

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

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

<sup>1</sup>University of Canterbury

<sup>2</sup>Smithsonian Conservat Biol Inst

<sup>3</sup>The University of Sydney

<sup>4</sup>Malaghan Institute of Medical Research

<sup>5</sup>University of Otago

<sup>6</sup>The University of Auckland

August 2, 2021

## Abstract

Structural variants (SVs) are large rearrangements (> 50 bp) within the genome that impact gene function and the content and structure of chromosomes. As a result, SVs are a significant source of functional genomic variation, 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 variation in threatened species via single nucleotide polymorphism (SNP) datasets, SVs remain understudied despite their potential influence on fitness 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 also leverage the existing literature—predominantly in human health, agriculture and eco-evolutionary biology—to identify approaches for readily characterizing SVs and consider how integrating these into the conservation genomics toolbox may transform the way we manage some of the world’s most threatened species.

**Running Head:** Structural variants and conservation genomics

## Structural variants: a new tool in the conservation genomics toolbox

Over the past 10 years, the field of conservation genomics has largely focused on transitioning from a handful of microsatellite markers to thousands of genome-wide single nucleotide polymorphisms (SNPs) (Allendorf, Hohenlohe, & Luikart, 2010; Mable, 2019). Whether characterized using reduced-representation data or whole-genome sequence (WGS) methods, SNP-based estimates of genome-wide variation are being used to better inform conservation by delineating species, including cryptic species (e.g., Binks, Steane, & Byrne, 2021; Quattrini et al., 2019; but see Stanton et al., 2019), detecting hybridization and introgression (Dufresnes & Dubey, 2020; Peters et al., 2016; but see Forsdick et al., 2021; Hauser, Athrey, & Leberg, 2021 preprint), identifying conservation units (Liddell et al., 2020), informing conservation translocations (Dresser, Ogle, & Fitzpatrick, 2017; Glasscock, Grueber, Belov, & Hogg, 2021), guiding conservation breeding programs (e.g., Galla et al., 2020; Wright et al., 2020) and identifying the genomic basis of adaptive traits (Duntsch et al.,

2020). However, SNPs constitute only one component of genome-wide variation (Ho, Urban, & Mills, 2020). Structural variants (SVs) are genomic rearrangements such as insertions, deletions, duplications, inversions and translocations, and are generally defined as  $\geq 50$  bp in length (Figure 1). Variation in chromosome structure was one of the earliest types of genetic variation studied (e.g., McClung, 1905; Sturtevant, 1921), and a recent focus on the impacts of these genomic rearrangements on sequence variation has found that SVs affect more overall genome content and intersect with genes more often than SNPs (Catanach et al., 2019; Chakraborty, Emerson, Macdonald, & Long, 2019; Chiang et al., 2017; Frayling, 2014; Pang et al., 2010).

Structural variants can have significant evolutionary impacts on species. These impacts include genomic incompatibilities between ancestral and novel variants that may influence hybridization and introgression (Weissensteiner et al., 2020) and may even lead to speciation (Davey et al., 2016; Todesco et al., 2020). In addition, emerging evidence indicates that SVs can underlay fine-scale population structure, and may facilitate local adaptation (Cayuela et al., 2021; Dorant et al., 2020; Tigano et al., 2021). These findings—which are generally the result of integrating SNPs and SVs—highlight the opportunity for both variant types to inform conservation. Nevertheless, the broad application of SVs in conservation remains challenging in part due to the uncertainty in how best to detect and genotype all SV types at the population scale (Mérot, Oomen, Tigano, & Wellenreuther, 2020). Further, although SVs are more likely to impact fitness traits than SNPs (Pang et al., 2010), the overall impacts of SVs on population persistence is relatively unexplored (Oomen et al., 2020).

