SVs have been found to affect overall genome content, and intersect with genes more often than SNPs (Catanach et al., 2019; Chakraborty et al., 2019; Chiang et al., 2017; Frayling, 2014; Pang et al., 2010). Further, characterizing SVs may help resolve complex variants (i.e., multiple overlapping events) that may inhibit the discovery of causal variants (Alonge et al., 2020; McHale et al., 2012). Structural variants may also suppress recombination and facilitate the evolution of supergenes that underlie complex traits (Huynh et al., 2011; Jay et al., 2018; K.-W. Kim et al., 2017). For example, sperm swimming speed in zebra finch is determined by inversion haplotypes with heterokaryotypic males producing faster sperm than homokaryotypic males (K.-W. Kim et al., 2017; Knief et al., 2017). In addition, a supergene resulting from an inversion has been found to determine mating strategy and morphology in the Eurasian ruff (Küpper et al., 2016), it is notable that this is likely a lethal recessive variant (Lamichhaney et al., 2016). Finally, complex trait variation has been found to underlie behavioural differences among white-striped and tan-striped morphs in White-throated sparrows (Merritt et al., 2020). Here again, a large inversion has resulted in a supergene with significant morphological and behavioural differences, however, the aggressiveness noted in the white-striped morph is monogenic (Merritt et al., 2020).

To date, there are generally two approaches to investigating the genomic-basis of complex traits: 1) Comparative genomics where well characterized groups (either populations or species) are used to identify highly differentiated genomic regions (Alonge et al., 2020; McHale et al., 2012; vonHoldt et al., 2017; Weissensteiner et al., 2020); and 2) Association studies where associations between specific markers (ie., SNPs and/or SVs) and phenotype are assessed (Chakraborty et al., 2019). Comparative approaches using SVs have revealed signatures of domestication (e.g., aquaculture salmon and dogs; Bertolotti et al., 2020; Cagan & Blass, 2016; vonHoldt et al., 2017), and the relative contribution of SVs to species evolution (e.g., Atlantic cod, Corvids, *Heliconia* butterflies, Sunflowers; Berg et al., 2017; Joron et al., 2011, 2006; Rieseberg, Whitton, & Gardner, 1999). Quality genome annotations–for focal species themselves or their close relatives–are crucial to determine whether candidate regions contain genes or regulatory regions in close association with SVs (Alonge et al., 2020; Bertolotti et al., 2020; vonHoldt et al., 2017). In the case of domestic dogs, the detection of a SV associated with specific behavioural traits potentially selected for during domestication was facilitated by species level comparisons with wolves (vonHoldt et al., 2017). These associations were made possible by having clearly defined and well-documented phenotypic traits, and comparisons between populations/species with divergent demographic and evolutionary histories.

Whole-genome SNP-based association studies have been extensively applied to model organisms, agriculturally significant species and humans (Goddard & Hayes, 2009; Mackay, Stone, & Ayroles, 2009; Mauricio, 2001; Stranger, Stahl, & Raj, 2011). However, for many SNP-based association studies, candidate loci do not explain the vast majority of known trait heritability (i.e., the proportion of trait variation that is due to genetic rather than environmental differences between individuals; Clarke & Cooper, 2010; Eichler et al., 2010; Manolio et al., 2009). This "missing heritability" is attributed to genetic variation not explained by candidate loci, that is to say that causal variants may remain uncharacterized (Manolio et al., 2009). Another hypothesis to explain this missing heritability is that individually rare SVs are a significant source of complex trait variation (Eichler et al., 2010; Frazer, Murray, Schork, & Topol, 2009), this is a persuasive argument given the challenges associated with accurately identifying and genotyping SVs (discussed in *Structural variant discovery and genotyping* above), and that association studies perform poorly when causal variants are rare (Dickson, Wang, Krantz, Hakonarson, & Goldstein, 2010; Spencer, Su, Donnelly, & Marchini, 2009; but see Wray, Purcell, & Visscher, 2011). The inclusion of SVs into association studies is promising as evidence suggests that SVs have a larger effect and may be deleterious more frequently than SNPs (Chakraborty et

al., 2019; Conrad et al., 2010; Cridland, Macdonald, Long, & Thornton, 2013; Emerson, Cardoso-Moreira, Borevitz, & Long, 2008; Rogers et al., 2015). For example, a quantitative trait loci (QTL) study for *Drosophila melanogaster* found that about half of all candidate genes underpinning mapped QTL were impacted by SVs, and that a large proportion of genes contained multiple rare SVs (Chakraborty et al., 2019).

Integrating structural variants into applied conservation

The central paradigm of conservation genomic management, based on minimizing inbreeding in the short term and maximizing genome-wide diversity as a proxy for evolutionary potential in the long-term, is a practice supported by decades of empirical research. Incorporating selection–especially selection against maladaptive complex traits–has the potential to reduce genome-wide diversity and the ability for threatened populations and species to respond to novel selection pressures in a changing world (Kardos & Shafter, 2018; Hoelzel, Bruford, & Fleischer, 2019). We acknowledge several recent studies have challenged the significance of genome-wide diversity in population and species persistence (e.g., Robinson et al., 2016; Robinson et al., 2019; Morin et al. 2020; Kyriazis, Wayne, & Lohmueller, 2021; Teixeira & Huber 2021). Here, we highlight the growing need to determine how best to mitigate the impacts of maladaptive complex traits without compromising genome-wide diversity.

Whereas *ad hoc* management against maladaptive traits that are clearly Mendalian-inherited does occur for intensively managed threatened species (e.g., California condor; Moran et al. 2021), it is challenging for conservation practitioners to manage maladaptive complex traits (e.g., disease susceptibility, reduced fertility, developmental abnormalities) with confidence. In the hypothetical scenario that a complex trait is clearly associated with a deleterious SV, it is conceivable that individuals harbouring such variants could also be managed on an *ad hoc* basis. For example, if one of two siblings had a deleterious variant then that individual would likely be excluded from active management. However, the long-term consequences of hypothetical decisions like these on genome-wide diversity are unclear. Therefore, conservation practitioners will benefit from the integration of SV data and associated metadata with individual-based models (e.g., Languth et al. 2017) to forecast the consequences of selection against complex traits underpinned by deleterious SVs on genome-wide diversity to better inform conservation management actions. To achieve this, strong collaboration between conservation practitioners, population ecologists (e.g., modelers), quantitative geneticists, and conservation genomicists will be required to make the most effective models possible (Hohenlohe, Funk, & Rajora, 2020).

