A Beginner's Guide to Structural Variants in Eco-Evolutionary Population Genomics: Everything You Wanted to Know

Preprint · June 2025

DOI: 10.22541/au.174853973.36642913/v2

7 authors, including:

Katarina C Stuart
Macquarie University
29 PUBLICATIONS **261** CITATIONS
SEE PROFILE

Rebekah A. Oomen
University of New Brunswick
107 PUBLICATIONS **1,550** CITATIONS
SEE PROFILE

Anna Tigano
University of British Columbia - Okanagan
39 PUBLICATIONS **1,600** CITATIONS
SEE PROFILE

Maren Wellenreuther
University of Auckland
195 PUBLICATIONS **4,827** CITATIONS
SEE PROFILE

A Beginner's Guide to Structural Variants in Eco-Evolutionary Population Genomics: Everything You Wanted to Know

Katarina Stuart[1,2], Rebekah Oomen[3,3,4], Anna Tigano[5], Maren Wellenreuther[2,6], Jana Wold[7], David L. Field[1], and Claire Mérot[8]

[1]Applied Biosciences, Macquarie University, Sydney, Australia
[2]School of Biological Sciences, University of Auckland, Auckland, New Zealand
[3]University of New Brunswick Saint John, Saint John, New Brunswick, Canada
[4]Centre for Coastal Research, University of Agder, Kristiansand, Norway
[5]Department of Fisheries and Oceans, Nanaimo, Canada
[6]The New Zealand Institute of Plant and Food Research, Nelson, New Zealand
[7]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand
[8]CNRS, UMR6553 ECOBIO, Université de Rennes, Rennes, France

June 11, 2025

1. Introduction

Characterizing genomic variation is fundamental to address many ecological and evolutionary questions. The 'evolution' of DNA sequencing methods over time has enabled the identification of new aspects of genetic diversity at every step. In particular, ecological and evolutionary genomics have flourished with the increased availability of high quality reference genomes (Formenti et al., 2022), and attainability of whole genome sequencing (WGS) (Fuentes-Pardo & Ruzzante, 2017). Resequencing entire genomes, rather than a small portion through reduced-representation approaches, provides a rich source of information and has led to a proliferation of methods to investigate evolutionary and demographic processes. We can now identify signatures of balancing selection in the genome (Stern & Lee, 2020), reconstruct demographic history in the near and distant past with unprecedented resolution (Nadachowska-Brzyska et al., 2015; Santiago et al., 2020), and characterize the roles of genome structure and recombination in the levels and distribution of genomic variation across the genome (Akopyan et al., 2025; Tigano et al., 2021). Many genomic analysis methods are currently catered to SNP variation, and WGS in particular has greatly expanded our view of genome wide variability within and between species.

The increasing accessibility and coverage of WGS data has also enabled the direct identification of larger genetic variants, known as structural variants (SVs) (Alkan et al., 2011), enabling a deeper understanding of their role in ecology and evolution (Mérot, Oomen, et al., 2020). The growing breadth of population level SV studies has quickly revealed both the ubiquity and magnitude of SVs' contributions towards intraspecific genomic diversity across species. A higher proportion of the genome is covered by SVs than SNPs (e.g. 5x (Mérot et al., 2023), 3x (Catanach et al., 2019), 8x (Hämälä et al., 2021)). SVs can also strongly affect fitness traits (e.g., 24% increased heritability estimates (Zhou et al., 2022)). Continued research into SVs will undoubtedly reveal more about the fundamental mechanisms of evolution across the tree of life. The inclusion of SVs into research fields that have previously focused heavily on SNPs will aid with the interpretation of complex genomic patterns and processes (Dallaire et al., 2023; Oomen et al., 2020) and a more complete picture of intraspecific and interspecific genetic variation.

Because of this growing appreciation of SVs, researchers are increasingly interested in reanalysing existing

WGS datasets or obtaining new data to examine SVs in their species of interest. This can be a daunting task because of the genetic resources required, as well as the technical and biological complexity of SV data analysis and interpretation. Furthermore, SVs are an extremely diverse category of variants, and analysing SVs like SNPs - as a single group - limits insights from their diverse subtypes and complex roles in genetic diversity.

In this review, we aim to answer practical questions for those new to the study of SVs, guide study design and analytical best practices for those seeking to analyse population level SVs within eco-evolutionary studies, and suggest future avenues of inquiry. First, we summarise the differences between SVs and SNPs, as well as between diverse types of SVs, and discuss how these specific properties may interact with eco-evolutionary processes. We then focus on the practical side of SV analysis, providing a framework for identifying and analysing SVs from WGS data. Accompanying the global movement providing high quality reference genomes (e.g. Ebenezer et al., 2022; Lewin et al., 2022; Mc Cartney et al., 2024), this review aims to reduce barriers to analyzing the full spectrum of genomic variation by incorporating SVs in eco-evolutionary studies.

2. How do we define and classify SVs?

'Structural Variant' is a broad term that encompasses all variants other than single nucleotide variants (SNV, which include SNPs), including whole chromosome and genome duplications (Scherer et al., 2007). They are generally defined as variation in the presence, absence, number, orientation, or position of a sequence. Some studies define a variant as 'structural' if they affect more than a minimum defined length, typically 50 bp. This threshold-based definition is inherited from SV detection software. Earlier methods for calling small SVs termed insertions-deletions ('INDELs') based on short reads ignored variants longer than this threshold, a length range that SV detection software now often targets. However, these variants exist along a continuous spectrum of length that extends from 2+ bp to many megabases in length (Mérot, Oomen, et al., 2020) and any length threshold is arbitrary and limits capturing the diversity of evolutionarily relevant SVs (Recuerda & Campagna, 2024). Practically, the length range of SVs within a study is ultimately dictated by the limits of variant calling algorithms and input data, with longer reads allowing for detection of larger SVs (Mahmoud et al., 2019). Therefore, a good practice is to clearly state the length range targeted by a study, particularly when it guides the definition of 'SV' therein.

SVs are also defined by their sequence change relative to a reference genome. This usually includes deletions (DEL), insertions (INS), duplications (DUP), inversions (INV), fusions (FUS), and translocations (TRA) (Alkan et al., 2011). While this categorization is meaningful from a bioinformatic perspective, other intrinsic characteristics may be more relevant from a biological point of view, such as how they originate and evolve over time. SVs can evolve by many mechanisms, including errors in meiotic recombination like incomplete crossover (e.g., due to age or toxins), improper DNA repair, or replication issues like template switching or slippage (Carvalho & Lupski, 2016; Currall et al., 2013). SVs can also be caused by the activity of transposable elements (TEs), which are repetitive genetic elements that can originate from the genome itself or from external viruses, and have the ability to move and replicate themselves within the genome (Bourque et al., 2018). When a TE replicates or moves itself within the host genome, it is creating genetic variation, and thus this TE will be an SV. However, the presence and nature of repeat sequences (which include TEs) can impact SV formation processes. TE replication, for example, results in different regions of the genome having similar sequences which can induce non-allelic homologous recombination and the formation of larger rearrangements (Klein & O'Neill, 2018). This mechanism is supported by the presence of segmental duplications and TEs at the breakpoints of large SVs (Harringmeyer & Hoekstra, 2022; Meyer et al., 2024). However, even when SVs and TEs are closely associated within a genomic region, it is not always clear whether TEs are involved in the formation of a SV, as large and complex SVs may also accumulate TEs regardless of their formation mechanism (Jay et al., 2021; Munasinghe et al., 2023).

While the activity of TEs generates SVs, TEs are defined by their specific sequence identity (i.e., the nucleotide sequence that defines a variant, which distinguishes them from surrounding DNA). The sequences of TEs evolve and diversify alongside their host genomes: TEs are generally phylogenetically grouped into 'families' sharing similar sequences, with sequence divergence within a TE family being an indicator of divergence

time (Bourque et al., 2018). TEs can become fixed over time due to selection or drift within the population or species. At this point they are no longer a polymorphic variant, so they are not considered SVs despite their sequence still identifying them as a TE. Thus, not all SVs are TEs, and likewise, not all TEs are SVs.

In this review, we use 'SV' to refer collectively to structural variants of all types and lengths, including those of TE and non-TE origin, unless otherwise specified. Note that many studies focus on specific subsets of SVs, using different terms such as copy number variation (CNV), INDELs (insertions-deletions), presence-absence variation (PAV), chromosomal rearrangements (CRs, e.g. inversions, translocations, fusions, usually longer than 100s of kb), or microsatellites (which are CNVs, which are in turn INDELs). Similarly, TEs are referred to as transposons, jumping genes, mobile genetic elements (MGEs), mobile DNA, retrotransposons or DNA transposons. Overlooking this diverse terminology may lead to important research being missed.

2.1 Why do we study SV diversity when we already have genome wide SNP data?

One might question whether identifying SVs is necessary—can SNP sequence variation alone capture the patterns of genetic variation necessary for genomic analysis? Broadly, relative patterns of genetic diversity between individuals and populations (e.g. population structure), when calculated using both SNPs and SVs often correlate (Tigano et al., 2024; Tigano & Russello, 2022) although differential patterns are sometimes observed (Dorant et al., 2020; Tigano et al., 2024). Moreover, SVs can influence population structure (Tepolt et al., 2022) and, by underlying ecotypes, large rearrangements may be important to account for in species management (Oomen et al., 2020).

Much of the theory and analytical approaches used in population genomics has been developed around SNPs, and expanding the genomic toolkit to include SVs is still in its infancy (e.g. Barton & Zeng, 2018). This is particularly pertinent for linkage disequilibrium (LD) studies (e.g. inferring recombination landscapes, effective population size, etc) and for quantitative trait loci (QTL) mapping, where SNPs remain ideal due to their dense and genome-wide distribution. Because SVs, being larger, encompass more base pair changes overall, limiting the population genomic toolkit to SNP-based analysis will misrepresent genetic diversity, and in particular will underestimate the occurrence of strong effect variants within the genome. SVs can be the genomic basis of discrete morphotypes (Lamichhaney et al., 2016) and ecotypes (Li et al., 2024), and underlie many human diseases, likely contributing to the missing heritability issue (i.e. where trait heritability estimates are much lower than expected when calculated using SNPs only) (Groza, Chen, et al., 2024). An increasing number of studies on commercially-relevant species demonstrate that SVs underlie traits of economic interest (Jayakodi et al., 2020; Leonard et al., 2024). Therefore, important sources of genomic variation will be missed in ecology, evolution, and in applied research if SVs are ignored.

One may wonder to what extent the effect of SV can be predicted from neighbouring SNPs. This approach may be cost-efficient in some cases such as well-documented catalogues of variants (Blaj et al., 2022) or for diagnostic SNPs associated with large inversions (Fang & Edwards, 2024). However, in general, SNP calling pipelines often exclude SV signals inadvertently (or intentionally) by masking repeats or removing SNPs with anomalously high depth or systematic missingness, which removes false SNP signals (e.g. Dallaire et al., 2023; Jaegle et al., 2023) but ultimately underrepresents SV variation. Further, SVs are often found in regions of high recombination (Currall et al., 2013; Stuart, Tan, et al., 2025) and may respond to selection differently than SNPs (discussed below), which will impact the patterns of linkage disequilibrium between SVs and adjacent SNPs (Kato et al., 2006). The interplay of both biological and technical reasons means that whether SVs are anchored to nearby SNPs is not clear (Chia et al., 2012; Geibel et al., 2022), and inferring genotypes based on nearby SNPs for known SVs may be inaccurate.

3. What are the properties of SVs that matter for population genomics?

Understanding the population genomics of SVs requires recognizing their properties and how they might interact with eco-evolutionary processes differently than SNPs (**Fig. 1**). Most knowledge about genome-wide SVs comes from model organisms or humans (often in disease research), focusing on variants with strong effect or easily identifiable SVs like short DEL or large INV. However, understanding how a study system's unique history interacts with SV properties can guide the development of evolutionarily relevant hypotheses.

3.1 The origins and dynamics of SVs are diverse and poorly characterized

The formation rates of SVs, including TEs, have long been studied at the macro-evolutionary scale within the context of speciation and genome size evolution (Talla et al., 2017). The vast diversity in genome sizes across the tree of life reflects the highly variable rates of SV origin and accumulation, and consequently, their contributions to genome content (Chalopin et al., 2015).

The rate at which SVs arise is expected to be more variable than for SNPs, reflecting the wide diversity in SV types and lengths (**Fig. 1a** , Ho & Schaack, 2021). SVs are thought to arise at a lower frequency than SNPs, on average, in terms of singular SNP mutation/ SV formation events (e.g. 0.16 vs 70 events per genome (Belyeu et al., 2021)), even though novel SV formation can impact a greater number of nucleotides. However, some types of SVs may arise more frequently than SNPs (e.g. very small microsatellites (Vigouroux et al., 2002) or DUP (Katju & Bergthorsson, 2013)). The subset of SVs that are TEs are known for their high activity levels (Biémont & Vieira, 2006), though this rate varies considerably across TE families and across time (Ho et al., 2021).