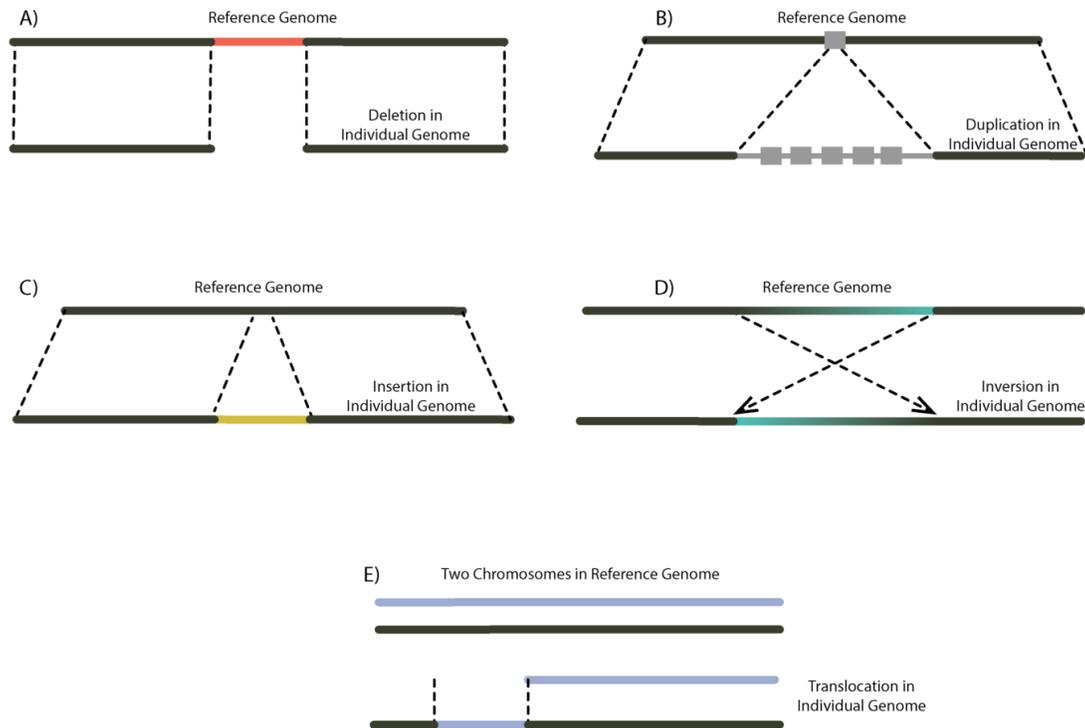


Figure 1: Structural Variant (SV) types and how they differ from a reference genome: A) deletion, where a segment of DNA is not present in an individual, but present in the reference; B) duplication, a rearrangement where there is more than one copy of a particular region of the genome, often in tandem. These can be either intrachromosomal, as shown here, or interchromosomal; 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 one chromosome has been transposed onto another.

Here, we leverage our cross-sector expertise in conservation (AWS, CJH, JG, JRW, KPK, MLL, SJG, TES), human health (DE), agriculture (JG), quantitative genetics (AWS) and bioinformatics (DE, JG, JRW, MLL) to consider the significance of SVs in population persistence and species extinction risk, and discuss opportunities and challenges for relating SVs to fitness traits in small populations. We identify approaches for reliably characterizing SVs with an emphasis on leveraging new and existing genomic resources for species of conservation concern. We also describe how the incorporation of genome assemblies for multiple reference individuals can be used to improve SV discovery and genotyping, even for large and/or complex SVs (Alonge et al., 2020; Eizenga et al., 2020; Gao et al., 2019; Golicz, Batley, & Edwards, 2016; Tettelin et al., 2005). We contend that the study of SVs has the potential to improve conservation outcomes by enhancing our understanding of genome-wide variation and traits related to fitness.

## Small population paradigm and structural variants

Like SNPs, SVs vary within and among individuals, and SV variation is subject to genetic drift, selection, mutation and gene flow, and is impacted by demography, particularly effective population size ( $N_e$ ). Species and/or populations with large  $N_e$  should harbor a higher number of SVs, whereas populations with small  $N_e$  should harbor less (e.g., fewer SV counts in the critically endangered Hawaiian crow versus non-threatened conspecifics; Weissensteiner et al., 2020). Although also influencing the mutation rates and genomic locations of SNP variants, the genome composition and architecture (e.g., overall chromosome repeat content) has a much larger impact on the number, type, location and mutation rate of SVs (Collins et al., 2020). Further, while both SNPs and SVs can be neutral, their size relative to single base substitutions mean that SVs are more likely to overlap gene and regulatory regions and impact fitness. The evolutionary dynamics of SVs in large populations will therefore be determined by the interplay between the functional effects of these SVs, if any, and variation in mutation rates of different SV sizes and types, which may lead to enrichment of SVs of particular types and/or in particular genomic regions (Conrad & Hurles, 2007).