Concluding remarks

To date, most conservation genomic studies have been SNP-based. Whereas SNPs impact single nucleotides, SVs can capture multiple genes and gene regions simultaneously and are therefore more likely to impact fitness. Indeed, mounting evidence indicates that SVs play an important role in determining complex traits for both model and non-model species. Emerging approaches for SV discovery and genotyping, and relating SVs to complex traits, are now in reach for intensively managed threatened species with readily accessible high-quality genomic resources and well-curated metadata, and we are confident that discoveries related to the role of SVs in determining complex traits of conservation interest are imminent. However, we readily acknowledge that establishing a robust SV call set verified through the use of multiple sequencing platforms (e.g., using a pangenomic approach) and/or aligned *de novo* short- or long-read assemblies for population-level genotyping will be untenable for many threatened species, especially those with large and complex genomes, for the foreseeable future. Nonetheless, we anticipate a growing number of empirical studies will explore how to best mitigate the impact of maladaptive complex traits, including those underpinned by SVs, for the recovery of some of the world's rarest species.

Glossary

Breakpoint: A double stranded break in the DNA, occurs on either side of an inversion, or a translocation.

Chromatin Conformation Capture (3C) : A suite of molecular biology techniques for capturing the spatial organisation of chromatin within the nucleus of the cell. Common iterations include Hi-C, Pore-C and Omni-C.

Complex traits : Phenotypes that are polygenic and influenced by both genetic and environmental factors.

Complex structural variants: An instance where multiple SV classes occur in close proximity, such as an inversion with different deletions at breakpoints.

Deletion : A chromosomal rearrangement where a sequence of DNA is missing in an individual but present in the reference. The inverse of an insertion.

Duplication

Functional diversity : Genomic variation in regions underpinning phenotypic traits, may occur directly in genes or within regulatory elements. Variation in these regions can have consequences for individual fitness.

Genome-wide diversity: Genomic variation inclusive of functional and neutral diversity. The sum of all variation within the genome.

Gene dosage : The number of copies of a gene, may vary among individuals should mutations inactivate, delete or duplicate genes.

Genome graph : A representation of multiple possible sequences (i.e., a pangenome).

Haploinsufficiency : When one chromosomal copy of a gene is inactivated or deleted completely and the remaining copy is unable to produce sufficient gene product to maintain normal function.

Insertion : A chromosomal rearrangement where a sequence of DNA is 'inserted' into the original sequence. May also be defined as a sequence of DNA present in an individual that is missing in the reference. The inverse of a deletion.

Inversion : A chromosomal rearrangement where a section of the DNA is reversed in order in an individual compared to a reference.

Neutral diversity : Genomic variation that does not impact fitness, most variation in the genome is neutral.

Pangenome : A reference genome approach where the genomes of multiple individuals are assembled and aligned to form genome graphs. A pangenome is divided into variable 'accessory' and invariable 'core' components, with the goal to characterize all the variation present within a group of interest. Currently the scale of pangenome projects range from populations/strains to subfamilies.

Regulatory element : Regions of non-coding DNA involved in the regulation of gene expression, the two primary regulatory elements are enhancers and promoters.

Structural Variant (SV)

Topologically Associated Domain (TAD) : Region of the genome that physically interacts with itself more frequently than with regions outside of the TAD.

Translocation : A chromosomal rearrangement where a segment of one nonhomologous chromosome is transferred to another.

Acknowledgements

We wish to thank the Guest Editors of the 'Whole-genome sequences in Molecular Ecology' Special Issue for the invitation to contribute this Perspective. We are grateful to the Kākāpō125+ Consortium led by Genomics Aotearoa for providing the opportunity to explore structural variants in a conservation context. We extend our thanks to the University of Canterbury's Conservation, Systematics and Evolution Research Team (ConSERT), including Roger Moraga at Tea Break Bioinformatics for robust dialogue–and Kylie Price and Joanna Roberts for their support and guidance–on this topic. We acknowledge the NSF Track 2 EPSCoR Program (award number OIA-1826801) for their support of SJG. We also thank three reviewers for their insightful comments on a previous version of the manuscript.

Data accessibility and benefit sharing

Not applicable.

References

Ahn, S.-M., Kim, T.-H., Lee, S., Kim, D., Ghang, H., Kim, D.-S., ... Kim, S.-J. (2009). The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. Genome Research, 19(9), 1622–1629. doi: 10.1101/gr.092197.109

Akçakaya, H. R., & Sjögren-Gulve, P. (2000). Population Viability Analyses in Conservation Planning: An Overview. Ecological Bulletins, (48), 9–21.

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

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., ... Lippman, Z. B. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell, 182(1), 145-161.e23. doi: 10.1016/j.cell.2020.05.021

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. Genome Biology, 21(1), 30. doi: 10.1186/s13059-020-1935-5

Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., ... Antonelli, A. (2020). A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. Frontiers in Genetics, 10. doi: 10.3389/fgene.2019.01407

Antonides, J., Ricklefs, R., & DeWoody, J. A. (2017). The genome sequence and insights into the immunogenetics of the bananaquit (Passeriformes: Coereba flaveola). Immunogenetics, 69(3), 175–186. doi: 10.1007/s00251-016-0960-8

Ballou, J. D., & Lacy, R. C. (1995). Identifying genetically important individuals for management of genetic variation in pedigreed populations. In J. D. Ballou, M. Gilpin, & T. J. Foose (Eds.), Population management for survival and recovery.