Many eco-evolutionary and population genomics models for SNPs assume constant mutation rates (e.g. Nadachowska-Brzyska et al., 2015). While such gradual models are known to be a simplification (Bergeron et al., 2023; Heasley et al., 2021), this may be particularly problematic when applied to SV formation rates due to their varied genomic contexts (Loewenthal et al., 2022; Petrov, 2002). Additionally, TEs specifically are well known to multiply in bursts often triggered by processes such as population expansion or hybridization events (Bergman & Bensasson, 2007). These events, which may appear as an excess of low-frequency TEs insertions, could be falsely interpreted as periods of strong purifying selection under an equilibrium model that assumes a balance between novel variant loss and gain (Bourgeois & Boissinot, 2019). Examining the origin and frequency of SNPs and different groups of SVs can offer nuanced insights into historical admixture, demography, and the eco-evolutionary contexts shaping variant fitness.

3.2 SVs are less likely to be neutral due to their length

Theoretical predictions state that the larger an SV, the higher the likelihood that it will induce a functional, likely negative, impact on fitness (Hämälä et al., 2021; Scott et al., 2021). Thus, we may expect that the distribution of fitness effects (DFE) of SVs to have a lower proportion of nearly neutral variants than SNPs (**Fig. 1b**). The maximum negative fitness consequence will be the same for SNPs and SVs (i.e. the variant is lethal), however, the rarity of large SVs suggests lethal deleterious SVs may be more common than deleterious SNPs (Eichler, 2019). Conversely, the magnitude of potential beneficial effect of SVs most likely extends beyond that of SNPs, with large inversion polymorphisms being a common example of beneficial variation (Berdan et al., 2023). The DFE reflects the interaction of formation rates, genome interactions, and selection regime. Even closely related species can exhibit differences in the fitness effects of SNPs, and SVs may be even more diverse in this regard (James et al., 2023).

Larger variants are more likely to have functional impacts, and thus effect size generally scales with variant length (Collins et al., 2020), with smaller SVs behaving more similarly to SNPs, though exceptions exist (Metzgar et al., 2000). The fitness effects of an SV also depend on the type of sequence change, with the addition, removal, or rearrangement of genomic sequence likely to be under varying strengths of selection in different contexts (Gaut et al., 2018; Loewenthal et al., 2022). The location of the variant is also important, for example intronic INS are less likely to be disruptive compared to DEL (Petrov, 2002). Finally, the impact of an SV may change over time. All SVs, and INV in particular, are prone to genetic load, owing to localised supressed recombination and reduced Ne (Hämälä et al., 2021), which may increase the mutational load they confer over generations (Jay et al., 2021).

This relationship between length and fitness consequence for SVs is also likely to interact with many biological processes in different ways compared to SNPs. Large effect SVs are predicted to be more resistant than SNPs to genetic swamping (when an allele is replaced by a more common variant due to gene flow from a larger population, Sakamoto et al., 2024), while effective population size affects whether smaller neutral or overdominant INV are favored (Connallon and Olito 2021). Because of differences in underlying mutation

rates and subsequent fitness effects, the evolutionary dynamics of SV may vary across population and species divergence continuums. For example, within plants, DUPs and TRA have been found to accumulate with increasing phylogenetic distance, suggesting they may be common SV classes that differentiate sister taxa, meanwhile INV differences were more stochastic and highly variable (Ferguson et al., 2024; Hirabayashi & Owens, 2023).

Consequently, the different characteristics of SVs such as length and type are important to consider, alongside selective and demographic processes, when quantifying common population genomic patterns such as site frequency spectrums and genetic diversity metrics (Collins et al., 2020). However, due to the unavoidable detection biases present in many SV studies to date (see Section 5.2), a comprehensive framework for population-level expectations of variant types, lengths, and frequencies has yet to be developed.

3.3 SVs can alter population level dynamics in ways that SNPs cannot

Large SVs create non-homologous sequences when occurring in the heterozygous form, which may interfere with, or entirely inhibit, recombination (**Fig. 1c**). Although INV are the most studied in this regard (Hoffmann & Rieseberg, 2008), other types of SVs, such as fusions, large CNVs or complex rearrangements, may have similar effects on recombination. This is caused by the reduction of chiasma or locally elevated LD in association with non-INV SVs (Dumas & Britton-Davidian, 2002; Trickett & Butlin, 1994; Wellband et al., 2019). SVs frequently arise in recombination hotspots (Currall et al., 2013), but can subsequently impede or inhibit recombination (Morgan et al., 2017). Regions with high recombination rates are more effective at purging deleterious variants (Kent et al., 2017; Morgan et al., 2017), however selection within gene-rich regions may favor the persistence of SVs that suppress recombination, to maintain advantageous haplotypes.

Recombination suppression has minimal effects on genome functionality but significantly influences population dynamics and the level of genetic variation (Yeaman, 2013). Whereas the rest of the genome is homogenized by recombination, the rearranged region exhibits two sets of haplotypes, with reduced or not gene flow between them and locally reduced effective population size (Ne) (see Faria et al., 2019 for further discussion). Large SVs inhibiting recombination may accumulate additional variants over time, often increasing deleterious effects (Berdan et al., 2022; Jay et al., 2021; Mahmoud et al., 2019). Conversely, they can link beneficial variants into 'supergenes,' particularly in INV (Wellenreuther & Bernatchez, 2018). Reduced recombination can also enhance the spread of beneficial variants, such as from range cores to edges in expanding populations (Peischl et al., 2015). This promotes selective advantages and the spread of SVs, including across species boundaries and in parallel adaptation across species ranges (Battlay, Craig, et al., 2024; Jay et al., 2018; Nicolas et al., 2025; Westram et al., 2022).

3.4 SVs can alter genome organization and functionality in ways that SNPs cannot

SVs can alter the 3D organization of DNA, for example by altering the boundaries of topologically associating domains (TAD), which helps to segregate interacting sequence features such as target genes and their cis-regulatory sequences (**Fig. 1d** , Spielmann et al., 2018). In this way, even small SVs that alter TAD boundaries may change chromatin structure, changing gene regulation without impacting the gene sequence itself. The role of SVs in chromatin accessibility can be examined using ATAC-seq (Ruggieri et al., 2022), though most work has been done to date in disease or developmental studies. In comparison, chromosome conformation capture can be used to investigate 3D interactions between separate regions of the genome, and has been used to demonstrate changes in 3D structure and TAD boundaries, which in turn impact the position of chromosome during meiosis which further favours other fusions (Vara et al., 2021).

When considering genome functionality, it is important to consider the unique properties of SVs that are also TEs. Similar to other SVs, TEs can create novel genetic sequences during their removal or insertion, and impact genome functionality (Klein & O'Neill, 2018). Unique to TEs, however, is that they contain the sequence motifs enabling transcription into mRNA within the host genome (Bourque et al., 2018). This means that the TE-derived sequence encoding replication may be exapted by the host genome through co-option or domestication. Ironically, such events appear to commonly aid the evolution of genome immunity and invasion defence, which in turn will help the genome to defend against TE replication (Jangam et al.,

2017). The expression of TE-derived sequences are integral to transcriptome function, including long non-coding RNAs and cis-regulatory networks (**Fig. 1h** , Bourque et al., 2018). RNA-seq data can reveal active TE families (Jin et al., 2015) and identify TE-derived sequence, highlighting co-option events (Oliveira et al., 2023).

Genomes have evolved defenses against TE invasions, creating an evolutionary arms race between TE activity and host suppression. Mechanisms like enzymatic control, epigenetic modifications (**Fig. 1d**), and small RNAs silence TEs (Bourgeois & Boissinot, 2019). However, stress can disrupt these controls, weakening regulation, triggering bursts of activity, and potentially causing genetic innovation (Capy et al., 2000; Stapley et al., 2015). Admixture of genetically distinct lineages can even cause 'genome shock,' activating TEs (Dion-Côté et al., 2014). Understanding interactions between genome processes, eco-evolutionary dynamics, and SVs will require careful sequence annotation to interpret patterns and outcomes effectively.

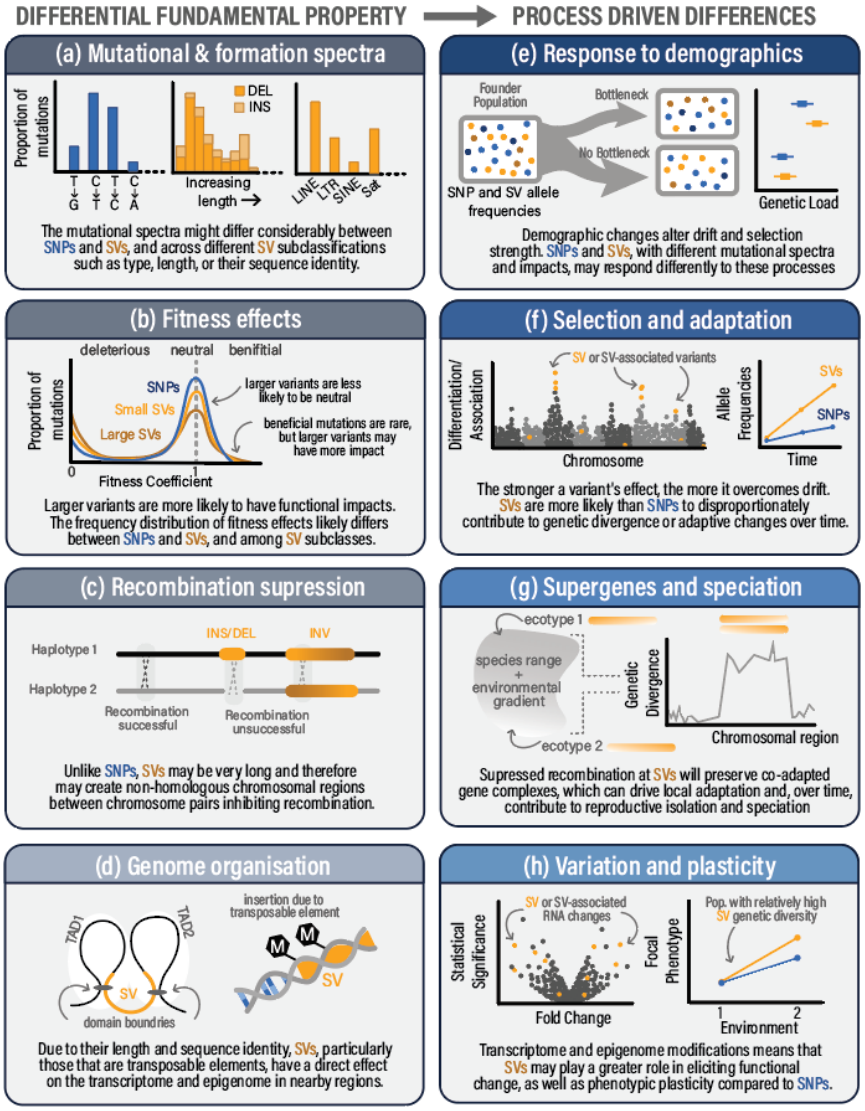

Figure 1 | This infographic illustrates how the fundamental properties of single nucleotide polymorphisms (SNPs) and structural variants (SVs) differ and how these differences interact with biologically important processes, influencing evolutionary and conservation outcomes.

Panel (**a**) depicts possible dimensions across which different SV formation spectra could be investigated. Panel (**b**) presents a theoretical frequency distribution of fitness effects for SVs compared to SNPs. Panel (**c**) depicts recombination suppression. Panel (**d**) depicts ways in which structural variants can induce large changes to genome organization and interactions through, for example, a SV disrupting boundaries between topologically associating domains (TADs) and a TE insertion causing methylome changes. Panel (**e**) depicts how SVs (yellow) may respond differently to SNPs (blue) to demographic changes. Panel (**f**) captures how SVs may overcontribute to adaptation, as seen through GWAS or GEA analysis, or when tracking adaptive allele/haplotype frequency changes over time. Panel (**g**) depicts how recombination suppression can interact with selection regime to create different ecotypes. Panel (**h**) depicts how SVs could disproportionately contribute to inter-individual variation and plasticity.

4. What do we know so far about the eco-evolutionary dynamics of structural variants?

Emerging research into population dynamics of SVs is increasing our understanding of their role in eco-evolutionary processes such as speciation, adaptation, and diversification (Berdan et al., 2023; Mérot et al., 2020; Serrato-Capuchina & Matute, 2018; Zhang et al., 2021), as well as in various specific contexts including in birds (Recuerda & Campagna, 2024), crop evolution (Gaut et al., 2018), insecticide resistance (Weetman et al., 2018), plant pathogens (Hartmann, 2022), biological invasions (Stuart, Santure, et al., 2025), and for different variant subclassifications such as TEs (Bourgeois & Boissinot, 2019), and inversions (Huang & Rieseberg, 2020; Wellenreuther & Bernatchez, 2018). Here we summarize the primary themes across this literature (see**Table 1** and appendix) and additional empirical research on SVs that is most relevant for eco-evolutionary research applied to population genomics datasets.

4.1 Demography likely plays a strong role in shaping correlations between SVs and SNPs

Genomic patterns of diversity and differentiation can be concordant or differ between sequence and structural variation. The level of correlation between different groups of genome-wide variants is important to know to best inform conservation research, such as cryopreservation, biobanking, or assisted migration choices (Bolton et al., 2022). For example, in Scandinavian wolves *Canis lupus lupus* , both SNPs and SVs contribute to mutational load and respond similarly to genetic rescue (Smeds et al., 2024), whereas within endangered hihi *Notiomystis cincta,* mutational load across SNPs and SVs and its impact on lifetime fitness are only moderately correlated with one another (Stuart, Tan, et al., 2025).