In contrast, populations of conservation concern are likely to be fragmented or chronically small, and the evolutionary dynamics of both SNP and SV variation will be dominated by genetic drift and founder effects that interplay with the small  $N_e$  to both decrease the overall level genetic variation and reduce the efficacy of natural (purifying) selection (Henn, Botigué, Bustamante, Clark, & Gravel, 2015). Decreasing genetic variation leads to further reductions in  $N_e$  which increases the probability of extinction through both the loss of adaptive variants and the increase in frequency of deleterious variants. This can lead to a downward spiral of further loss of diversity and increasing extinction risk, a process termed an ‘extinction vortex’ (Gilpin & Soule 1986). Further, inbreeding is unavoidable in small populations, leading to higher frequencies of deleterious recessive individuals in the population, with this inbreeding depression leading to a further fitness burden on the threatened species (Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016; Stoffel, Johnston, Pilkington, & Pemberton, 2021).

While both SNPs and SVs can have deleterious fitness effects, their size and potential to overlap gene regions mean that SVs may be enriched in the pool of deleterious variants that are maintained in small populations. While SVs of very large deleterious effect may be purged from the population, small populations may selectively accumulate slightly deleterious SVs. This enrichment indicates that SVs likely underpin some negative fitness traits associated with inbreeding depression. Although there are some clues as to the evolutionary dynamics of SVs in small populations, further study is critical for resolving the significance of SVs in population persistence and species extinction risk.

## Relating structural variants to fitness traits

Advancements in WGS technologies coupled with increased computational capacities, have renewed interest in the role of SVs in determining trait differences (Chiang et al., 2017; Pang et al., 2010; Sadowski et al.,

2019; Yi & Ju, 2018). For example, characterizing SVs in agricultural species has led to the identification of variants associated with economically significant traits in crops like grapevine (Zhou et al., 2019), maize (Yang et al., 2019), soybean (Liu et al., 2020) and tomato (Alonge et al., 2020), and identified specific genes and gene regions associated with domestication in vertebrates (Bertolotti et al., 2020; Cagan & Blass, 2016; vonHoldt et al., 2017). The impact of these SVs on trait variation may be direct or indirect. For example, SVs can lead to direct changes to gene coding regions and gene regulation that change the function or expression of proteins (Derek M. Bickhart & Liu, 2014; Collins et al., 2020). In contrast, in some cases where a particular SV is associated with a trait, the SV may not be the causal variant, but rather increase the likelihood of *de novo* causal variants nearby. 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). There is also growing evidence that SVs can further impact the gene regulatory landscape by altering the formation of topologically associating domains (TADs), genomic regions that physically interact with themselves more frequently than with regions elsewhere (Sadowski et al., 2019; Shanta et al., 2020).

There is growing evidence that SVs—in particular, the suppression of recombination and subsequent evolution of ‘supergenes’, or tightly linked co-adapted alleles—can impact fitness traits in natural populations (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). Moreover, it is notable that this inversion is likely a lethal recessive variant (Lamichhaney et al., 2016). Similarly, a large inversion resulting in a supergene underlies significant morphological and behavioral differences among white-striped and tan-striped morphs in White-throated sparrows, with aggressiveness being monogenic in the white-striped morph (Merritt et al., 2020). Although many fitness traits are likely to be polygenic, including traits of conservation interest such as 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), those impacted by supergenes are likely to have relatively simple inheritance patterns which will enable their characterization and management.

To date, there are generally two approaches to investigate the genomic basis of traits, both of which require clearly defined traits and well-curated data sets and high-quality, well annotated, reference genomes: 1) Comparative genomics, where well-characterized groups (individuals, 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 traits are assessed (Chakraborty et al., 2019). For example, association studies and/or 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, *Heliconius* butterflies, Sunflowers; Berg et al., 2017; Joron et al., 2011, 2006; Rieseberg, Whitton, & Gardner, 1999; Weissensteiner et al., 2020).

Whole-genome SNP-based association studies have been extensively applied to model organisms, agriculturally significant species and humans, although they are increasingly applied to wild and non-model species (Santure & Garant, 2018). However, for many SNP-based association studies, significantly associated 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 often attributed to genetic variation not explained by genotyped loci, that is to say that causal variants may remain uncharacterized (Manolio et al., 2009). Another hypothesis to explain this missing heritability is that SVs are a significant source of trait variation (Eichler et al., 2010; Frazer, Murray, Schork, & Topol, 2009). Further, although a given SV is likely to be linked to a nearby SNP (Wilder, Palumbi, Conover, & Therkildsen, 2020), they may not be in strong linkage disequilibrium with each other, so that the SNP will not capture the impact of the SV on the trait (Pang et al., 2010). Structural variants may also lead to challenges in aligning reads and calling SNPs in a region, in

turn preventing the identification of SVs from sequence data (Pang et al., 2010). The inclusion of SVs into association studies is therefore likely to lead to better power to detect causative variants, particularly given advances in GWAS methods, including machine-learning approaches, that can more effectively prioritize variants in a region and account for non-additive interactions (Briec, Ono, Drinan, & Naish, 2015; Ramzan, Gültas, Bertram, Cavero, & Schmitt, 2020). This is particularly promising as evidence suggests that SVs have larger phenotypic effects and may be deleterious more often 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 locus (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).