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., & Edwards, D. (2020). Plant pan-genomes are the new reference. Nature Plants, 6(8), 914–920. doi: 10.1038/s41477-020-0733-0

Becker, T., Lee, W.-P., Leone, J., Zhu, Q., Zhang, C., Liu, S., ... Malhotra, A. (2018). FusorSV: An algorithm for optimally combining data from multiple structural variation detection methods. Genome Biology, 19(1), 38. doi: 10.1186/s13059-018-1404-6

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., ... Aury, J.-M. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nature Plants, 4(11), 879–887. doi: 10.1038/s41477-018-0289-4

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., ... Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. Heredity, 119(6), 418–428. doi: 10.1038/hdy.2017.54

Bertolotti, A. C., Layer, R. M., Gundappa, M. K., Gallagher, M. D., Pehlivanoglu, E., Nome, T., ... Macqueen, D. J. (2020). The structural variation landscape in 492 Atlantic salmon genomes. Nature Communications, 11(1), 5176. doi: 10.1038/s41467-020-18972-x

Bickhart, D. M., McClure, J. C., Schnabel, R. D., Rosen, B. D., Medrano, J. F., & Smith, T. P. L. (2020). Symposium review: Advances in sequencing technology herald a new frontier in cattle genomics and genome-enabled selection. Journal of Dairy Science, 103(6), 5278–5290. doi: 10.3168/jds.2019-17693

Brandies, P. A., Wright, B. R., Hogg, C. J., Grueber, C. E., & Belov, K. (2020). Characterization of reproductive gene diversity in the endangered Tasmanian devil. Molecular Ecology Resources, 00, 1–12. doi: 10.1111/1755-0998.13295

Buchanan, J. A., & Scherer, S. W. (2008). Contemplating effects of genomic structural variation. Genetics in Medicine, 10(9), 639–647. doi: 10.1097/GIM.0b013e318183f848

Cagan, A., & Blass, T. (2016). Identification of genomic variants putatively targeted by selection during dog domestication. BMC Evolutionary Biology, 16(1), 10. doi: 10.1186/s12862-015-0579-7

Cam, S. L., Daguin-Thiebaut, C., Bouchemousse, S., Engelen, A. H., Mieszkowska, N., & Viard, F. (2020). A genome-wide investigation of the worldwide invader Sargassum muticum shows high success albeit (almost) no genetic diversity. Evolutionary Applications, 13(3), 500–514. doi: 10.1111/eva.12837

Cameron, D. L., Di Stefano, L., & Papenfuss, A. T. (2019). Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nature Communications, 10(1), 3240. doi: 10.1038/s41467-019-11146-4

Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., & Wellenreuther, M. (2019). The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost Chrysophrys auratus. Molecular Ecology, 28(6), 1210–1223. doi: 10.1111/mec.15051

Chakraborty, M., Emerson, J. J., Macdonald, S. J., & Long, A. D. (2019). Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. Nature Communications, 10(1), 4872. doi: 10.1038/s41467-019-12884-1

Chander, V., Gibbs, R. A., & Sedlazeck, F. J. (2019). Evaluation of computational genotyping of structural variation for clinical diagnoses. GigaScience, 8(9). doi: 10.1093/gigascience/giz110

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding Depression and its Evolutionary Consequences. Annual Review of Ecology and Systematics, 18, 237–268.

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., ... Mardis, E. R. (2009). BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. Nature Methods, 6(9), 677–681. doi:10.1038/nmeth.1363

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., ... Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics, 32(8), 1220–1222. doi:10.1093/bioinformatics/btv710

Chiang, C., Scott, A. J., Davis, J. R., Tsang, E. K., Li, X., Kim, Y., ... Hall, I. M. (2017). The impact of structural variation on human gene expression. Nature Genetics, 49(5), 692–699. doi:10.1038/ng.3834

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., ... Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences, 106(45), 19096–19101. doi: 10.1073/pnas.0910672106

Clarke, A. J., & Cooper, D. N. (2010). GWAS: Heritability missing in action? European Journal of Human Genetics, 18(8), 859–861. doi:10.1038/ejhg.2010.35

Claussen, U. (2005). Chromosomics. Cytogenetic and Genome Research, 111(2), 101–106. doi:10.1159/000086377

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., ... Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. Nature, 464(7289), 704–712. doi: 10.1038/nature08516

Cridland, J. M., Macdonald, S. J., Long, A. D., & Thornton, K. R. (2013). Abundance and Distribution of Transposable Elements in Two Drosophila QTL Mapping Resources. Molecular Biology and Evolution, 30(10), 2311–2327. doi:10.1093/molbev/mst129

Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., & Salzberg, S. L. (1999). Alignment of whole genomes. Nucleic Acids Research, 27(11), 2369–2376. doi:10.1093/nar/27.11.2369

Depledge, D. P., Srinivas, K. P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D. G., ... Wilson, A. C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nature Communications, 10(1), 754. doi:10.1038/s41467-019-08734-9

Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H., & Goldstein, D. B. (2010). Rare Variants Create Synthetic Genome-Wide Associations. PLOS Biology, 8(1), e1000294. doi:10.1371/journal.pbio.1000294

Dohm, J. C., Peters, P., Stralis-Pavese, N., & Himmelbauer, H. (2020). Benchmarking of long-read correction methods. NAR Genomics and Bioinformatics, 2(2). doi: 10.1093/nargab/lqaa037

Dresser, C. M., Ogle, R. M., & Fitzpatrick, B. M. (2017). Genome scale assessment of a species translocation program. Conservation Genetics, 18(5), 1191–1199. doi: 10.1007/s10592-017-0970-6

Dufresnes, C., & Dubey, S. (2020). Invasion genomics supports an old hybrid swarm of pool frogs in Western Europe. Biological Invasions, 22(2), 205–210. doi: 10.1007/s10530-019-02112-8

Ebler, J., Clarke, W. E., Rausch, T., Audano, P. A., Houwaart, T., Korbel, J., ... Marschall, T. (2020). Pangenome-based genome inference. BioRxiv, 37.