However, some research suggests that demographic processes may affect the coupling between sequence and structural variation (**Fig. 1e**), such as that found in the genetically bottlenecked invasive species European starling *Sturnus vulgaris* (Stuart et al., 2023) and fruit fly *Drosophila melanogaster* (Scarpa et al., 2024), within populations that experience intraspecific hybridization (e.g. butterfly *Lycaeides sp.,* Zhang et al., 2023)*, or among species resulting from a rapid radiation* (Recuerda et al., 2025). Non-recombining blocks, often due to large rearrangements, may also retain lineage-specific haplotype upon secondary contact, revealing a demographic history that is no longer visible in unlinked SNPs (Dallaire et al., 2025; Ishigohoka et al., 2024). Increased accumulation of TEs relative to synonymous SNPs was seen in small populations of lyrate rockcress *Arabidopsis lyrate* , demonstrating that weakened selection due to Ne reductions influences SNPs and TEs differently (Lockton et al., 2008). As with SNPs, it is crucial to investigate the conditions under which different demographic processes influence the (de)coupling of sequence and structural variation, as this remains largely unexplored.

4.2 SVs provide the substrate for rapid evolution

SVs are an important source of significant novel genetic variation and can be involved in the formation of new genes (Stewart & Rogers, 2019). SVs also contribute standing genetic variation, fuelling adaptation upon rapid environmental change (Barrett & Schluter, 2008). Changed selection pressures will alter the distribution of fitness effects (**Fig. 1b**), with strong-effect variants like SVs more likely to overcome drift and drive trait shifts within a population (**Fig. 1f**). This has been observed in invasive species, where SVs provide allelic variation on which selection acts within non-native ranges (e.g. Kreiner et al., 2019; Tepolt et al., 2022). Duplications (DUP), including of genes, whole chromosomes, and genomes (polyploidy), particularly facilitate

rapid evolution as they create genetic redundancy enabling the development of secondary functions while maintaining original functions (Magadum et al., 2013). For example, this process had been implicated in insecticide resistance in multiple mosquito species (Weetman et al., 2018).

Large INVs create jointly inherited regions that can contain multiple co-adapted genes. These "adaptive cassettes" can enable rapid responses to changes in environmental conditions (Jay et al., 2018; Ma et al., 2024; Mérot, Llaurens, et al., 2020; Therkildsen et al., 2019) and ultimately contribute to speciation (Faria et al., 2019). Because INVs only rearrange sequence content, they are less likely than DELs to be deleterious because of their initial sequence change (Berdan et al., 2022). INVs are often associated with adaptations to contrasting environmental conditions, sexual polymorphisms, and reproductive isolation (Wellenreuther & Bernatchez, 2018).

The self-replicating abilities of TEs provide them with an additional avenue for introducing rapid genomic change within a genome. Introgressed genomic regions containing novel TE families can facilitate TE invasions in a new host genome that lacks co-evolved defense mechanisms, creating a rapid and substantial source of novel genomic variation. This process is thought to have driven global fruit fly *D. melanogaster* invasions (Scarpa et al., 2024). Ultimately, to determine whether SVs (including TEs) contribute to rapid evolution beyond random chance, we need to better profile the rates at which they arise to distinguish the relative contributions of high rates of variation and strong selection to the evolutionary dynamics of TEs. Experimental evolution may provide complementary information about whether rapid adaptive change is biased toward standing or novel structural variation.

4.3 SVs are often implicated in parallel evolution

Parallel evolution, where a trait is selected for independently within separate lineages, may occur at the phenotypic or genetic level. While the former often involves complex polygenic architectures, genetic-level parallel evolution is more likely when a trait is controlled by a single gene or a few genes, and when the same mutations—or structurally similar variants—are repeatedly favoured by selection (Nosil et al., 2024). INVs can be very geographically and taxonomically widespread (Nicolas et al., 2025; Reeve et al., 2024), and are often implicated in studies of parallel adaptation (Westram et al., 2022). A 30-year island experiment demonstrated parallel adaptation in marine snails *Littorina saxatilis* facilitated by INVs (Garcia Castillo et al., 2024) and similar patterns have been observed in dune sunflowers *Helianthus sp.* (Todesco et al., 2020). How well these INV studies can inform the dynamics of other SV classes likely depends on their DFEs; as discussed, inversions may be less subject to immediate purifying selection since they do not alter the amount of genetic material.

Under similar artificial selection, large-effect SVs provide a frequent mechanism for parallel evolution (Therkildsen et al., 2019). Demography interacts with this process, and it has been demonstrated in simulations that large effective population sizes enhance the probability of parallel selection on strong-effect variants (MacPherson & Nuismer, 2017). This pattern has been well documented in invasive species, where the same standing SVs have been repeatedly selected in independent populations across the globe, as seen in the common ragweed *Ambrosia artemisiifolia* (Battlay, Craig, et al., 2024), white clover (Battlay, Hendrickson, et al., 2024) and common myna *Acridotheres tristis* (Atsawawaranunt et al., 2025). Open questions include the interplay between drift and selection in shaping SV-driven parallel evolution and how SV type, length, and content influence their role in it.

4.4 Multiple high-effect SV haplotypes may be maintained through balancing selection

Examples abound of SVs maintained through balancing selection, particularly for rearrangements longer than 100s of Kb (Berdan et al., 2022). One of the most common examples of balancing selection is spatially balanced polymorphism, where large-effect SVs are maintained across environmental gradients (**Fig. 1g**). In this scenario, migration load prevents the fixation of either SV haplotype, as each variant provides an adaptive advantage in different parts of the gradient (Tepolt et al., 2022; Wellenreuther & Bernatchez, 2018). Temporally cyclical variation in the selection regime also preserves SV polymorphism, such as the cyclical morphotype variations seen in both manipulated and field experiments with *Timema cristinae* stick insects

(Nosil et al., 2024). The main hypothesis explaining why long and large-effect SVs may be particularly involved in those cases of adaptation to spatially and temporally fluctuating selection is that they gain or capture a set of alleles co-adapted to similar environmental conditions, thereby transforming a polygenic architecture in a superlocus architecture (Kirkpatrick & Barton, 2006). Because they capture many genes and accumulate deleterious variants, non-recombining large SVs may also be under other mechanisms of balancing selection such as antagonistic pleiotropy (where a single genetic variant has both beneficial and detrimental effects depending on environmental conditions or life stages) and overdominance (where the heterozygote has a higher fitness than both homozygotes). Empirical examples include the seaweed fly *Coelopa frigida* , in which an overdominant inversion modulates a life-history trade-off (Mérot, Llaurens, et al., 2020). Investigating the extent to which SVs of other types and sizes are maintained by balancing selection will reveal whether the widespread examples of balanced inversions represent a biological pattern or a discovery bias.

Table 1 | Example studies that compare different groups of variants (e.g. SNPs, all SVs, INDELs, inversions, TEs etc.) across population genomics data in an eco-evolutionary context. Sequencing data types are listed in brackets next to the variant type they were used to identify RR – reduced representation data, SR – short-read, LR – long-read data, G – genome assembly, PG - pangenome graphs. See Table S1 for a more exhaustive bibliography, and the Supplementary Materials for literature search methods.

Focal species (taxa group)	**Variant types compared (read data)**
Two Lycaeides butterfly taxa (*L. melissa* and Jackson Hole *Lycaeides*)	SNPs DEL, INS, DUP, INV (LR)
Chocolate tree (*Theobroma cacao*)	SNPs DEL, INS, DUP, INV (G)
Atlantic salmon (*Salmo salar*)	SNPs Indels (SR) SVs - DEL, INS, DUP, INV (LR
American lobster (*Homarus americanus*)	SNPs CNV (RR)
Australasian snapper (*Chrysophrys auratus*)	SNPs DEL, INS, DUP, INV (LR + SR)
European starling (*Sturnus vulgaris*)	SNPs DEL, INS, DUP, INV (SR)
Mediterranean grass (*Brachypodium distachyon*)	SNPs TEs (SR)
Kokanee salmon (*Oncorhynchus nerka*)	SNPs SVs (SR)
Human (*Homo sapiens*)	DEL, INS, DUP, INV, Complex variants (SRs)
Cotton (*Gossypium sp.*)	INV, INS, DEL (Gs)
Eudicots plants (32 pairs of closely related species)	INV (Gs)
House finch (*Haemorhous mexicanus*)	SNPs, INDEL (<50bp) SVs (>=50bp) (PG)
Eucalyptus melliodora and E. sideroxylon	SVs (Gs and SR)

5. What data do I need to study SVs?

5.1 The growing accessibility of reference genomes and pangenome graphs

The quality of the reference genome assembly affects the accuracy of SV detection. A high-quality reference genome typically based on long reads, which are ideal to cover repetitive regions, is assembled into chromosomes and has few gaps. High quality reference genomes are increasingly available due in large part to genome sequencing consortia (e.g., African Biogenome Project, Darwin Tree of Life, Vertebrate Genomes Project, European Reference Genome Atlas, Canada Biogenome Project, DNA Zoo) and are promoting SV research (Formenti et al., 2022; Wold et al., 2021).

The use of a single reference may be limiting as it only captures variation within one individual. Many SVs, especially unbalanced ones, cannot be identified if the sequence they affect is not present in the reference genome. This is evident in the identification of INSs and DELs: while DELs can be easily spotted as a gap in the alignment, INSs can be missed if absent from the reference genome. Furthermore, SVs tend to occur in more dynamic genomic regions that are likely to harbour complex genetic variation, which is particularly difficult to detect if absent from the reference genome (Günther & Nettelblad, 2019; Martiniano et al., 2020; Stuart, Tan, et al., 2025). Pangenome graphs address this bias by incorporating multiple individuals' genomic

variation in a single data structure (Eizenga et al., 2020). Variable regions represent 'accessory' sequences, while homologous regions shared among all individuals in the genome graph represent the 'core' genome, providing a more comprehensive reference for mapping (Secomandi et al., 2025 for full review). Pangenome graphs are demonstrating their usefulness in humans and agriculturally significant species (Eizenga et al., 2020; Khan et al., 2020; Leonard et al., 2024; Li et al., 2024) and are increasing for non-model species (Edwards et al., 2025; Recuerda et al., 2025; Ruggieri et al., 2022; Secomandi et al., 2023). SV analysis tools are adjusting to mapping on pangenome graphs (Groza et al., 2023). Challenges include the resource intensiveness of assembling multiple contiguous de-novo genomes and working with SV and population data on pangenomes (Fang & Edwards, 2024).

5.2 Population-level analysis of SVs with short and long reads

SVs can be identified through discordant mapping of re-sequenced individuals to a reference genome (**Fig. 2** , **Box 1**). Identifying and genotyping population level SVs ideally uses high coverage (generally [?]15-20) short-read (SR) and/or long-read (LR) WGS data across a sufficient number of individuals (**Fig. 2**) (Mahmoud et al., 2019; Xi et al., 2010). Most software infer the location and type of SV based on read coverage and discordant read mapping (Xi et al., 2010), though machine learning, de Bruijn graphs, and local assemblies are increasingly available tools (Cameron et al., 2019; Chen et al., 2014; Cleal & Baird, 2022). Deletions are the easiest SV to detect due to their clear signals with SR and LR (**Fig. B1**). INSs, particularly TE insertions, are best detected with LR WGS (Zhao et al., 2021).

Many SV pipelines split analysis into an identification step, which builds a catalog of SVs, and a genotyping step (Alkan et al., 2011). Genotyping improves SV calls when applied to the same dataset by providing a final matrix of SVs with less missing data. Genotyping can also be applied on a different dataset. LR WGS or alignment of genome assemblies are used for discovery (allowing larger SVs to be detected, even across repetitive regions) and SR WGS, including low (<5x) and medium (5-15x) SR data, is used for genotyping as the reduced cost allows more individuals to be included (e.g. Jin et al., 2023; Lecomte et al., 2023). Recent tools will perform genotyping by building a variant-aware genome graph, which represents targeted SVs as bubbles on which short or long reads can be mapped for genotyping, see e.g. Paragraph (Chen et al., 2019), vg-tools (Hickey et al., 2020), SV-jedi-graph (Romain & Lemaitre, 2023), and benchmarking studies such as Duan et al., (2022). Promising methods alo suggest that SVs may be imputed to provide a more comprehensive genotyped dataset (Gundappa et al., 2025; Lou et al., 2021). In such two-step approaches, the choice of discovery individuals will affect the bias in variant discovery and an ideal panel would include different genetic lineages, ecotypes, and sexes.

Benchmarking studies and reviews are useful to choose software for SV discovery and genotyping and select the most appropriate best one(s) for the biological system, data, and objectives at hand (Ahsan et al., 2023; Helal et al., 2024; Lin et al., 2023). Different datasets and SV callers present different weaknesses and strengths (**Box 2**). Most SV detection programs have a length detection threshold. Specialized software also exists to identify *de -novo* TE insertions within WGS data (e.g. Chen et al., 2023; Li et al., 2018).

Beyond picking the right tool, further, filtering and curation steps may be needed depending on the degree of confidence required. For example, focusing on a few putatively adaptive SVs may require more stringency than a broad scale analysis (**Box 1**).