## Structural variant discovery and genotyping in threatened species

### A high-quality reference genome is an invaluable investment

While linkage mapping and cytological techniques can be used to characterize SVs (Deakin et al., 2019), here we focus on the approaches most likely to be accessible to the conservation genetics community. Partnerships with global genome consortia (e.g., Vertebrate Genomes Project, Rhie et al., 2021; Earth BioGenome Project, Lewin et al., 2018; Nature and Zoonomia Consortium, Genereux et al., 2020) are providing increased accessibility to highly contiguous, well-annotated genome assemblies generated from both short and long read data and long-range scaffolding approaches for many species, including those of conservation concern 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, well annotated reference genome, nor do we recommend this action for all threatened species. Where there is a clear hypothesis that SVs may be impacting a particular gene or region, comparative approaches (Alonge et al., 2020; Weissensteiner et al., 2020), and target capture methods (vonHoldt et al., 2017) may be the most cost-effective, especially if a reference genome from a closely related species can be leveraged instead. However, as described below, investment in a high-quality reference genome is a near necessity for accurate and precise SV discovery and genotyping, particularly if the goal is to characterize a broad range of SV types for downstream functional analyses (e.g., investigating the genomic basis of traits linked to fitness). Further, if the goal is to characterize a large proportion of SVs—including large and/or complex SVs—across the genome, integration of multiple reference genomes provide a powerful tool for accurately characterizing genome-wide variation and may facilitate genotyping across multiple variant classes (i.e., both SNPs and SVs; Figure 2).

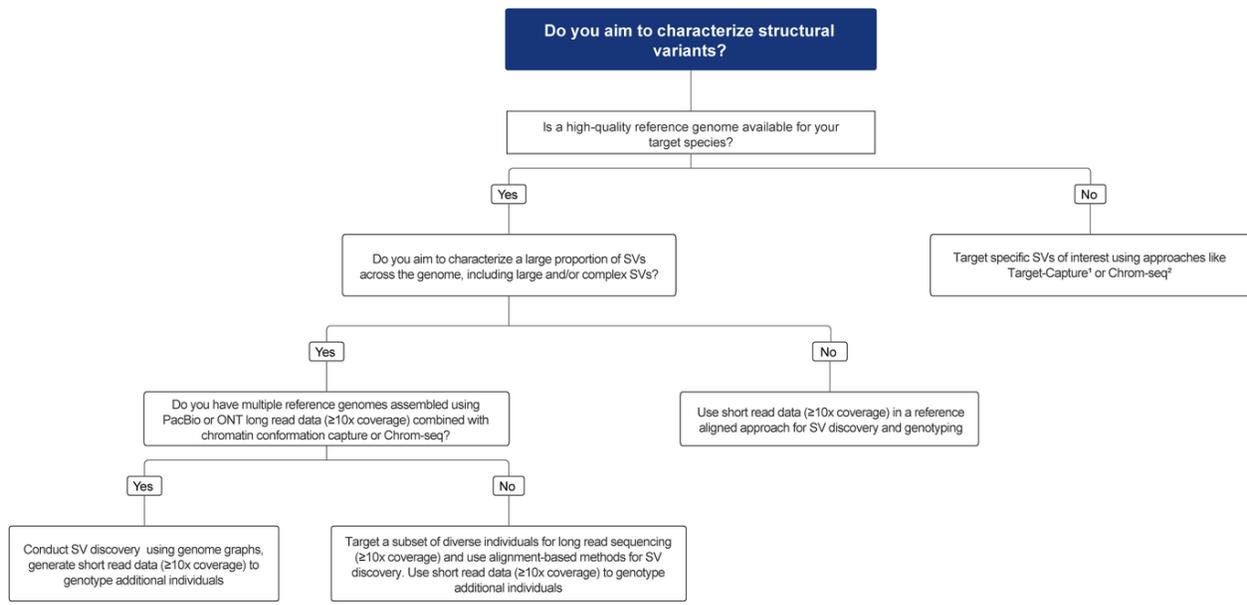


Figure 2: Decision tree for characterizing structural variants (SVs), including large and/or complex SVs. See text for details. Chromatin conformation capture methods include Hi-C, Omni-C, and Pore-C. PacBio = Pacific Biosciences; ONT = Oxford Nanopore Technologies. <sup>1</sup>Choi et al., (2009); <sup>2</sup>Iannucci et al., (2021).