Eggertsson, H. P. (2017). Graphtyper enables population-scale genotyping using pangenome graphs. Nature Genetics, 49(11), 11.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics, 11(6), 446–450. doi: 10.1038/nrg2809

Eizenga, J. M., Novak, A. M., Sibbesen, J. A., Heumos, S., Ghaffaari, A., Hickey, G., ... Garrison, E. (2020). Pangenome Graphs. Annual Review of Genomics and Human Genetics, 21(1), 139–162. doi:10.1146/annurev-genom-120219-080406

Ekblom, R., Brechlin, B., Persson, J., Smeds, L., Johansson, M., Magnusson, J., ... Ellegren, H. (2018). Genome sequencing and conservation genomics in the Scandinavian wolverine population. Conservation Biology, 32(6), 1301–1312. doi:10.1111/cobi.13157

Elbers, J. P., Clostio, R. W., & Taylor, S. S. (2017). Population genetic inferences using immune gene SNPs mirror patterns inferred by microsatellites. Molecular Ecology Resources, 17(3), 481–491. doi:10.1111/1755-0998.12591

Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O., & Long, M. (2008). Natural Selection Shapes Genome-Wide Patterns of Copy-Number Polymorphism in Drosophila melanogaster. Science, 320(5883), 1629–1631. doi:10.1126/science.1158078

Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., & Funk, W. C. (2020). Genomic and Fitness Consequences of Genetic Rescue in Wild Populations. Current Biology, 30(3), 517-522.e5. doi:10.1016/j.cub.2019.11.062

Forsdick, N. J., Martini, D., Brown, L., Maloney, R. F., Steeves, T. E., & Knapp, M. (2021). Genomic sequencing confirms absence of introgression despite past hybridisation between a common and a critically endangered bird. BioRxiv, 2020.09.28.316299. doi:10.1101/2020.09.28.316299

Fox, R. A., & Millam, J. R. (2014). Personality traits of pair members predict pair compatibility and reproductive success in a socially monogamous parrot breeding in captivity. Zoo Biology, 33(3), 166–172. doi:10.1002/zoo.21121

Frayling, T. M. (2014). Genome-wide association studies: The good, the bad and the ugly. Clinical Medicine, 14(4), 428–431. doi:10.7861/clinmedicine.14-4-428

Frazer, K. A., Murray, S. S., Schork, N. J., & Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. Nature Reviews Genetics, 10(4), 241–251. doi:10.1038/nrg2554

Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoeppner, M. P., Maloney, R. F., ... Steeves, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. Evolutionary Applications, 13(5), 991–1008. doi:10.1111/eva.12916

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., ... Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nature Genetics, 51(6), 1044–1051. doi:10.1038/s41588-019-0410-2

Giglio, R. M., Ivy, J. A., Jones, L. C., & Latch, E. K. (2018). Pedigree-based genetic management improves bison conservation. The Journal of Wildlife Management, 82(4), 766–774. doi: https://doi.org/10.1002/jwmg.21433

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics, 10(6), 381–391. doi: 10.1038/nrg2575

Golicz, A. A., Batley, J., & Edwards, D. (2016). Towards plant pangenomics. Plant Biotechnology Journal, 14(4), 1099–1105. doi:10.1111/pbi.12499

Gooley, R. M., Tamazian, G., Castaneda-Rico, S., Murphy, K. R., Dobrynin, P., Ferrie, G. M., ... Koepfli, K.-P. (2020). Comparison of genomic diversity and structure of sable antelope (Hippotragus niger) in zoos, conservation centers, and private ranches in North America. Evolutionary Applications, 13(8), 2143–2154. doi:10.1111/eva.12976

Grogan, K. E., Sauther, M. L., Cuozzo, F. P., & Drea, C. M. (2017). Genetic wealth, population health: Major histocompatibility complex variation in captive and wild ring-tailed lemurs (Lemur catta). Ecology and Evolution, 7(19), 7638–7649. doi:10.1002/ece3.3317

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature Communications, 11(1), 1001. doi:10.1038/s41467-020-14803-1

Hajirasouliha, I., Hormozdiari, F., Alkan, C., Kidd, J. M., Birol, I., Eichler, E. E., & Sahinalp, S. C. (2010). Detection and characterization of novel sequence insertions using paired-end next-generation sequencing. Bioinformatics, 26(10), 1277–1283. doi:10.1093/bioinformatics/btq152

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.

doi:10.1111/eva.12149

Hauser, S., Athrey, G., & Leberg, P. (2021). Waste not, want not: Microsatellites remain an economical and informative technology for conservation genetics. Authorea. doi:10.22541/au.161175207.74315618/v1

Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural Variation in the Sequencing Era: Comprehensive Discovery and Integration. Nature Reviews. Genetics, 21(3), 171–189. doi:10.1038/s41576-019-0180-9

Hoelzel, A. R., Bruford, M. W., & Fleischer, R. C. (2019). Conservation of adaptive potential and functional diversity. Conservation Genetics, 20(1), 1–5. doi:10.1007/s10592-019-01151-x

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2020). Population genomics for wildlife conservation and management. Molecular Ecology, 00, 1–21. doi:10.1111/mec.15720

Howe, K., & Wood, J. M. (2015). Using optical mapping data for the improvement of vertebrate genome assemblies. GigaScience, 4(s13742-015-0052-y). doi:10.1186/s13742-015-0052-y

Hurgobin, B., & Edwards, D. (2017). SNP Discovery Using a Pangenome: Has the Single Reference Approach Become Obsolete? Biology, 6(1), 21. doi:10.3390/biology6010021

Huynh, L. Y., Maney, D. L., & Thomas, J. W. (2011). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). Heredity, 106(4), 537–546. doi: 10.1038/hdy.2010.85

Iannucci, A., Makunin, A. I., Lisachov, A. P., Ciofi, C., Stanyon, R., Svartman, M., Trifonov, V. A. (2021). Bridging the Gap between Vertebrate Cytogenetics and Genomics with Single-Chromosome Sequencing (ChromSeq). Genes, 12(1), 124. doi: 10.3390/genes12010124

Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., & McVean, G. (2012). De novo assembly and genotyping of variants using colored de Bruijn graphs. Nature Genetics, 44(2), 226–232. doi: 10.1038/ng.1028