5.3 Alternative methods and future prospects

In addition to SR and LR sequencing, other methods have been used for SV discovery, including reduced-representation sequencing (Dorant et al., 2020; Huang et al., 2025; Tepolt et al., 2022), RNA-seq (Ma et al., 2018), optical mapping (Budurlean et al., 2024), linkage mapping (Akopyan et al., 2022), and reference genome alignments (Hirabayashi & Owens, 2023) (see Mahmoud et al., 2019 for review). The latter may be one of the most powerful approaches to detect SVs, depending on the quality of reference genomes compared. Pangenome graphs offer great advantages over synteny-based methods for the identification of SVs (Fang & Edwards, 2024). However, due to the costs of high-quality reference genomes, assembly-based SV analysis can not readily scale up to population level. Assembly comparisons are often used for studies at the macro-

evolutionary level involving few samples per species (Hirabayashi & Owens, 2023) and, for micro-evolutionary studies, to build a catalogue of SVs that will be genotyped in a larger number of individuals using short-reads (Merot et al., 2023) or reduced-representation sequences (Fang & Edwards, 2024).

For SV population genomics, cost-effective methods that conserve long-range contact information, such as linked-read sequencing (e.g., haplotagging; Meier et al., 2021), are promising. Short read libraries are prepared to carry a tag specific to a long molecule, allowing detection of anomalies in structure across individuals. Pipelines and tools to discover and genotype SVs with linked reads are emerging (Dimens, 2025; Romain & Lemaitre, 2025). Several studies demonstrate the value of linked reads for population-level analyses of SVs, especially inversions (Meier et al., 2021; Orteu et al., 2024). Comparable inferences can be made with Hi-C data (Chang et al., 2024), which also provides access to the 3D conformation of the genome (Schopflin et al., 2022) , although a big challenge remains with cost effective scaling to population level.

Expanding our understanding of how genetic diversity arises and is maintained may shift genomic research beyond individual genotypes. One way forward involves making greater use of information from broader analytical methods (such as haplotags) to define haplotype blocks that capture both SNPs and SVs. Haplotype blocks have been increasingly used in defining Ancestral Recombination Graphs (ARGs), providing the complete ancestry information of a sample of genomes (Shipilina et al., 2023). The full ARG, which provides information on the relationships between samples defined by a unique coalescence and recombination event, contains much richer information than individual SNP loci. These non-recombining haplotype blocks, which may contain any number of SNPs and SVs, will reside as the set of genomic regions that descend from a particular edge within the ARG. The advent of new computationally efficient methods for estimating haplotype structure, and for estimating the ARG for larger number of samples (Rasmussen et al., 2014; Speidel et al., 2019; Wohns et al., 2022), may also provide further avenues for jointly identifying SVs and understanding how their evolutionary dynamics may differ from SNPs at the population scale. Further developments will likely come from bridging methodologies, from linked read sequencing, utilisation of pan-genome graphs, and reducing SV detection biases, to enable access to rich sources of data from genealogical information contained in the complete ARG.

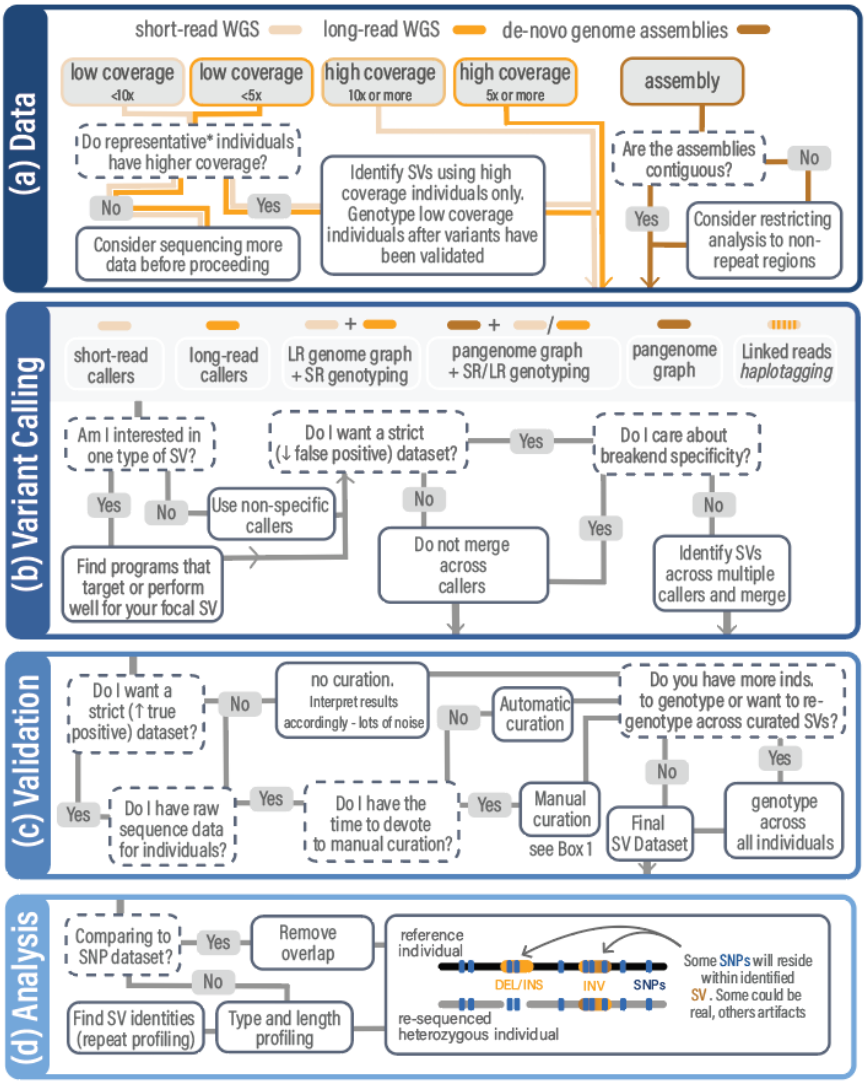

Figure 2 | Decision point workflow for genome-wide*de novo* identification of SVs using multiple genome assemblies, as well as whole genome resequencing (WGS) short-read (SR) and long-read (LR) data. Questions are depicted in dashed boxes, while decisions or analytical actions are depicted in solid lined boxes. This includes panel **(a)** how much genome coverage is typically considered low or high coverage for SR and LR WGS. *Representative individuals aim for different populations or familiar lineages to have equal and complete representation in the dataset. Panel **(b)** links the input data type to the type of variant calling approach needed for SV identification, as well as questions around cross-caller merging. Panel**(c)** depicts considerations around when to assess identified SVs using validation techniques. Panel **(d)** highlights a few key considerations and proximate analysis that should be performed on identified SVs. Input data type is indicated by different colours within panel **(a)** only, and is simplified to one colour for panels**(b-d)** .

BOX 1 | Reducing false positives in SV datasets and dealing with uncertainty

Several approaches can increase confidence in the final SV set, such as filtering steps that use minimum read coverages and quality metrics (**Fig. 2**). Merging SVs across multiple variant callers can also remove false positives (Jeffares et al., 2017). However, program performance varies depending on raw data and variant

types (e.g. they have often been designed for INS/DEL and perform poorly on TRA/INV) so this approach has limited sensitivity.

Even with high coverage, high quality reference and resequencing data, there will be noise creating uncertainty about the proportion of true variants. Conversely, some SVs that do not meet filtering criteria might reflect genetic differences between individuals that are too complex for SV identification methods to capture (Groza, Schwendinger-Schreck, et al., 2024). The balance between false positives and false negatives should be considered in light of the focal research question. The balance between false positives and false negatives should be considered in light of the focal research question. For question that aims at capturing a general pattern, it may be possible to accept inherent noise (high false positives, lower false negatives and reduced risk of viewer selection bias) and use shallow filtering or automated curation. Conversely, focusing on SVs of interest, on their breakpoint, and their effect on focal phenotypes may require additional analysis, such as manual curation or even qPCR (Recuerda & Campagna, 2024). Validation steps thus depend on the degree of confidence required for each study.

A common validation step is manual curation, which refines high-confidence SV lists by visually assessing local read mapping around SV breakpoints (**Fig. B1**) (Smeds et al., 2024). SVs are detected through various signals and programs (Alkan et al., 2011), but non-uniform read mapping can produce systematic errors. Manual curation improves accuracy and mitigates batch effects, aided by efficient tools (Belyeu et al., 2021).

We summarize key SV signatures in SR (**Fig. B1**). For details on short and long read SV identification, see Xi et al., (2010). For raw data signatures, refer to: DELs and DUPs (supplementary materials, David et al., 2024), INVs (Figure 3 in Wu et al., 2020), and INSs (**Fig. B1e,** Li et al., 2018). Manual curation plots may not always show clear variant signatures, but data exploration can help to identify true positive patterns. Coverage and read mapping in heterozygous and homozygous individuals can aid in system-specific calibration of interpretation. We also provide empirical SR datasets for exploring SV signals (https://github.com/katarinastuart/Rv3_SVecoevolutionary/) (Dryad link will also be made upon acceptance). While not traditional training sets, these real datasets help first-time curators recognize diverse read signals and assess false positives.

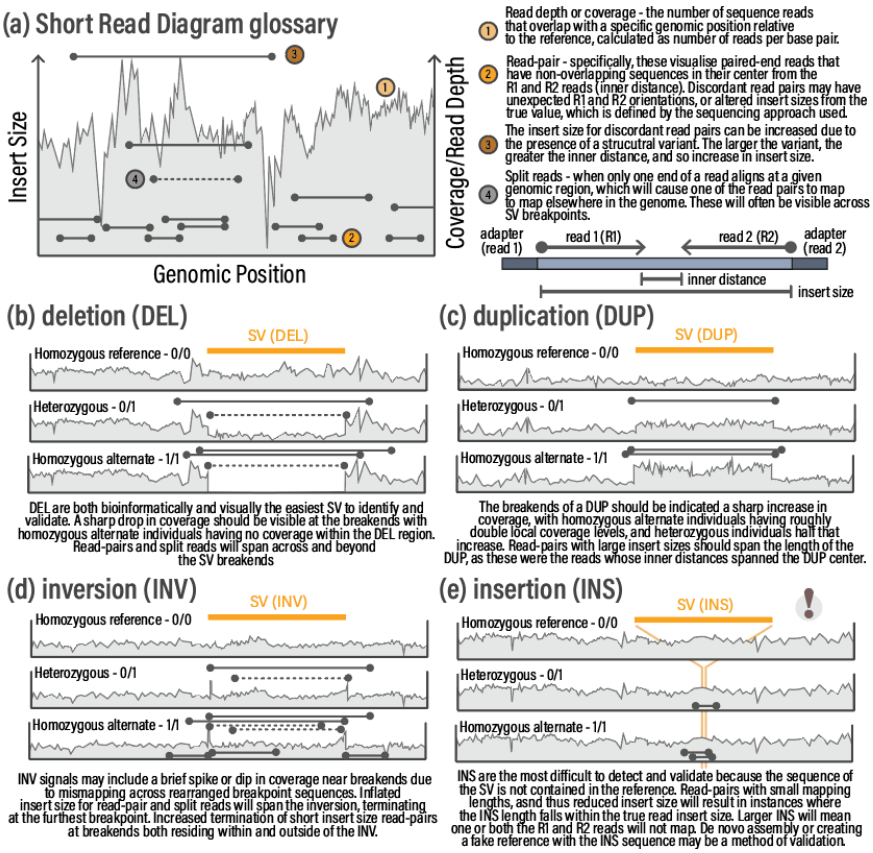

Figure B1 | Diagram of evidence for each SV type for validation using raw reads, where the alternate haplotype/allele is the focal SV. Panel **(a)** summarizes the diagrammatic features used for SV validation and depicts how inner distance is inferred from paired end reads. The following panels depict minimum ideal scenarios for **(b)** deletions, **(c)** duplications, **(d)** inversions, and **(e)** insertions.

6. What are the key challenges and considerations in SV analysis?

Researchers experienced in population genomics analyses but new to SVs can still encounter unexpected pitfalls (**Box 2**). In addition, here we outline key conceptual areas that are relevant for analysis of SVs and direct comparisons between SNPs and SVs.

6.1 Artifacts from batch effects and sequence read quality

There is a growing number of multi-year WGS data-sets generated across multiple sequencing batches. Variant calling and filtering protocols developed for SNPs effectively minimize the impact of batch effects (Lou & Therkildsen, 2022; Tom et al., 2017). Differing levels of DNA and sequence quality, coverage, read lengths, or insert sizes (particularly important for TE identification) are all factors introducing bias in SV identification (Chen et al., 2023). SVs are more prone to batch effects (variation between independently sequenced samples attributed to non-biological effects that creates systematic differences between groups) and are often encoded with only a few quality or mapping metrics that can be used to correct them or other artifacts (Stuart et al., 2023; Zarate et al., 2020). To alleviate batch effects, their severity can be determined through analysis of raw sequence variation (Lou & Therkildsen, 2022) and mitigation measures are taken before SV identification.

6.2 Uniform *vs.* adaptive methodologies for comparing genome-wide variants

A SV dataset can be filtered or subdivided by SV type (**Fig 3a** , examples in Hämälä et al., 2021), length (**Fig. 3b** , examples in Alonge et al., 2020), sequence identities (Fig. 3c, examples in Munasinghe et al., 2023), or gene proximity (Alonge et al., 2020; Smeds et al., 2024) to focus on a specific class of variants or to enable comparisons across variant types. Comparing SNPs and SVs or different SV types provides insights into how eco-evolutionary processes differ across variant types. However, counts of variant types can differ vastly in a dataset. SVs are generally one order of magnitude fewer than SNPs (Catanach et al., 2019) and DELs are far more common than large INVs or TRAs (Tigano & Russello, 2022; Wold et al., 2023). Combined with different pipelines and analytical methods for variant calling, these differences can hamper quantitative comparisons and introduce statistical biases.