## Alignment-based SV discovery methods

Once a high-quality reference genome has been assembled for a species of interest, SVs can be detected using alignment-based methods where short- and/or long read sequence data for multiple individuals are aligned to the reference without prior assembly. These 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; Korbelt 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. Split-reads identify regions where alignments map across a breakpoint, which 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 relied on 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 only a few SV types and tend to underperform against generalist programs that incorporate at least three algorithms for SV detection (e.g., Delly, Lumpy, Manta, SvABA; Chen et al., 2016; Layer, Chiang, Quinlan, & Hall, 2014; Rausch et al., 2012; 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), especially when data is derived from short-read sequencing methods (Figure 3; see more below). In order to address this error, an ensemble approach for SV characterization has been extensively applied in human data and in domesticates (Du et al., 2021; Ho et al., 2020; Zhou et al., 2019). With an ensemble approach, multiple SV callers are integrated into a single pipeline as a means to create more certainty around SV discovery, with only variants that intersect

multiple SV callers retained (Becker et al., 2018; Mohiyuddin et al., 2015; Zarate et al., 2018 preprint). Further, because validated SV call sets have been developed for humans, it is straightforward to benchmark appropriate program combinations for SV discovery (Collins et al., 2020; Ho et al., 2020; Parikh et al., 2016; Zook et al., 2020).

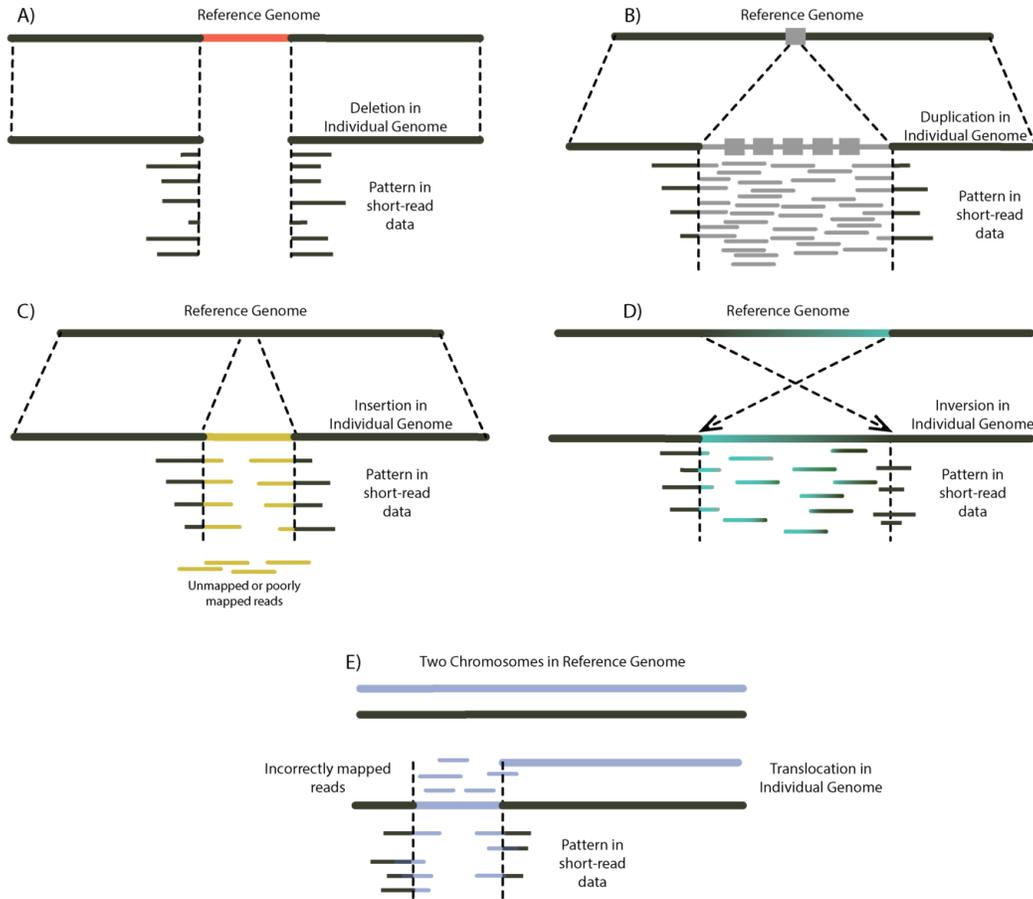


Figure 3: Structural Variant (SV) types and common problems associated with short-read sequence data: A) deletion, called when reads do not map to, and/or are split across, a given region on the reference genome. Generally the most straightforward SV to detect with short-read data, but complex rearrangements 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 or different SV types 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), unless large sample sizes are used (Collins et al., 2020).