Jay, P., Whibley, A., Frezal, L., Rodriguez de Cara, M. A., Nowell, R. W., Mallet, J., ... Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. Current Biology, 28(11), 1839-1845.e3. doi: 10.1016/j.cub.2018.04.072

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., ... ffrench-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature, 477(7363), 203–206. doi: 10.1038/nature10341

Joron, M., Papa, R., Beltran, M., Chamberlain, N., Mavarez, J., Baxter, S., ... Jiggins, C. D. (2006). A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. PLOS Biology, 4(10), e303. doi: 10.1371/journal.pbio.0040303

Kardos, M., & Shafer, A. B. A. (2018). The Peril of Gene-Targeted Conservation. Trends in Ecology & Evolution, 33(11), 827–839. doi: 10.1016/j.tree.2018.08.011

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology, 37(8), 907–915. doi: 10.1038/s41587-019-0201-4

Kim, K.-W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L. L., Griffith, S. C., ... Slate, J. (2017). A sex-linked supergene controls sperm morphology and swimming speed in a songbird. Nature Ecology & Evolution, 1(8), 1168–1176. doi: 10.1038/s41559-017-0235-2

Kirk, H., & Freeland, J. R. (2011). Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. International Journal of Molecular Sciences, 12(6), 3966–3988. doi: 10.3390/ijms12063966

Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., ... Kempenaers, B. (2017). A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. Nature Ecology & Evolution, 1(8), 1177–1184. doi: 10.1038/s41559-017-0236-1

Kong, S., & Zhang, Y. (2019). Deciphering Hi-C: From 3D genome to function. Cell Biology and Toxicology, 35(1), 15–32. doi: 10.1007/s10565-018-09456-2

Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., ... Snyder, M. (2007). Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. Science, 318(5849), 420–426. doi: 10.1126/science.1149504

Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., & Kamatani, Y. (2019). Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biology, 20(1), 117. doi: 10.1186/s13059-019-1720-5

Kupper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., ... Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. Nature Genetics, 48(1), 79–83. doi: 10.1038/ng.3443

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2020). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. BioRxiv, 678524. doi: 10.1101/678524

Lacy, R. C., Ballou, J. D., & Pollak, J. P. (2012). PMx: Software package for demographic and genetic analysis and management of pedigreed populations: PMx software for pedigree analysis. Methods in Ecology and Evolution, 3(2), 433–437. doi: 10.1111/j.2041-210X.2011.00148.x

Lacy, R. C., & Pollak, J. P. (2021). Vortex: A stochastic simulation of the extinction process (Version 10.5.5). Brookfield, Illinois, USA: Chicago Zoological Society.

Lado, S., Elbers, J. P., Doskocil, A., Scaglione, D., Trucchi, E., Banabazi, M. H., ... Burger, P. A. (2020). Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes. Communications Biology, 3(1), 1–8. doi: 10.1038/s42003-020-1098-7

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., ... Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nature Genetics, 48(1), 84–88. doi: 10.1038/ng.3430

Landguth, E. L., Bearlin, A., Day, C. C., & Dunham, J. (2017). CDMetaPOP: An individual-based, eco-evolutionary model for spatially explicit simulation of landscape demogenetics. Methods in Ecology and Evolution, 8(1), 4–11. doi: 10.1111/2041-210X.12608

Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., ... Liu, S. (2020). Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacbio Sequel II system and ultralong reads of Oxford Nanopore. BioRxiv, 2020.02.13.948489. doi: 10.1101/2020.02.13.948489

Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY: A probabilistic framework for structural variant discovery. Genome Biology, 15(6), R84. doi: 10.1186/gb-2014-15-6-r84

Levy-Sakin, M., Pastor, S., Mostovoy, Y., Li, L., Leung, A. K. Y., McCaffrey, J., ... Kwok, P.-Y. (2019). Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nature Communications, 10(1), 1025. doi: 10.1038/s41467-019-08992-7

Lew, R. M., Finger, A. J., Baerwald, M. R., Goodbla, A., May, B., & Meek, M. H. (2015). Using Next-Generation Sequencing to Assist a Conservation Hatchery: A Single-Nucleotide Polymorphism Panel for the Genetic Management of Endangered Delta Smelt. Transactions of the American Fisheries Society, 144(4), 767–779. doi: 10.1080/00028487.2015.1037016

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., Feng, X., & Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. Genome Biology, 21(1), 265. doi: 10.1186/s13059-020-02168-z

Li, R., Fu, W., Su, R., Tian, X., Du, D., Zhao, Y., ... Jiang, Y. (2019). Towards the Complete Goat Pan-Genome by Recovering Missing Genomic Segments From the Reference Genome. Frontiers in Genetics, 10. doi: 10.3389/fgene.2019.01169

Li, S., Li, B., Cheng, C., Xiong, Z., Liu, Q., Lai, J., ... Yan, J. (2014). Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. Genome Biology, 15(12), 557. doi: 10.1186/s13059-014-0557-1

Liddell, E., Cook, C. N., Sunnucks, P., Liddell, E., Cook, C. N., & Sunnucks, P. (2020). Evaluating the use of risk assessment frameworks in the identification of population units for biodiversity conservation. Wildlife Research, 47(3), 208–216. doi: 10.1071/WR18170

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., ... Dekker, J. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. 326, 6.