While some common population genomics analyses (e.g. PCA) can yield similar results across diverse variant count numbers (Lecomte et al., 2023; Tigano & Russello, 2022), the sensitivity of others to sample size can produce misleading results (Lin et al., 2013). Selecting similar numbers of SNPs and SVs would address this bias but discard potentially informative data and fail to address the issue of large sample sizes not being fit for p-value testing (Lin et al., 2013). Using indicators that are not impacted by sample size, such as effect sizes (Sullivan & Feinn, 2012), or looking at comparative statistics or distributions of variants within continuous genome intervals can overcome this problem.

6.3 A lack of neutral framework for SVs

We lack theoretical and empirical frameworks for understanding the neutral and functional expectations for SVs. For example, discerning between synonymous and nonsynonymous SNPs is central to the estimation of mutation load or selection strength, but estimating the effect of a SV on gene function is far less intuitive. While direct genic overlap could be used to estimate deleterious effects, a SV in a non-coding region of the genome could alter the 3D structure of the genome and potentially have greater fitness consequences (Spielmann et al., 2018). Issues arising from a lack of neutral SV expectations include determining a neutral baseline for estimating the distribution of fitness effects (DFE). Intergenic SVs are a reasonable starting point (Fang & Edwards, 2024), yet can overestimate neutral SVs. Robust statistical frameworks that account for the unique properties of SVs are thus urgently needed. Simulations-based testing the limits of applying SNP-based methods on SVs will also be useful to account for both biological and technical differences. SV data impact such analysis In the meantime, SNP-based pipelines can be applied with caution and conservative approaches, and could ideally be tested with simulations accounting for SV effects (Lotterhos, 2019). This warning is particularly true for analysis that assume neutrality, such as demographic inference using the site frequency spectrum (Nadachowska-Brzyska et al., 2015), which can be biased by SVs (Dallaire et al., 2023).

6.4 Beyond biallelic structural variants

Most analysis in population and evolutionary genomics that incorporates SNP data are limited to biallelic variants, despite tri- and tetra-allelic SNPs being known to play significant roles in accounting for inter-individual genetic differences across many species (Hodgkinson & Eyre-Walker, 2010; Sopniewski & Catullo, 2024). Allelic variation can be much higher at SV loci, especially in repetitive SVs such as TEs, microsatellites, and CNV, or complex ones such as nested inversions (e.g. McComish et al., 2024; Munasinghe et al., 2023). As most analytical approaches target only bi-allelic SVs (Mahmoud et al., 2019), levels of structural variation are likely underestimated. Advancing methods that allow for variation in non-biallelic variants (e.g. (Saitou et al., 2022; Tigano et al., 2018) and at complex loci (e.g. pangenome graphs) may increase the statistical power of evolutionary inferences from SVs.

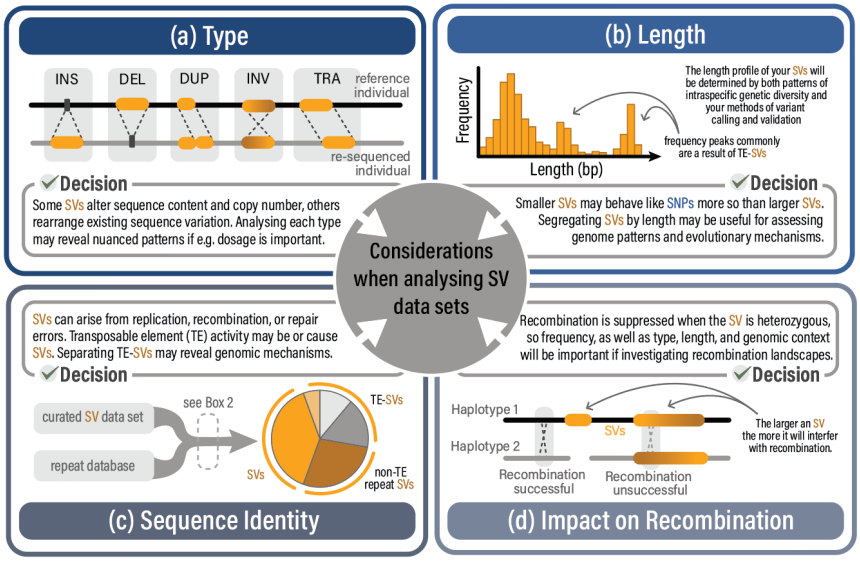

Figure 3 | Subclassifications of different structural variants (SVs) that may need to be considered when conducting analysis, as they could impact pattern inference of different eco evolutionary processes. Panel **(a)** depicts a diagram of different SV types, and suggests that conducting analysis separately on each type may be important if considering the volume and nature of genetic change is important for a study. Panel **(b)** depicts an example frequency distribution of SV lengths, which is extremely important to consider when trying to estimate the impact of a variant. Smaller SVs may have similar fitness impacts to SNPs, while larger SVs, though being less common, are more likely to have differential impact. Panel **(c)** depicts determining the sequence identity of SVs using a repeat database such as Repbase (Bao et al., 2015), or a de-novo species-specific repeat database using programs like EarlGrey (Baril et al., 2022). Understanding the sequence identity of a SV will also help to reveal information about how the SV was created. Panel **(d)** depicts how recombination may be suppressed in non-homologous regions created by heterozygous SVs within an individual. SVs that are lengths of hundreds of kbs to a few Mb may significantly affect recombination in heterozygotes, and different SV types may be more or less likely to inhibit recombination. While there has not yet been comprehensive characterisation of which lengths and types of SVs most impact recombination, researchers should be mindful that some SVs will impact their local recombination environment when in the heterozygous form, and others not.

Box 2 | Top 10 pitfalls when analyzing structural variants within a population genomics context and how to avoid them.

SV genotypes are stored in the same file format as SNPs (e.g., VCF, Plink BED), which can be input to many tools developed for sequence variation. Performing most analyses on SV data is straightforward, provided using it as input does not violate program assumptions. However, some differences between SNPs and SVs might change how you interact with the data and programs.

1. *SV calling may provide inaccurate or missing genotypes.* Because the read quality and quantity required for genotyping an SV are lower than for calling one, re-genotyping all samples after incorporating all detected variants across the population is recommended to reduce missing data and improve accuracy. This applies to all datasets including LR.
2. *SV caller and dataset have assets and limits* . Many programs have been developed and tested with a specific type of data, such as a minimum amount of coverage, and with a focus on specific SVs, usually DEL/INS. Moreover, most software are primarily tested on human datasets or idealized simulated datasets. Hence, it is important to check whether the data fits the requirements of the caller and analyse output carefully. For example, one typically finds less INV than INS, but it is unclear whether

this is due to lower power to detect them. The same applies to long SVs, which are likely rarer but also may be more difficult to detect.

3. *SVs are intervals and not positions:* Because of the diversity in SVs, there is extra information embedded in common file types (e.g. VCF) including variant type and length. Programs may not detect extra information embedded in such files and manual checking is often necessary. For example, some programs rely on the field length, while others use start stop; the length can be 0 for inversions or encoded as positive or negative values. In addition to a VCF, it may be useful to create a BED interval file that captures the chromosome and breakpoint positions of each SV. In doing so, it is important to ensure whether the location is encoded in 0-indexing (as used in BED files) or 1-indexing (as used in VCF files).
4. *The type of an SV is defined by the reference*: INS and DEL could be interchanged because an insertion in one sample may be a deletion in another. If detected against a different reference genome, an insertion in one population or species may appear as a deletion in another. Misclassification or lack of clarity about whether a variant is an INS or DEL can influence evolutionary and functional interpretations, particularly in studies of gene content, genome size variation, and fitness impacts. The ideal way of analysing INS/DEL from an evolutionary perspective is to infer the ancestral state, e.g., by polarizing the variants relative to a closely related species.
5. *The sequence information of INS is absent from the reference genome*: The sequence identity of most SVs can be found by identifying the overlapped BED interval file and the reference genome, except for INSs, whose sequences need to be found during variant calling. SV callers usually encode INS sequences into the output genetic file, but sometimes in the ALT field or elsewhere. INS are generally better provided by LR than SR, as it is challenging to assemble the inserted sequence from short read data (but see Rizk et al., 2014).
6. *Some programs may not accept the alternate allele encoding used for SVs*: SV VCFs are generally more complex than SNP VCFs. For example, REF and ALT can be encoded in different ways depending on the SV caller or include special characters incompatible with downstream programs. For downstream programs which do not use allele identities, this information can sometimes be overwritten with dummy REF and ALT alleles to force programs to accept the VCF without consequence to the analysis.
7. *SVs are better understood in the context of repeat annotation and repeat databases*: Because of the high proportion of repeats in SVs, understanding the abundance and distribution of repeat classes, including TEs, in the study species' genome is necessary. This can be done using pre-existing repeat databases (e.g. Bao et al., 2015), to annotate the genome, and the set of SVs with the program repeatmasker (Smit et al., 2013). More advanced approaches involve development of species-specific repeat databases, e.g., using RepeatModeler2 (Flynn et al., 2020). Knowing the repeat profile of SVs, and which ones are also TEs, will provide insights into formation mechanisms and putative impacts of different groups of SVs.
8. *A SV may not be a homogeneous unit*: It may be relevant, particularly for long SVs, to consider the breakpoints and the whole sequence separately because each can bear different information about the evolution of the variant. For example, analyzing sequence identity at breakpoints can reveal putative formational origin (Harringmeyer & Hoekstra, 2022), while large SVs can have multiple repeat sequences inside them which have accumulated afterwards (Jay et al., 2021). Hierarchical classification can be powerful to retrace the evolution of a variant (see Figure 4, Munasinghe et al., 2023).
9. *Recombination suppression areas are not necessarily inversions*: INV are often indirectly detected from the analysis of sequence variation, as suppressed local recombination generally leaves a signal of high LD and divergence (e.g. local PCAs, (Todesco et al., 2020). However, these patterns should always be taken with caution since it is only suggestive of a putative SV, possibly INV or a more complex rearrangement, and can alternatively represent haplotype blocks of different ancestry that have not been homogenised by recombination. Even for a true INV, the position of the variant remains imprecise because recombination suppression can expand up to a few 100s of kb beyond breakpoints (Li et al., 2023).
10. *Effective SV discovery and genotyping with pangenome graphs depends on underlying assembly quality*:

The ability of a pangenome graph to profile population-level SVs accurately relies on input assemblies being contiguous and error free. Difficult to resolve regions of the genome, such as those with high repeat content, are therefore difficult to profile. Pangenome graphs are essential for resolving complex structural variants, though bioinformatic approaches have limitations and are actively under development (Rice et al., 2023).

7. What are some exciting or necessary avenues of study for SVs?

Novel avenues of inquiry are opening for studies into SVs as they are analysed across a greater variety of biological systems. Qualitatively exploring the landscape of SVs along the genome and/or across populations is an important first step, as many taxa are lacking even a basic picture of how SVs are distributed in the genome, and how SVs capture population structure and diversity. This can involve creating standard profiles of SV lengths, types, and consider SV in the light of the repeat profiles. For population genomics, expanding analysis to SVs alongside SNPs will provide a more comprehensive capture of genetic diversity, which may, or may not, follow consistent patterns of population structure, distribution of fitness effects (DFE) (with comparable neutral ref), heterozygosity, and spatial-temporal distribution.

More generally, targeting SVs within an ecological and evolutionary framework, accounting for the different properties of SVs discussed above, and applied with the best practices open new research directions including:

- Investigating SVs in small populations to assess how a weakened selection regime influences their frequency and effects
- Examining how SVs interact with demographic changes, such as bottlenecks and population expansions.
- Exploring the relationship between SVs, recombination, and population structure across spatial and environmental gradients, including admixture and hybrid zones.
- Developing improved classification methods for SVs to enhance interpretability (e.g., does SV length matter? Can we infer formation processes?).
- Developing theoretical frameworks to predict how SV properties influence biological and evolutionary processes, including their DFE and multi-allelic variation. Theoretical models, simulations, and predictive tools must consider the bidirectional interactions between SVs and biological processes—complex dynamics that SNP-focused studies have largely overlooked (Fig. 1).
- Advancing statistical methods to account for SV-specific properties in commonly used analyses, ensuring that models appropriately capture their unique patterns and evolutionary consequences.
- Conducting multi-species studies to compare SV diversity (quantity) and SV properties (quality).
- Enhancing bioinformatics tools to improve SV call accuracy and best practices, including the potential development of a genotype likelihood framework for SVs.
- Investigate SV accumulation and fixation along a gradient of divergence, and at different phylogenetic scales, to consider significance of structural diversity on longer evolutionary timescales

8. Summary and conclusion

As more diverse study systems explore SVs at the population level, our understanding of how their prevalence varies across species lineages and evolutionary histories will grow. Throughout the history of the field of genetics and genomics, every new type of genetic variant (e.g. chromosomal banding, allozymes, microsatellites, mtDNA, SNPs) provided a new, and often expanded snapshot of the genetic diversity that is present within a species. We should treat SVs similarly, as even those obtained from pangenome graphs will likely underestimate intraspecific diversity in regions of the genome that are hard to assemble. Contrary to previous genetic markers that often replaced what was used before for genetic inferences, SVs are not going to replace SNPs but investigating their properties will advance our understanding of ecological and evolutionary genomics processes. Although technical restrictions and sources of uncertainty still require thoughtful interpretation of results, the study of SVs in ecology and evolution is opening exciting avenues of scientific inquiry in different study systems. When accompanied by methodological, statistical, and theoretical developments, there will undoubtedly be many exciting discoveries forthcoming which will have important implications for evolutionary theory, conservation analysis, and species' adaptation to rapidly changing environments.