When validated SV call sets are lacking, we recommend using a generalist SV discovery program that combines multiple SV discovery algorithms to target a range of SVs at once. These generalist programs enable a combination of assembly-based methods alongside read depth, read pairs and/or split-read approaches (i.e., GRIDSS, Cameron et al., 2017; Manta, Chen et al., 2016). 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). Using a generalist program has an additional benefit that it overcomes the challenge faced by many ensemble methods, where distinguishing between true variants and false positives is difficult due to the significant overlap in false positives across methods (Cameron et al., 2019). However, conservation geneticists should be aware that the computational resources required to characterize SVs at the population scale using generalist SV discovery programs are substantial (e.g., for a diploid 1.15Gb genome with 170 individuals sequenced to [?]25x coverage, 72 physical cores, 460 Gb RAM, >3 Tb storage was required to implement paired-, split-read and assembly-based SV discovery algorithms; JRW personal observation).

## Multiple reference genomes improve genome-wide structural variant discovery and genotyping

A pangenome is the aggregate characterization of genomic variation present in a group of interest, including species and populations (e.g., variation 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 trait diversity in humans (Pang et al., 2010) and agriculturally significant species (e.g., cattle, goats, soybean, and maize; D. M. Bickhart et al., 2020; Della Coletta, Qiu, Ou, Hufford, & Hirsch, 2021; Golicz et al., 2016; 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; Song et al., 2020; Soto et al., 2020; Weissensteiner et al., 2020; Zhou et al., 2019), 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; Li, 2018; Rakocevic et al., 2019; Tettelin et al., 2005) that are encompassed in analyses of copy number variation (CNV) and presence-absence variation (PAV) (Della Coletta et al., 2021). 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 resolve complex variants (i.e., multiple overlapping events) that may otherwise go undetected in alignment-based approaches, hampering the discovery of causal variants (Alonge et al., 2020; McHale et al., 2012).

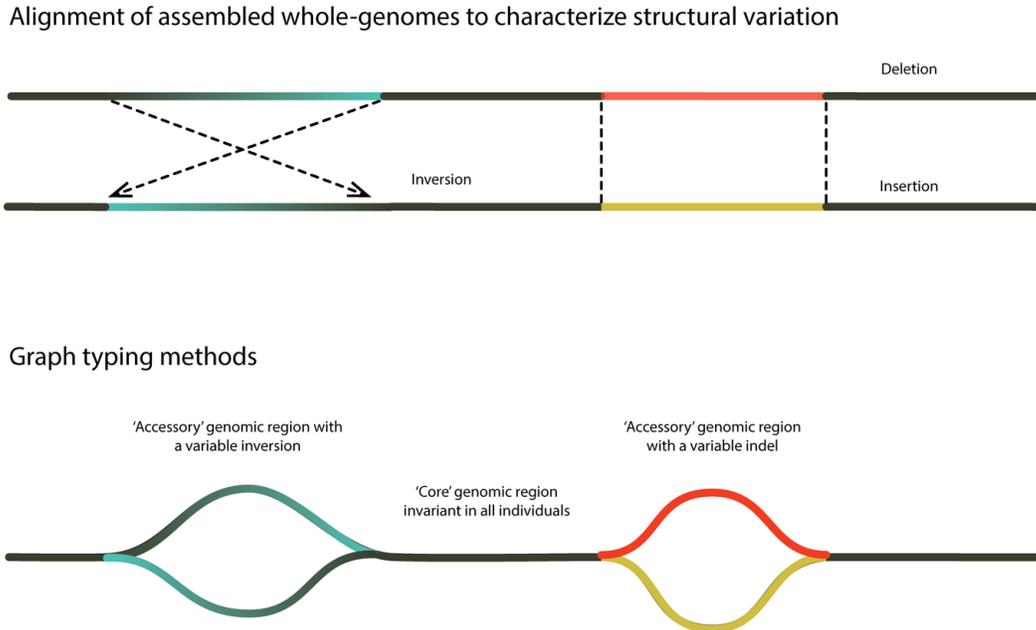


Figure 4: Characterization of SVs using assembled whole-genomes and genome graphs using multiple reference genomes: (a) Schematic of structural variant (SV) discovery using alignment of assembled whole genomes. (b) Schematic of genome graphs, which can then be used to characterize a pangenome and facilitate 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. The use of multiple reference genomes can reduce reference bias and are better able to capture insertions, inversion haplotypes and complex SVs (i.e., genomic regions where multiple SVs occur in close proximity).