Lindsay, E. A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., ... Baldini, A. (2001). Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. Nature, 410(6824), 97–101. doi: 10.1038/35065105

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., ... Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. Cell, 182(1), 162-176.e13. doi: 10.1016/j.cell.2020.05.023

Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. Nature Reviews Genetics, 21(10), 597–614. doi: 10.1038/s41576-020-0236-x

Low, W. Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D. M., ... Williams, J. L. (2020). Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. Nature Communications, 11(1), 2071. doi: 10.1038/s41467-020-15848-y

Lujan, N. K., Weir, J. T., Noonan, B. P., Lovejoy, N. R., & Mandrak, N. E. (2020). Is Niagara Falls a barrier to gene flow in riverine fishes? A test using genome-wide SNP data from seven native species. Molecular Ecology, 29(7), 1235–1249. doi: 10.1111/mec.15406

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., ... Mundlos, S. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell, 161(5), 1012–1025. doi: 10.1016/j.cell.2015.04.004

Lupski, J. R., Wise, C. A., Kuwano, A., Pentao, L., Parke, J. T., Glaze, D. G., ... Patel, P. I. (1992). Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. Nature Genetics, 1(1), 29–33. doi: 10.1038/ng0492-29

Mable, B. K. (2019). Conservation of adaptive potential and functional diversity: Integrating old and new approaches. Conservation Genetics, 20(1), 89–100. doi: 10.1007/s10592-018-1129-9

Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: Challenges and prospects. Nature Reviews Genetics, 10(8), 565–577. doi: 10.1038/nrg2612

Mahmoud, M., Gobet, N., Cruz-Davalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. Genome Biology, 20(1), 246. doi: 10.1186/s13059-019-1828-7

Malo, A. F., Martinez-Pastor, F., Alaks, G., Dubach, J., & Lacy, R. C. (2010). Effects of Genetic Captive-Breeding Protocols on Sperm Quality and Fertility in the White-Footed Mouse1. Biology of Reproduction, 83(4), 540–548. doi: 10.1095/biolreprod.110.085316

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. Nature, 461(7265), 747–753. doi: 10.1038/nature08494

Mauricio, R. (2001). Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. Nature Reviews Genetics, 2(5), 370–381. doi: 10.1038/35072085

McClung, C. E. (1905). The chromosome complex of orthopteran spermatocytes. The Biological Bulletin, 9(5), 304–340. doi: 10.2307/1535568

McHale, L. K., Haun, W. J., Xu, W. W., Bhaskar, P. B., Anderson, J. E., Hyten, D. L., ... Stupar, R. M. (2012). Structural Variants in the Soybean Genome Localize to Clusters of Biotic Stress-Response Genes. Plant Physiology, 159(4), 1295–1308. doi: 10.1104/pp.112.194605

Merritt, J. R., Grogan, K. E., Zinzow-Kramer, W. M., Sun, D., Ortlund, E. A., Yi, S. V., & Maney, D. L. (2020). A supergene-linked estrogen receptor drives alternative phenotypes in a polymorphic songbird. Proceedings of the National Academy of Sciences, 117(35), 21673–21680. doi: 10.1073/pnas.2011347117

Merscher, S., Funke, B., Epstein, J. A., Heyer, J., Puech, A., Lu, M. M., ... Kucherlapati, R. (2001). TBX1 Is Responsible for Cardiovascular Defects in Velo-Cardio-Facial/DiGeorge Syndrome. Cell, 104(4), 619–629. doi: 10.1016/S0092-8674(01)00247-1

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., ... Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature, 585(7823), 79–84. doi: 10.1038/s41586-020-2547-7

Mohiyuddin, M., Mu, J. C., Li, J., Bani Asadi, N., Gerstein, M. B., Abyzov, A., ... Lam, H. Y. K. (2015). MetaSV: An accurate and integrative structural-variant caller for next generation sequencing. Bioinformatics, 31(16), 2741–2744. doi: 10.1093/bioinformatics/btv204

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C.-K. K., ... Edwards, D. (2017). The pangenome of hexaploid bread wheat. The Plant Journal, 90(5), 1007–1013. doi: 10.1111/tpj.13515

Moran, B. M., Thomas, S. M., Judson, J. M., Navarro, A., Davis, H., Sidak-Loftis, L., ... Steiner, C. C. (2021). Correcting parentage relationships in the endangered California Condor: Improving mean kinship estimates for conservation management. Ornithological Applications, (duab017). doi: 10.1093/ornithapp/duab017

Morin, P. A., Archer, F. I., Avila, C. D., Balacco, J. R., Bukhman, Y. V., Chow, W., ... Jarvis, E. D. (2021). Reference genome and demographic history of the most endangered marine mammal, the vaquita. Molecular Ecology Resources, 21(4), 1008–1020. doi: 10.1111/1755-0998.13284

Morrison, C. E., Hogg, C. J., Gales, R., Johnson, R. N., & Grueber, C. E. (2020). Low innate immune-gene diversity in the critically endangered orange-bellied parrot (Neophema chrysogaster). Emu - Austral Ornithology, 120(1), 56–64. doi: 10.1080/01584197.2019.1686994

Murchison, E. P., Schulz-Trieglaff, O. B., Ning, Z., Alexandrov, L. B., Bauer, M. J., Fu, B., ... Stratton, M. R. (2012). Genome Sequencing and Analysis of the Tasmanian Devil and Its Transmissible Cancer. Cell, 148(4), 780–791. doi: 10.1016/j.cell.2011.11.065

Ogden, R., Chuven, J., Gilbert, T., Hosking, C., Gharbi, K., Craig, M., ... Senn, H. (2020). Benefits and pitfalls of captive conservation genetic management: Evaluating diversity in scimitar-horned oryx to support reintroduction planning. Biological Conservation, 241, 108244. doi: 10.1016/j.biocon.2019.108244

Oyler-McCance, S. J., Cornman, R. S., Jones, K. L., & Fike, J. A. (2015). Genomic single-nucleotide polymorphisms confirm that Gunnison and Greater sage-grouse are genetically well differentiated and that the Bi-State population is distinct. The Condor, 117(2), 217–227. doi: 10.1650/CONDOR-14-174.1

Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., ... Scherer, S. W. (2010). Towards a comprehensive structural variation map of an individual human genome. Genome Biology, 11(5), R52. doi: 10.1186/gb-2010-11-5-r52

Parikh, H., Mohiyuddin, M., Lam, H. Y. K., Iyer, H., Chen, D., Pratt, M., ... Salit, M. (2016). svclassify: A method to establish benchmark structural variant calls. BMC Genomics, 17(1), 64. doi: 10.1186/s12864-016-2366-2

Paten, B., Novak, A. M., Eizenga, J. M., & Garrison, E. (2017). Genome graphs and the evolution of genome inference. Genome Research, 27(5), 665–676. doi: 10.1101/gr.214155.116