Acknowledgments:

People

Thank you to Iris Jia, Avneet Kaur, Florent Sylvestre, Philippe Hénault, Xavier Dallaire, Laurie Lecomte, Timothy Healy, Nadège Bélouard, and Branden Judson for feedback on an earlier draft of the manuscript.

Funding

KCS is supported by a Genomics Aotearoa ECR Grant, and a Marsden Grant (UOA1911) from the New Zealand Royal Society Te Aparangi. RAO is supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant, the New Brunswick Innovation Foundation, and a Swedish Research Council Starting Grant. CM is supported by an ERG-Stg EVOL-SV. MW is supported by a Marsden Grant (MFP-PAF2301) from the New Zealand Royal Society Te Aparangi and funding from the New Zealand Ministry of Business, Innovation and Employment (CAWX2304, PFR2430). AT is supported by Fisheries and Oceans Canada. CM is supported by an ERC-Stg (EVOL-SV).

REFERENCES

Ahsan, M. U., Liu, Q., Perdomo, J., Fang, L., & Wang, K. (2023). A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nature Methods*, *20*(8), 1143–1158. https://doi.org/10.1038/s41592-023-01932-w

Akopyan, M., Tigano, A., Jacobs, A., Wilder, A. P., Baumann, H., & Therkildsen, N. O. (2022). Comparative linkage mapping uncovers recombination suppression across massive chromosomal inversions associated with local adaptation in Atlantic silversides. *Molecular Ecology*, *31*(12), 3323–3341. https://doi.org/10.1111/mec.16472

Akopyan, M., Tigano, A., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2025). Genetic differentiation is constrained to chromosomal inversions and putative centromeres in locally adapted populations with higher gene flow. *Molecular Biology and Evolution*, msaf092. https://doi.org/10.1093/molbev/msaf092

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*(5), 363–376. https://doi.org/10.1038/nrg2958

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T. H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A. L., Tieman, D. M., Klee, H., Kirsche, M., … Lippman, Z. B. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell*, *182*(1), 145-161.e23. https://doi.org/10.1016/j.cell.2020.05.021

Atsawawaranunt, K., Stuart, K. C., Whibley, A., Ewart, K. M., Major, R. E., Johnson, R. N., & Santure, A. W. (2025). Parallel Signatures of Diet Adaptation in the Invasive Common Myna Genome. *Molecular Ecology*, *34*(2), e17607. https://doi.org/10.1111/mec.17607

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 11. https://doi.org/10.1186/s13100-015-0041-9

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Barton, H. J., & Zeng, K. (2018). New Methods for Inferring the Distribution of Fitness Effects for INDELs and SNPs. *Molecular Biology and Evolution*, *35*(6), 1536–1546. https://doi.org/10.1093/molbev/msy054

Battlay, P., Craig, S., Putra, A. R., Monro, K., Silva, N. P. D., Wilson, J., Bieker, V. C., Kabir, S., Shamaya, N., Boheemen, L. van, Rieseberg, L. H., Stinchcombe, J. R., Fournier-Level, A., Martin, M. D., & Hodgins, K. A. (2024). *Rapid parallel adaptation in distinct invasions of Ambrosia artemisiifolia is driven by large-effect structural variants* (p. 2024.09.18.613801). bioRxiv. https://doi.org/10.1101/2024.09.18.613801

Battlay, P., Hendrickson, B. T., Mendez-Reneau, J. I., Santangelo, J. S., Albano, L. J., Wilson, J., Caizergues, A. E., King, N., Puentes, A., Tudoran, A., Violle, C., Vasseur, F., Patterson, C. M., Foster, M., Stamps, C., Innes, S. G., Allio, R., Angeoletto, F., Anstett, D. N., ... Kooyers, N. J. (2024). *Structural variants underlie parallel adaptation following global invasion* (p. 2024.07.09.602765). bioRxiv. https://doi.org/10.1101/2024.07.09.602765

Belyeu, J. R., Brand, H., Wang, H., Zhao, X., Pedersen, B. S., Feusier, J., Gupta, M., Nicholas, T. J., Brown, J., Baird, L., Devlin, B., Sanders, S. J., Jorde, L. B., Talkowski, M. E., & Quinlan, A. R. (2021). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *American Journal of Human Genetics*, *108*(4), 597–607. https://doi.org/10.1016/j.ajhg.2021.02.012

Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., Gilbert, K. J., Jay, P., Kapun, M., Lotterhos, K. E., Mérot, C., Durmaz Mitchell, E., Pascual, M., Peichel, C. L., Rafajlović, M., Westram, A. M., Schaeffer, S. W., Johannesson, K., & Flatt, T. (2023). How chromosomal inversions reorient the evolutionary process. *Journal of Evolutionary Biology*, *36*(12), 1761–1782. https://doi.org/10.1111/jeb.14242

Berdan, E. L., Blanckaert, A., Butlin, R. K., Flatt, T., Slotte, T., & Wielstra, B. (2022). Mutation accumulation opposes polymorphism: Supergenes and the curious case of balanced lethals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1856), 20210199. https://doi.org/10.1098/rstb.2021.0199

Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., Hoffman, J. I., Li, Z., St. Leger, J., Shao, C., Stiller, J., Gilbert, M. T. P., Schierup, M. H., & Zhang, G. (2023). Evolution of the germline mutation rate across vertebrates. *Nature*, *615*(7951), 285–291. https://doi.org/10.1038/s41586-023-05752-y

Bergman, C. M., & Bensasson, D. (2007). Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *104*(27), 11340–11345. https://doi.org/10.1073/pnas.0702552104

Biémont, C., & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature*, *443*(7111), Article 7111. https://doi.org/10.1038/443521a

Blaj, I., Tetens, J., Bennewitz, J., Thaller, G., & Falker-Gieske, C. (2022). Structural variants and tandem repeats in the founder individuals of four F2 pig crosses and implications to F2 GWAS results. *BMC Genomics*, *23*(1), 631. https://doi.org/10.1186/s12864-022-08716-0

Bolton, R. L., Mooney, A., Pettit, M. T., Bolton, A. E., Morgan, L., Drake, G. J., Appeltant, R., Walker, S. L., Gillis, J. D., & Hvilsom, C. (2022). Resurrecting biodiversity: Advanced assisted reproductive technologies and biobanking. *Reproduction & Fertility*, *3*(3), R121–R146. https://doi.org/10.1530/RAF-22-0005

Bourgeois, Y., & Boissinot, S. (2019). On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes*, *10*(6), Article 6. https://doi.org/10.3390/genes10060419

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Budurlean, L., Tukaramrao, D. B., Zhang, L., Dovat, S., & Broach, J. (2024). Integrating Optical Genome Mapping and Whole Genome Sequencing in Somatic Structural Variant Detection. *Journal of Personalized Medicine*, *14*(3), 291. https://doi.org/10.3390/jpm14030291

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. https://doi.org/10.1038/s41467-019-11146-4

Capy, P., Gasperi, G., Biémont, C., & Bazin, C. (2000). Stress and transposable elements: Co-evolution or useful parasites? *Heredity*, *85*(2), 101–106. https://doi.org/10.1046/j.1365-2540.2000.00751.x

Carvalho, C. M. B., & Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews. Genetics*, *17*(4), 224–238. https://doi.org/10.1038/nrg.2015.25

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. https://doi.org/10.1111/mec.15051

Chalopin, D., Naville, M., Plard, F., Galiana, D., & Volff, J.-N. (2015). Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. *Genome Biology and Evolution*, *7*(2), 567–580. https://doi.org/10.1093/gbe/evv005

Chang, L., Xie, Y., Taylor, B., Wang, Z., Sun, J., Armand, E. J., Mishra, S., Xu, J., Tastemel, M., Lie, A., Gibbs, Z. A., Indralingam, H. S., Tan, T. M., Bejar, R., Chen, C. C., Furnari, F. B., Hu, M., & Ren, B. (2024). Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nature Biotechnology*, 1–14. https://doi.org/10.1038/s41587-024-02447-1

Chen, J., Basting, P. J., Han, S., Garfinkel, D. J., & Bergman, C. M. (2023). Reproducible evaluation of transposable element detectors with McClintock 2 guides accurate inference of Ty insertion patterns in yeast. *Mobile DNA*, *14*(1), 8. https://doi.org/10.1186/s13100-023-00296-4

Chen, K., Chen, L., Fan, X., Wallis, J., Ding, L., & Weinstock, G. (2014). TIGRA: A targeted iterative graph routing assembler for breakpoint assembly. *Genome Research*, *24*(2), 310–317. https://doi.org/10.1101/gr.162883.113

Chen, S., Krusche, P., Dolzhenko, E., Sherman, R. M., Petrovski, R., Schlesinger, F., Kirsche, M., Bentley, D. R., Schatz, M. C., Sedlazeck, F. J., & Eberle, M. A. (2019). Paragraph: A graph-based structural variant genotyper for short-read sequence data. *Genome Biology*, *20*(1), 291. https://doi.org/10.1186/s13059-019-1909-7

Chia, J.-M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., Elshire, R. J., Gaut, B., Geller, L., Glaubitz, J. C., Gore, M., Guill, K. E., Holland, J., Hufford, M. B., Lai, J., Li, M., Liu, X., Lu, Y., McCombie, R., … Ware, D. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics*, *44*(7), 803–807. https://doi.org/10.1038/ng.2313

Cleal, K., & Baird, D. M. (2022). Dysgu: Efficient structural variant calling using short or long reads. *Nucleic Acids Research*, *50*(9), e53. https://doi.org/10.1093/nar/gkac039

Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. V., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., … Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, *581*(7809), Article 7809. https://doi.org/10.1038/s41586-020-2287-8

Currall, B. B., Chiang, C., Talkowski, M. E., & Morton, C. C. (2013). Mechanisms for Structural Variation in the Human Genome. *Current Genetic Medicine Reports*, *1*(2), 81–90. https://doi.org/10.1007/s40142-013-0012-8

Dallaire, X., Bouchard, R., Hénault, P., Ulmo-Diaz, G., Normandeau, E., Mérot, C., Bernatchez, L., & Moore, J.-S. (2023). Widespread Deviant Patterns of Heterozygosity in Whole-Genome Sequencing Due to Autopolyploidy, Repeated Elements, and Duplication. *Genome Biology and Evolution*, *15*(12), evad229. https://doi.org/10.1093/gbe/evad229

Dallaire, X., Normandeau, E., Brazier, T., Harris, L., Hansen, M. M., Mérot, C., & Moore, J.-S. (2025). Leveraging Whole Genomes, Mitochondrial DNA and Haploblocks to Decipher Complex Demographic Histories: An Example From a Broadly Admixed Arctic Fish. *Molecular Ecology*, *34*(10), e17772. https://doi.org/10.1111/mec.17772

David, G., Bertolotti, A., Layer, R., Scofield, D., Hayward, A., Baril, T., Burnett, H. A., Gudmunds, E., Jensen, H., & Husby, A. (2024). Calling Structural Variants with Confidence from Short-Read Data in Wild Bird Populations. *Genome Biology and Evolution*, *16*(4), evae049. https://doi.org/10.1093/gbe/evae049

Dimens, P. V. (2025). *Pdimens/harpy* [Python]. https://github.com/pdimens/harpy (Original work published 2022)

Dion-Côté, A.-M., Renaut, S., Normandeau, E., & Bernatchez, L. (2014). RNA-seq Reveals Transcriptomic Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species. *Molecular Biology and Evolution*, *31*(5), 1188–1199. https://doi.org/10.1093/molbev/msu069

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandeau, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Molecular Ecology*, *29*(24), 4765–4782. https://doi.org/10.1111/mec.15565

Duan, X., Pan, M., & Fan, S. (2022). Comprehensive evaluation of structural variant genotyping methods based on long-read sequencing data. *BMC Genomics*, *23*(1), 324. https://doi.org/10.1186/s12864-022-08548-y

Dumas, D., & Britton-Davidian, J. (2002). Chromosomal rearrangements and evolution of recombination: Comparison of chiasma distribution patterns in standard and robertsonian populations of the house mouse. *Genetics*, *162*(3), 1355–1366. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1462339/

Ebenezer, T. E., Low, R. S., O'Neill, E. C., Huang, I., DeSimone, A., Farrow, S. C., Field, R. A., Ginger, M. L., Guerrero, S. A., Hammond, M., Hampl, V., Horst, G., Ishikawa, T., Karnkowska, A., Linton, E. W., Myler, P., Nakazawa, M., Cardol, P., Sánchez-Thomas, R., … Field, M. C. (2022). Euglena International Network (EIN): Driving euglenoid biotechnology for the benefit of a challenged world. *Biology Open*, *11*(11), bio059561. https://doi.org/10.1242/bio.059561

Edwards, S. V., Fang, B., Khost, D., Kolyfetis, G. E., Cheek, R. G., DeRaad, D. A., Chen, N., Fitzpatrick, J. W., McCormack, J. E., Funk, W. C., Ghalambor, C. K., Garrison, E., Guarracino, A., Li, H., & Sackton, T. B. (2025). *Comparative population pangenomes reveal unexpected complexity and fitness effects of structural variants* (p. 2025.02.11.637762). bioRxiv. https://doi.org/10.1101/2025.02.11.637762