For species of conservation concern, individuals selected for reference genomes may not be the most representative in that some are chosen to showcase individuals known to the public (e.g., animal ambassadors) or are the result of opportunistic sampling. Regardless, in most cases, it is unlikely that a single individual accurately represents the genomic variation within the species as a whole, even in highly inbred species (Gao et al., 2019; McHale et al., 2012). 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 crop species such as soy, wheat and *Brassica* spp. where many more SVs are resolved, leading to a paradigm shift in crop improvement (Della Coletta et al., 2021; Golicz et al., 2016; Liu et al., 2020; Montenegro et al., 2017; Song 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. In lieu of a conventional pangenome, whereby all genomic variation is captured, the targeted sequencing and assembly of multiple representative individuals to construct genome graphs may provide significant improvements over the use of a single reference genome of comparable assembly quality (Figure 2). For a related approach see Tigano et al., (2021).

## SV discovery and genotyping with short-read sequence data

Moderate coverage (e.g., ~10x) short-read sequencing is becoming the predominant approach in conservation genomics (Cam et al., 2020; Galla et al., 2020; Lado et al., 2020; Lew et al., 2015; Lujan, Weir, Noonan,

Lovejoy, & Mandrak, 2020; Oyler-McCance, Cornman, Jones, & Fike, 2015, p.; Robinson et al., 2016). Moderate coverage datasets are generally cost effective and appropriate for characterizing SNPs using a single reference genome, whereas a minimum of 30x coverage is often recommended for de novo SV discovery and genotyping (Ahn et al., 2009; Kosugi et al., 2019; Sims, Sudbery, Iltott, 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, 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 average read depth to 30x, the likelihood of distinguishing between sequencing error, read mapping errors and true genomic variation increases as well. However, due to the increased costs, such high coverage short-read datasets may be challenging to generate for many species of conservation concern. The question then becomes how best to fully utilize moderate coverage short-read datasets to investigate SVs. Recent studies demonstrate that an average read depth of 10x is sufficient for population-scale comparisons, provided sampling is representative and a high-quality reference genome is available (Collins et al., 2020; Du et al., 2021; Zhou et al., 2019). Regardless of coverage, it is important to note that genotypes called in short-read SV discovery programs are prone to high error rates (Chander, Gibbs, & Sedlazeck, 2019). One strategy to alleviate this is to conduct SV discovery across all individuals to establish a SV call set and then use dedicated programs for SV genotyping (e.g., BayesTyper, SVTyper, Paragraph)(Chander et al., 2019). In addition, SV genotyping programs have variable performance across a range of SV sizes and types. As such, many programs are unable to genotype across the whole range of SVs that may have been called during discovery. A final consideration is that genotyping programs are still an area of active development. Although they may alleviate some of the false discovery rates prevalent in SV discovery programs, they do not provide a definitive solution (Chander et al., 2019). 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 (e.g., Patel et al., 2014; Pilipenko et al., 2014). This is particularly relevant for intensively managed threatened species, as pedigree relationships are often known, and suitable trios can be identified for sequencing. Alternatively, if sufficient resources are available, the sensitivity of long-read sequence data may be leveraged multiple reference genomes approach for SV discovery and may facilitate the use of lower coverage short-read resequencing data for SV genotyping, but this largely remains untested. Target-capture methods for genes or regulatory regions of interest can also provide an affordable alternative to WGS for species with large genomes and projects with tight budgets (Andermann et al., 2020). For example, vonHoldt et al., (2017) used a targeted-sequencing approach combined with PCR validation to characterize a region on chromosome 6 under positive selection in domestic dog breeds, which significantly reduced sequence coverage requirements.

## SV discovery and genotyping beyond short-read sequence data

When characterizing genomic features, especially SVs, 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. Further, when PacBio sequencing providers are overseas, ONT offers the advantage that data can be generated in-house, which may, for example, be more responsive to the needs

and aspirations of Indigenous Peoples (Collier-Robinson, Rayne, Rupene, Thoms, & Steeves, 2019; Galla et al., 2016) and/or alleviate the need for export permits. Despite the 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, well annotated reference genome, this improves confidence in read mapping across the genome (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).