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B., & Loose, M. (2020). Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. BioRxiv, 2020.02.03.926956. doi: 10.1101/2020.02.03.926956

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadia-Cardoso, A., ... Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. Nature Ecology & Evolution, 3(12), 1731–1742. doi: 10.1038/s41559-019-1044-6

Peters, J. L., Lavretsky, P., DaCosta, J. M., Bielefeld, R. R., Feddersen, J. C., & Sorenson, M. D. (2016). Population genomic data delineate conservation units in mottled ducks (Anas fulvigula). Biological Conservation, 203, 272–281. doi: 10.1016/j.biocon.2016.10.003

Quattrini, A. M., Wu, T., Soong, K., Jeng, M.-S., Benayahu, Y., & McFadden, C. S. (2019). A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. BMC Evolutionary Biology, 19(1), 116. doi: 10.1186/s12862-019-1427-y

Rabier, R., Robert, A., Lacroix, F., & Lesobre, L. (2020). Genetic assessment of a conservation breeding program of the houbara bustard (Chlamydotis undulata undulata) in Morocco, based on pedigree and molecular analyses. Zoo Biology, 39(6), 422–435. doi: 10.1002/zoo.21569

Ralls, K., Sunnucks, P., Lacy, R. C., & Frankham, R. (2020). Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. Biological Conservation, 251, 108784. doi: 10.1016/j.biocon.2020.108784

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. Nature, 592(7856), 737–746. doi: 10.1038/s41586-021-03451-0

Rice, E. S., Koren, S., Rhie, A., Heaton, M. P., Kalbfleisch, T. S., Hardy, T., ... Smith, T. P. L. (2020). Continuous chromosome-scale haplotypes assembled from a single interspecies F1 hybrid of yak and cattle. GigaScience, 9(giaa029). doi: 10.1093/gigascience/giaa029

Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid Zones and the Genetic Architecture of a Barrier to Gene Flow Between Two Sunflower Species. Genetics, 152(2), 713–727. doi: 10.1093/genetics/152.2.713

Robinson, J. A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B. Y., vonHoldt, B. M., Marsden, C. D., ... Wayne, R. K. (2016). Genomic Flatlining in the Endangered Island Fox. Current Biology, 26(9), 1183–1189. doi: 10.1016/j.cub.2016.02.062

Robinson, J. A., Raikkonen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Science Advances, 5(5), eaau0757. doi: 10.1126/sciadv.aau0757

Roelke, M. E., Martenson, J. S., & O'Brien, S. J. (1993). The consequences of demographic reduction and genetic depletion in the endangered Florida panther. Current Biology, 3(6), 340–350. doi: 10.1016/0960-9822(93)90197-V

Rogers, R. L., Cridland, J. M., Shao, L., Hu, T. T., Andolfatto, P., & Thornton, K. R. (2015). Tandem Duplications and the Limits of Natural Selection in Drosophila yakuba and Drosophila simulans. PLOS ONE, 10(7), e0132184. doi: 10.1371/journal.pone.0132184

Sadowski, M., Kraft, A., Szalaj, P., Wlasnowolski, M., Tang, Z., Ruan, Y., & Plewczynski, D. (2019). Spatial chromatin architecture alteration by structural variations in human genomes at the population scale. Genome Biology, 20(1), 148. doi: 10.1186/s13059-019-1728-x

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology, 94(3), 441–448. doi: 10.1016/0022-2836(75)90213-2

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., & Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. Journal of Molecular Biology, 143(2), 161–178. doi: 10.1016/0022-2836(80)90196-5

Savage, J. L., Crane, J. M. S., Team, K. R., & Hemmings, N. (2020). Low hatching success in the critically endangered kākāpō (Strigops habroptilus) is driven by early embryo mortality not infertility. BioRxiv, 2020.09.14.295949. doi: 10.1101/2020.09.14.295949

Schäfer, D., Vincent, H., Fischer, M., & Kempel, A. (2020). The importance of genetic diversity for the translocation of eight threatened plant species into the wild. Global Ecology and Conservation, 24, e01240. doi: 10.1016/j.gecco.2020.e01240

Schwartz, D. C., Li, X., Hernandez, L. I., Ramnarain, S. P., Huff, E. J., & Wang, Y. K. (1993). Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science, 262(5130), 110–114. doi: 10.1126/science.8211116

Scott, P. A., Allison, L. J., Field, K. J., Averill-Murray, R. C., & Shaffer, H. B. (2020). Individual heterozygosity predicts translocation success in threatened desert tortoises. Science, 370(6520), 1086–1089. doi: 10.1126/science.abb0421

Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018). Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. Nature Reviews Genetics, 19(6), 329–346. doi: 10.1038/s41576-018-0003-4

Shanta, O., Noor, A., Chaisson, M. J. P., Sanders, A. D., Zhao, X., Malhotra, A., ... Human Genome Structural Variation Consortium (HGSVC). (2020). The effects of common structural variants on 3D chromatin structure. BMC Genomics, 21(1), 95. doi: 10.1186/s12864-020-6516-1

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. Nature Reviews Genetics, 15(2), 121–132. doi: 10.1038/nrg3642

Soto, D. C., Shew, C., Mastoras, M., Schmidt, J. M., Sahasrabudhe, R., Kaya, G., ... Dennis, M. Y. (2020). Identification of Structural Variation in Chimpanzees Using Optical Mapping and Nanopore Sequencing. Genes, 11(3), 276. doi: 10.3390/genes11030276

Spencer, C. C. A., Su, Z., Donnelly, P., & Marchini, J. (2009). Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. PLOS Genetics, 5(5), e1000477. doi: 10.1371/journal.pgen.1000477

Stranger, B. E., Stahl, E. A., & Raj, T. (2011). Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. Genetics, 187(2), 367–383. doi: 10.1534/genetics.110.120907

Sturtevant, A. H. (1921). Linkage Variation and Chromosome Maps. Proceedings of the National Academy of Sciences of the United States of America, 7(7), 181–183.