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 organization 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 organization of chromatin across genomes (Kong & Zhang, 2019 for review) and have 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. This information is particularly beneficial for ‘non-model’ species where gene annotations, regulatory element annotations and gene regulatory networks are un- or poorly- characterized, leading to challenges in predicting the impacts of SVs on gene expression. For example, 3C-based approaches combined with SV detection and analysis have revealed the molecular basis for human developmental diseases previously diagnosed through karyotypes (Melo et al., 2020) and resolved the mechanisms underlying reproductive disorders in goats (Guang-Xin et al., 2021).

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 utilizes 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 (Rhie et al., 2021), 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	Multiple reference genomes	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. Multiple reference genomes 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 weight <sup>a</sup>	High to ultra-high molecular weight <sup>a</sup>	Ultra-high molecular weight <sup>a</sup>
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	[?] 10x	[?] 10x	1. [?] 10x to generate multiple reference genomes 2. [?] 10x for population-level genotyping	> 30x
Method of SV discovery	Alignment-based, including read-pair, read depth, split-reads and local assembly	Alignment-based with local assembly	1. Alignment of multiple <i>de novo</i> assembled whole reference 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 (multiple references 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

## Concluding remarks

Structural variants provide an exciting opportunity to complement SNP-based approaches and expand our understanding of genome-wide variation. However, there remains much to learn about the integration of SVs in the small population paradigm. In the meantime, mounting evidence indicates that SVs play an important role in determining fitness traits for both model and non-model species. Whereas SNPs impact single nucleotides, SVs can capture multiple genes and gene regions simultaneously and are therefore more likely to impact fitness. Emerging approaches for SV discovery and genotyping, and relating SVs to fitness traits, are now in reach for threatened species, and we are confident that discoveries related to the role of SVs in determining fitness traits of conservation interest are imminent. Moreover, the development of SV-specific tools for analyses of within and between population variation will further enhance the utility of SVs in species of conservation concern (e.g., SV-Pop; Ravenhall, Campino, & Clark, 2019).

However, developing strategies to manage fitness traits will be challenging, as selection for or against them can lead to the reduction of genome-wide diversity (Kardos et al., 2021 preprint; Kardos & Shafer, 2018). Despite this, current management practice can include such actions. For example, individuals that carry, or express, negative fitness traits can be assessed in conservation breeding programs, followed by appropriate actions or recommendations (e.g., Attwater's prairie chicken; Grueber, Hogg, Ivy, & Belov, 2015; Hammerly, Morrow, & Johnson, 2013). Alternatively, when inheritance of these traits are known, individuals may be screened prior to selective pairing (e.g., California Condor; Moran et al., 2021). Regardless, as we learn more about the mechanisms underlying fitness traits, especially traits associated with negative fitness, there is a pressing need to establish comprehensive evidence-based management strategies that mitigate the impact of these traits while minimizing the loss of genome-wide diversity. Where we see the most promise for this is the emerging practice of integrating WGS data into individual-based models, which will enable different management actions to be modelled and compared (Seaborn et al., 2021). Such models will allow researchers to forecast the consequences of selection-based management of fitness traits underpinned by adaptive or deleterious SVs on genome-wide diversity, thus better informing 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 establish the most effective models possible (Hohenlohe, Funk, & Rajora, 2020).

## Glossary

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

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

**Complex structural variants:** An instance where multiple SV types 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:** A chromosomal rearrangement where there is more than one copy of a particular region of the genome. This can be anywhere from [?]50bp to whole chromosomes and even the whole genome (whole genome 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.

**Genome graph:** A representation of multiple possible sequences.

**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.

**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):** A genomic rearrangement [?]50bp in length. Common classes of structural variants include deletions, duplication, insertions, inversions and translocations.

**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. Retrieved from <https://www.jstor.org/stable/20113245>

- 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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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. Retrieved from <https://www.jstor.org/stable/2097132>
- 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: <https://doi.org/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: <https://doi.org/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., Hoepfner, 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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/10.1111/eva.12976>
- Grogan, K. E., Sauther, M. L., Cuzzo, 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: <https://doi.org/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
- Harrison, 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: <https://doi.org/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: <https://doi.org/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., ... Kempnaers, 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., Duskocil, 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., Hoepfner, 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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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., Chuvén, 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: <https://doi.org/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., Raikonen, 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., Illott, 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. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1084843/>
- 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., Maignani, 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
- Teixeira, 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., Bandopadhyay, 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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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