Teeling, E. C., Vernes, S. C., Dávalos, L. M., Ray, D. A., Gilbert, M. T. P., & Myers, E. and. (2018). Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. Annual Review of Animal Biosciences, 6(1), 23–46. doi: 10.1146/annurev-animal-022516-022811

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., ... Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proceedings of the National Academy of Sciences, 102(39), 13950–13955. doi: 10.1073/pnas.0506758102

Texeira, J. C., and Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences, 118(10),. doi: 10.1073/pnas.2015096118

Ulahannan, N., Pendleton, M., Deshpande, A., Schwenk, S., Behr, J. M., Dai, X., ... Imielinski, M. (2019). Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure. BioRxiv, 833590. doi: 10.1101/833590

vonHoldt, B. M., Shuldiner, E., Koch, I. J., Kartzinel, R. Y., Hogan, A., Brubaker, L., ... Udell, M. A. R. (2017). Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. Science Advances, 3(7), e1700398. doi: 10.1126/sciadv.1700398

Wala, J. A., Bandopadhayay, P., Greenwald, N. F., O'Rourke, R., Sharpe, T., Stewart, C., ... Beroukhim, R. (2018). SvABA: Genome-wide detection of structural variants and indels by local assembly. Genome Research, 28(4), 581–591. doi: 10.1101/gr.221028.117

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., ... Wang, J. (2008). The diploid genome sequence of an Asian individual. Nature, 456(7218), 60–65. doi: 10.1038/nature07484

Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., ... Wolf, J. B. W. (2020). Discovery and population genomics of structural variation in a songbird genus. Nature Communications, 11(1), 3403. doi: 10.1038/s41467-020-17195-4

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. Trends in Ecology & Evolution, 33(6), 427–440. doi: 10.1016/j.tree.2018.04.002

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., ... Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nature Biotechnology, 37(10), 1155–1162. doi: 10.1038/s41587-019-0217-9

Whibley, A., Kelley, J. L., & Narum, S. R. (2020). The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. Molecular Ecology Resources, 21(3), 641–652. doi: 10.1111/1755-0998.13312

Whitlock, R. (2014). Relationships between adaptive and neutral genetic diversity and ecological structure and functioning: A meta-analysis. The Journal of Ecology, 102(4), 857–872. doi: 10.1111/1365-2745.12240

Wilder, A. P., Palumbi, S. R., Conover, D. O., & Therkildsen, N. O. (2020). Footprints of local adaptation span hundreds of linked genes in the Atlantic silverside genome. Evolution Letters, 4(5), 430–443. doi: 10.1002/evl3.189

Willi, Y., Van Buskirk, J., & Hoffmann, A. A. (2006). Limits to the Adaptive Potential of Small Populations. Annual Review of Ecology, Evolution, and Systematics, 37(1), 433–458. doi: 10.1146/annurev.ecolsys.37.091305.110145

Willoughby, J. R., Fernandez, N. B., Lamb, M. C., Ivy, J. A., Lacy, R. C., & DeWoody, J. A. (2015). The impacts of inbreeding, drift and selection on genetic diversity in captive breeding populations. Molecular Ecology, 24(1), 98–110. doi: 10.1111/mec.13020

Woese, C., Sogin, M., Stahl, D., Lewis, B. J., & Bonen, L. (1976). A comparison of the 16S ribosomal RNAs from mesophilic and thermophilic bacilli: Some modifications in the sanger method for RNA sequencing. Journal of Molecular Evolution, 7(3), 197–213. doi: 10.1007/BF01731489

Wray, N. R., Purcell, S. M., & Visscher, P. M. (2011). Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results. PLOS Biology, 9(1), e1000579. doi: 10.1371/journal.pbio.1000579

Wright, B. R., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., & Grueber, C. E. (2020). A demonstration of conservation genomics for threatened species management. Molecular Ecology Resources, 20(6), 1526–1541. doi: 10.1111/1755-0998.13211

Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., ... Yan, J. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. Nature Genetics, 51(6), 1052–1059. doi: 10.1038/s41588-019-0427-6

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., & Ning, Z. (2009). Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics, 25(21), 2865–2871. doi: 10.1093/bioinformatics/btp394

Yi, K., & Ju, Y. S. (2018). Patterns and mechanisms of structural variations in human cancer. Experimental & Molecular Medicine, 50(8), 1–11. doi: 10.1038/s12276-018-0112-3

Yıldırım, Y., Tinnert, J., & Forsman, A. (2018). Contrasting patterns of neutral and functional genetic diversity in stable and disturbed environments. Ecology and Evolution, 8(23), 12073–12089. doi: 10.1002/ece3.4667

Yoon, S., Xuan, Z., Makarov, V., Ye, K., & Sebat, J. (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Research, 19(9), 1586–1592. doi: 10.1101/gr.092981.109

Yuan, Y., Chung, C. Y.-L., & Chan, T.-F. (2020). Advances in optical mapping for genomic research. Computational and Structural Biotechnology Journal, 18, 2051–2062. doi: 10.1016/j.csbj.2020.07.018

Zarate, S., Carroll, A., Krashenina, O., Sedlazeck, F. J., Jun, G., Salerno, W., ... Gibbs, R. (2018). Parliament2: Fast Structural Variant Calling Using Optimized Combinations of Callers. BioRxiv, 424267. doi: 10.1101/424267

Zhang, G. (2015). Bird sequencing project takes off. Nature, 522(7554), 34–34. doi: 10.1038/522034d

Zhihai, H., Jiang, X., Shuiming, X., Baosheng, L., Yuan, G., Chaochao, Z., ... Shilin, C. (2016). Comparative optical genome analysis of two pangolin species: Manis pentadactyla and Manis javanica. GigaScience, 5(1). doi: 10.1093/gigascience/giw001

Zook, J. M., Hansen, N. F., Olson, N. D., Chapman, L., Mullikin, J. C., Xiao, C., ... Salit, M. (2020). A robust benchmark for detection of germline large deletions and insertions. Nature Biotechnology, 38(11), 1347–1355. doi: 10.1038/s41587-020-0538-8