Eichler, E. E. (2019). Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *New England Journal of Medicine*, *381*(1), 64–74. https://doi.org/10.1056/NEJMra1809315

Eizenga, J. M., Novak, A. M., Sibbesen, J. A., Heumos, S., Ghaffaari, A., Hickey, G., Chang, X., Seaman, J. D., Rounthwaite, R., Ebler, J., Rautiainen, M., Garg, S., Paten, B., Marschall, T., Sirén, J., & Garrison, E. (2020). Pangenome Graphs. *Annual Review of Genomics and Human Genetics*, *21*(Volume 21, 2020), 139–162. https://doi.org/10.1146/annurev-genom-120219-080406

Fang, B., & Edwards, S. V. (2024). Fitness consequences of structural variation inferred from a House Finch pangenome. *Proceedings of the National Academy of Sciences*, *121*(47), e2409943121. https://doi.org/10.1073/pnas.2409943121

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Rafajlović, M., Panova, M., Ravinet, M., Johannesson, K., Westram, A. M., & Butlin, R. K. (2019). Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. *Molecular Ecology*, *28*(6), 1375–1393. https://doi.org/10.1111/mec.14972

Ferguson, S., Jones, A., Murray, K., Andrew, R. L., Schwessinger, B., Bothwell, H., & Borevitz, J. (2024). Exploring the role of polymorphic interspecies structural variants in reproductive isolation and adaptive divergence in Eucalyptus. *GigaScience*, *13*, giae029. https://doi.org/10.1093/gigascience/giae029

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., ... Zammit, G. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, *37*(3), 197–202. https://doi.org/10.1016/j.tree.2021.11.008

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, *26*(20), 5369–5406. https://doi.org/10.1111/mec.14264

Garcia Castillo, D., Barton, N., Faria, R., Larsson, J., Stankowski, S., Butlin, R., Johannesson, K., & Westram, A. M. (2024). Predicting rapid adaptation in time from adaptation in space: A 30-year field experiment in marine snails. *Science Advances*, *10*(41), eadp2102. https://doi.org/10.1126/sciadv.adp2102

Gaut, B. S., Seymour, D. K., Liu, Q., & Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. *Nature Plants*, *4*(8), Article 8. https://doi.org/10.1038/s41477-018-0210-1

Geibel, J., Praefke, N. P., Weigend, S., Simianer, H., & Reimer, C. (2022). Assessment of linkage disequilibrium patterns between structural variants and single nucleotide polymorphisms in three commercial chicken populations. *BMC Genomics*, *23*, 193. https://doi.org/10.1186/s12864-022-08418-7

Groza, C., Chen, X., Wheeler, T. J., Bourque, G., & Goubert, C. (2023). *GraffiTE: A Unified Framework to Analyze Transposable Element Insertion Polymorphisms using Genome-graphs* (p. 2023.09.11.557209). bioRxiv. https://doi.org/10.1101/2023.09.11.557209

Groza, C., Chen, X., Wheeler, T. J., Bourque, G., & Goubert, C. (2024). A unified framework to analyze transposable element insertion polymorphisms using graph genomes. *Nature Communications*, *15*(1), 8915. https://doi.org/10.1038/s41467-024-53294-2

Groza, C., Schwendinger-Schreck, C., Cheung, W. A., Farrow, E. G., Thiffault, I., Lake, J., Rizzo, W. B., Evrony, G., Curran, T., Bourque, G., & Pastinen, T. (2024). Pangenome graphs improve the analysis of structural variants in rare genetic diseases. *Nature Communications*, *15*(1), 657. https://doi.org/10.1038/s41467-024-44980-2

Gundappa, M. K., Robledo, D., Hamilton, A., Houston, R. D., Prendergast, J. G. D., & Macqueen, D. J. (2025). High performance imputation of structural and single nucleotide variants using low-coverage whole genome sequencing. *Genetics Selection Evolution*, *57*(1), 16. https://doi.org/10.1186/s12711-025-00962-6

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*, *15*(7), e1008302. https://doi.org/10.1371/journal.pgen.1008302

Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., dePamphilis, C. W., & Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. *Proceedings of the National Academy of Sciences*, *118*(35). https://doi.org/10.1073/pnas.2102914118

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*, *6*(12), 1965–1979. https://doi.org/10.1038/s41559-022-01890-0

Hartmann, F. E. (2022). Using structural variants to understand the ecological and evolutionary dynamics of fungal plant pathogens. *New Phytologist*, *234*(1), 43–49. https://doi.org/10.1111/nph.17907

Heasley, L. R., Sampaio, N. M. V., & Argueso, J. L. (2021). Systemic and rapid restructuring of the genome: A new perspective on punctuated equilibrium. *Current Genetics*, *67*(1), 57–63. https://doi.org/10.1007/s00294-020-01119-2

Helal, A. A., Saad, B. T., Saad, M. T., Mosaad, G. S., & Aboshanab, K. M. (2024). Benchmarking long-read aligners and SV callers for structural variation detection in Oxford nanopore sequencing data. *Scientific Reports*, *14*(1), 6160. https://doi.org/10.1038/s41598-024-56604-2

Hickey, G., Heller, D., Monlong, J., Sibbesen, J. A., Sirén, J., Eizenga, J., Dawson, E. T., Garrison, E., Novak, A. M., & Paten, B. (2020). Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biology*, *21*(1), 35. https://doi.org/10.1186/s13059-020-1941-7

Hirabayashi, K., & Owens, G. L. (2023). The rate of chromosomal inversion fixation in plant genomes is highly variable. *Evolution*, *77*(4), 1117–1130. https://doi.org/10.1093/evolut/qpad027

Ho, E. K. H., Bellis, E. S., Calkins, J., Adrion, J. R., Iv, L. C. L., & Schaack, S. (2021). Engines of change: Transposable element mutation rates are high and variable within Daphnia magna. *PLoS Genetics*, *17*(11), e1009827. https://doi.org/10.1371/journal.pgen.1009827

Ho, E. K. H., & Schaack, S. (2021). Intraspecific Variation in the Rates of Mutations Causing Structural Variation in Daphnia magna. *Genome Biology and Evolution*, *13*(12), evab241. https://doi.org/10.1093/gbe/evab241

Hodgkinson, A., & Eyre-Walker, A. (2010). Human Triallelic Sites: Evidence for a New Mutational Mechanism? *Genetics*, *184*(1), 233. https://doi.org/10.1534/genetics.109.110510

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics*, *39*, 21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Horvath, R., Minadakis, N., Bourgeois, Y., & Roulin, A. C. (2024). The evolution of transposable elements in Brachypodium distachyon is governed by purifying selection, while neutral and adaptive processes play a minor role. *eLife*, *12*. https://doi.org/10.7554/eLife.93284.2

Huang, K., Ostevik, K. L., Jahani, M., Todesco, M., Bercovich, N., Andrew, R. L., Owens, G. L., & Rieseberg, L. H. (2025). Inversions contribute disproportionately to parallel genomic divergence in dune sunflowers. *Nature Ecology & Evolution*, *9*(2), 325–335. https://doi.org/10.1038/s41559-024-02593-4

Huang, K., & Rieseberg, L. H. (2020). Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00296

Ishigohoka, J., Bascón-Cardozo, K., Bours, A., Fuß, J., Rhie, A., Mountcastle, J., Haase, B., Chow, W., Collins, J., Howe, K., Uliano-Silva, M., Fedrigo, O., Jarvis, E. D., Pérez-Tris, J., Illera, J. C., & Liedvogel, M. (2024). Distinct patterns of genetic variation at low-recombining genomic regions represent haplotype structure. *Evolution*, *78*(12), 1916–1935. https://doi.org/10.1093/evolut/qpae117

Jaegle, B., Pisupati, R., Soto-Jiménez, L. M., Burns, R., Rabanal, F. A., & Nordborg, M. (2023). Extensive sequence duplication in Arabidopsis revealed by pseudo-heterozygosity. *Genome Biology*, *24*(1), 44. https://doi.org/10.1186/s13059-023-02875-3

James, J., Kastally, C., Budde, K. B., González-Martínez, S. C., Milesi, P., Pyhäjärvi, T., Lascoux, M., & GenTree Consortium. (2023). Between but Not Within-Species Variation in the Distribution of Fitness Effects. *Molecular Biology and Evolution*, *40*(11), msad228. https://doi.org/10.1093/molbev/msad228

Jangam, D., Feschotte, C., & Betrán, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. *Trends in Genetics : TIG*, *33*(11), 817. https://doi.org/10.1016/j.tig.2017.07.011

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, *53*(3), 288–293. https://doi.org/10.1038/s41588-020-00771-1

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion.

Current Biology, *28*(11), 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., Lux, T., Kamal, N., Lang, D., Himmelbach, A., Ens, J., Zhang, X.-Q., Angessa, T. T., Zhou, G., Tan, C., Hill, C., Wang, P., Schreiber, M., Boston, L. B., … Stein, N. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, *588*(7837), 284–289. https://doi.org/10.1038/s41586-020-2947-8

Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., & Sedlazeck, F. J. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, *8*(1), 14061. https://doi.org/10.1038/ncomms14061

Jin, S., Han, Z., Hu, Y., Si, Z., Dai, F., He, L., Cheng, Y., Li, Y., Zhao, T., Fang, L., & Zhang, T. (2023). Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. *Molecular Plant*, *16*(4), 678–693. https://doi.org/10.1016/j.molp.2023.02.004

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics (Oxford, England)*, *31*(22), 3593–3599. https://doi.org/10.1093/bioinformatics/btv422

Katju, V., & Bergthorsson, U. (2013). Copy-number changes in evolution: Rates, fitness effects and adaptive significance. *Frontiers in Genetics*, *4*, 273. https://doi.org/10.3389/fgene.2013.00273

Kato, M., Sekine, A., Ohnishi, Y., Johnson, T. A., Tanaka, T., Nakamura, Y., & Tsunoda, T. (2006). Linkage disequilibrium of evolutionarily conserved regions in the human genome. *BMC Genomics*, *7*, 326. https://doi.org/10.1186/1471-2164-7-326

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160458. https://doi.org/10.1098/rstb.2016.0458

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., & Varshney, R. K. (2020). Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement. *Trends in Plant Science*, *25*(2), 148–158. https://doi.org/10.1016/j.tplants.2019.10.012

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*(1), 419–434. https://doi.org/10.1534/genetics.105.047985

Klein, S. J., & O'Neill, R. J. (2018). Transposable elements: Genome innovation, chromosome diversity, and centromere conflict. *Chromosome Research*, *26*(1), 5–23. https://doi.org/10.1007/s10577-017-9569-5

Kreiner, J. M., Giacomini, D. A., Bemm, F., Waithaka, B., Regalado, J., Lanz, C., Hildebrandt, J., Sikkema, P. H., Tranel, P. J., Weigel, D., Stinchcombe, J. R., & Wright, S. I. (2019). Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. *Proceedings of the National Academy of Sciences*, *116*(42), 21076–21084. https://doi.org/10.1073/pnas.1900870116

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., Chen, W., Liang, X., Huang, L., Wang, J., Liang, E., Wu, Q., Lee, S. M.-Y., Xu, X., Höglund, J., … Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics*, *48*(1), 84–88. https://doi.org/10.1038/ng.3430

Lecomte, L., Árnyasi, M., Ferchaud, A.-L., Kent, M., Lien, S., Stenløkk, K., Sylvestre, F., Mérot, C., & Bernatchez, L. (2023). *Investigating structural variant, indel and single nucleotide polymorphism differentiation between locally adapted Atlantic salmon populations* (p. 2023.09.12.557169). bioRxiv. https://doi.org/10.1101/2023.09.12.557169

Leonard, A. S., Mapel, X. M., & Pausch, H. (2024). Pangenome-genotyped structural variation improves molecular phenotype mapping in cattle. *Genome Research*, *34*(2), 300–309. https://doi.org/10.1101/gr.278267.123

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S.-M., Chen, L., Childers, A. K., Coddington, J. A., … Zhang, G. (2022). The Earth BioGenome Project 2020: Starting the clock. *Proceedings of the National Academy of Sciences*, *119*(4), e2115635118. https://doi.org/10.1073/pnas.2115635118

Li, H., Berent, E., Hadjipanteli, S., Galey, M., Muhammad-Lahbabi, N., Miller, D. E., & Crown, K. N. (2023). Heterozygous inversion breakpoints suppress meiotic crossovers by altering recombination repair outcomes. *PLOS Genetics*, *19*(4), e1010702. https://doi.org/10.1371/journal.pgen.1010702

Li, Y., Yao, J., Sang, H., Wang, Q., Su, L., Zhao, X., Xia, Z., Wang, F., Wang, K., Lou, D., Wang, G., Waterhouse, R. M., Wang, H., Luo, S., & Sun, C. (2024). Pan-genome analysis highlights the role of structural variation in the evolution and environmental adaptation of Asian honeybees. *Molecular Ecology Resources*, *24*(2), e13905. https://doi.org/10.1111/1755-0998.13905

Li, Z.-W., Hou, X.-H., Chen, J.-F., Xu, Y.-C., Wu, Q., González, J., & Guo, Y.-L. (2018). Transposable Elements Contribute to the Adaptation of Arabidopsis thaliana. *Genome Biology and Evolution*, *10*(8), 2140–2150. https://doi.org/10.1093/gbe/evy171

Lin, J., Jia, P., Wang, S., Kosters, W., & Ye, K. (2023). Comparison and benchmark of structural variants detected from long read and long-read assembly. *Briefings in Bioinformatics*, *24*(4), bbad188. https://doi.org/10.1093/bib/bbad188

Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research Commentary: Too Big to Fail: Large Samples and the p-Value Problem. *Information Systems Research*, *24*(4), 906–917. https://www.jstor.org/stable/24700283

Lockton, S., Ross-Ibarra, J., & Gaut, B. S. (2008). Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. *Proceedings of the National Academy of Sciences*, *105*(37), 13965–13970. https://doi.org/10.1073/pnas.0804671105

Loewenthal, G., Wygoda, E., Nagar, N., Glick, L., Mayrose, I., & Pupko, T. (2022). The evolutionary dynamics that retain long neutral genomic sequences in face of indel deletion bias: A model and its application to human introns. *Open Biology*, *12*(12), 220223. https://doi.org/10.1098/rsob.220223

Lotterhos, K. E. (2019). The Effect of Neutral Recombination Variation on Genome Scans for Selection. *G3 Genes|Genomes|Genetics*, *9*(6), 1851–1867. https://doi.org/10.1534/g3.119.400088

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, *30*(23), 5966–5993. https://doi.org/10.1111/mec.16077

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, *22*(5), 1678–1692. https://doi.org/10.1111/1755-0998.13559

Ma, C., Shao, M., & Kingsford, C. (2018). SQUID: Transcriptomic structural variation detection from RNA-seq. *Genome Biology*, *19*(1), 52. https://doi.org/10.1186/s13059-018-1421-5

Ma, L.-J., Cao, L.-J., Chen, J.-C., Tang, M.-Q., Song, W., Yang, F.-Y., Shen, X.-J., Ren, Y.-J., Yang, Q., Li, H., Hoffmann, A. A., & Wei, S.-J. (2024). Rapid and Repeated Climate Adaptation Involving Chromosome Inversions following Invasion of an Insect. *Molecular Biology and Evolution*, *41*(3), msae044. https://doi.org/10.1093/molbev/msae044

MacPherson, A., & Nuismer, S. L. (2017). The probability of parallel genetic evolution from standing genetic variation. *Journal of Evolutionary Biology*, *30*(2), 326–337. https://doi.org/10.1111/jeb.13006

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, *92*(1), 155–161. https://doi.org/10.1007/s12041-013-0212-8

Mahmoud, M., Gobet, N., Cruz-Dávalos, 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. https://doi.org/10.1186/s13059-019-1828-7

Martiniano, R., Garrison, E., Jones, E. R., Manica, A., & Durbin, R. (2020). Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. *Genome Biology*, *21*(1), 250. https://doi.org/10.1186/s13059-020-02160-7

Mc Cartney, A. M., Formenti, G., Mouton, A., De Panis, D., Marins, L. S., Leitão, H. G., Diedericks, G., Kirangwa, J., Morselli, M., Salces-Ortiz, J., Escudero, N., Iannucci, A., Natali, C., Svardal, H., Fernández, R., De Pooter, T., Joris, G., Strazisar, M., Wood, J. M. D., … Mazzoni, C. J. (2024). The European Reference Genome Atlas: Piloting a decentralised approach to equitable biodiversity genomics. *Npj Biodiversity*, *3*(1), 1–17. https://doi.org/10.1038/s44185-024-00054-6

McComish, B. J., Charleston, M. A., Parks, M., Baroni, C., Salvatore, M. C., Li, R., Zhang, G., Millar, C. D., Holland, B. R., & Lambert, D. M. (2024). Ancient and Modern Genomes Reveal Microsatellites Maintain a Dynamic Equilibrium Through Deep Time. *Genome Biology and Evolution*, *16*(3), evae017. https://doi.org/10.1093/gbe/evae017

Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., Box Power, O., Nadeau, N. J., Bridle, J. R., Rolian, C., Barton, N. H., McMillan, W. O., Jiggins, C. D., & Chan, Y. F. (2021). Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences*, *118*(25), e2015005118. https://doi.org/10.1073/pnas.2015005118

Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L., & Wellenreuther, M. (2020). Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature Communications*, *11*(1), 670. https://doi.org/10.1038/s41467-020-14479-7

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends in Ecology & Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Mérot, C., Stenløkk, K. S. R., Venney, C., Laporte, M., Moser, M., Normandeau, E., Árnyasi, M., Kent, M., Rougeux, C., Flynn, J. M., Lien, S., & Bernatchez, L. (2023). Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (Coregonus sp.) with long and short reads. *Molecular Ecology*, *32*(6), 1458–1477. https://doi.org/10.1111/mec.16468

Metzgar, D., Bytof, J., & Wills, C. (2000). Selection Against Frameshift Mutations Limits Microsatellite Expansion in Coding DNA. *Genome Research*, *10*(1), 72–80. https://doi.org/10.1101/gr.10.1.72

Meyer, L., Barry, P., Riquet, F., Foote, A., Der Sarkissian, C., Cunha, R. L., Arbiol, C., Cerqueira, F., Desmarais, E., Bordes, A., Bierne, N., Guinand, B., & Gagnaire, P.-A. (2024). Divergence and gene flow history at two large chromosomal inversions underlying ecotype differentiation in the long-snouted seahorse. *Molecular Ecology*, *33*(24), e17277. https://doi.org/10.1111/mec.17277

Morgan, A. P., Gatti, D. M., Najarian, M. L., Keane, T. M., Galante, R. J., Pack, A. I., Mott, R., Churchill, G. A., & de Villena, F. P.-M. (2017). Structural Variation Shapes the Landscape of Recombination in Mouse. *Genetics*, *206*(2), 603–619. https://doi.org/10.1534/genetics.116.197988

Munasinghe, M., Read, A., Stitzer, M. C., Song, B., Menard, C. C., Ma, K. Y., Brandvain, Y., Hirsch, C. N., & Springer, N. (2023). Combined analysis of transposable elements and structural variation in maize genomes reveals genome contraction outpaces expansion. *PLoS Genetics*, *19*(12), e1011086. https://doi.org/10.1371/journal.pgen.1011086

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Nicolas, L. A., Berdan, E. L., Colinet, H., Clouard, A., Wit, P. D., Glémin, S., & Mérot, C. (2025). *Parallel clines of chromosomal inversion in seaweed flies associated with thermal variation* (p. 2025.04.28.650981). bioRxiv. https://doi.org/10.1101/2025.04.28.650981

Nosil, P., de Carvalho, C. F., Villoutreix, R., Zamorano, L. S., Sinclair-Waters, M., Planidin, N. P., Parchman, T. L., Feder, J., & Gompert, Z. (2024). Evolution repeats itself in replicate long-term studies in the wild. *Science Advances*, *10*(21), eadl3149. https://doi.org/10.1126/sciadv.adl3149

Oliveira, D. S., Fablet, M., Larue, A., Vallier, A., Carareto, C. M. A., Rebollo, R., & Vieira, C. (2023). ChimeraTE: A pipeline to detect chimeric transcripts derived from genes and transposable elements. *Nucleic Acids Research*, *51*(18), 9764–9784. https://doi.org/10.1093/nar/gkad671

Oomen, R. A., Kuparinen, A., & Hutchings, J. A. (2020). Consequences of Single-Locus and Tightly Linked Genomic Architectures for Evolutionary Responses to Environmental Change. *Journal of Heredity*, *111*(4), 319–332. https://doi.org/10.1093/jhered/esaa020

Orteu, A., Kucka, M., Gordon, I. J., Ng'iru, I., van der Heijden, E. S. M., Talavera, G., Warren, I. A., Collins, S., ffrench-Constant, R. H., Martins, D. J., Chan, Y. F., Jiggins, C. D., & Martin, S. H. (2024). Transposable Element Insertions Are Associated with Batesian Mimicry in the Pantropical Butterfly Hypolimnas misippus. *Molecular Biology and Evolution*, *41*(3), msae041. https://doi.org/10.1093/molbev/msae041

Peischl, S., Kirkpatrick, M., & Excoffier, L. (2015). Expansion Load and the Evolutionary Dynamics of a Species Range. *The American Naturalist*, *185*(4), E81–E93. https://doi.org/10.1086/680220

Petrov, D. A. (2002). Mutational Equilibrium Model of Genome Size Evolution. *Theoretical Population Biology*, *61*(4), 531–544. https://doi.org/10.1006/tpbi.2002.1605

Rasmussen, M. D., Hubisz, M. J., Gronau, I., & Siepel, A. (2014). Genome-Wide Inference of Ancestral Recombination Graphs. *PLOS Genetics*, *10*(5), e1004342. https://doi.org/10.1371/journal.pgen.1004342

Recuerda, M., & Campagna, L. (2024). How structural variants shape avian phenotypes: Lessons from model systems. *Molecular Ecology*, *33*(11), e17364. https://doi.org/10.1111/mec.17364

Recuerda, M., Kraemer, S., Rosoni, J. R. R., Repenning, M., Browne, M., Cataudela, J. F., Giacomo, A. S. D., Kopuchian, C., & Campagna, L. (2025). *A pangenomic approach reveals the sources of genetic variation fueling the rapid radiation of Capuchino Seedeaters* (p. 2025.03.19.644153). bioRxiv. https://doi.org/10.1101/2025.03.19.644153

Reeve, J., Butlin, R. K., Koch, E. L., Stankowski, S., & Faria, R. (2024). Chromosomal inversion polymorphisms are widespread across the species ranges of rough periwinkles (Littorina saxatilis and L. arcana). *Molecular Ecology*, *33*(24), e17160. https://doi.org/10.1111/mec.17160

Rice, E. S., Alberdi, A., Alfieri, J., Athrey, G., Balacco, J. R., Bardou, P., Blackmon, H., Charles, M., Cheng, H. H., Fedrigo, O., Fiddaman, S. R., Formenti, G., Frantz, L. A. F., Gilbert, M. T. P., Hearn, C. J., Jarvis, E. D., Klopp, C., Marcos, S., Mason, A. S., … Warren, W. C. (2023). A pangenome graph reference of 30 chicken genomes allows genotyping of large and complex structural variants. *BMC Biology*, *21*(1), 267. https://doi.org/10.1186/s12915-023-01758-0

Rizk, G., Gouin, A., Chikhi, R., & Lemaitre, C. (2014). MindTheGap: Integrated detection and assembly of short and long insertions. *Bioinformatics*, *30*(24), 3451–3457. https://doi.org/10.1093/bioinformatics/btu545

Romain, S., & Lemaitre, C. (2023). SVJedi-graph: Improving the genotyping of close and overlapping structural variants with long reads using a variation graph. *Bioinformatics*, *39*(Supplement_1), i270–i278. https://doi.org/10.1093/bioinformatics/btad237

Romain, S., & Lemaitre, C. (2025). *SVJedi-tag—Forthcoming paper on new software* [Computer software].

Ruggieri, A. A., Livraghi, L., Lewis, J. J., Evans, E., Cicconardi, F., Hebberecht, L., Ortiz-Ruiz, Y., Montgomery, S. H., Ghezzi, A., Rodriguez-Martinez, J. A., Jiggins, C. D., McMillan, W. O., Counterman, B.

A., Papa, R., & Van Belleghem, S. M. (2022). A butterfly pan-genome reveals that a large amount of structural variation underlies the evolution of chromatin accessibility. *Genome Research*, *32*(10), 1862–1875. https://doi.org/10.1101/gr.276839.122

Saitou, M., Masuda, N., & Gokcumen, O. (2022). Similarity-Based Analysis of Allele Frequency Distribution among Multiple Populations Identifies Adaptive Genomic Structural Variants. *Molecular Biology and Evolution*, *39*(3), msab313. https://doi.org/10.1093/molbev/msab313

Sakamoto, T., Whiting, J. R., & Yeaman, S. (2024). *The persistence of locally adapted polymorphisms under mutation swamping* (p. 2024.06.18.599577). bioRxiv. https://doi.org/10.1101/2024.06.18.599577

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., & Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Molecular Biology and Evolution*, *37*(12), 3642–3653. https://doi.org/10.1093/molbev/msaa169

Scarpa, A., Pianezza, R., Wierzbicki, F., & Kofler, R. (2024). Genomes of historical specimens reveal multiple invasions of LTR retrotransposons in Drosophila melanogaster during the 19th century. *Proceedings of the National Academy of Sciences*, *121*(15), e2313866121. https://doi.org/10.1073/pnas.2313866121

Scherer, S. W., Lee, C., Birney, E., Altshuler, D. M., Eichler, E. E., Carter, N. P., Hurles, M. E., & Feuk, L. (2007). Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, *39*(7), Article 7. https://doi.org/10.1038/ng2093

Schöpflin, R., Melo, U. S., Moeinzadeh, H., Heller, D., Laupert, V., Hertzberg, J., Holtgrewe, M., Alavi, N., Klever, M.-K., Jungnitsch, J., Comak, E., Türkmen, S., Horn, D., Duffourd, Y., Faivre, L., Callier, P., Sanlaville, D., Zuffardi, O., Tenconi, R., … Mundlos, S. (2022). Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes. *Nature Communications*, *13*(1), 6470. https://doi.org/10.1038/s41467-022-34053-7

Scott, A. J., Chiang, C., & Hall, I. M. (2021). Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes. *Genome Research*, *31*(12), 2249–2257. https://doi.org/10.1101/gr.275488.121

Secomandi, S., Gallo, G. R., Rossi, R., Rodríguez Fernandes, C., Jarvis, E. D., Bonisoli-Alquati, A., Gianfranceschi, L., & Formenti, G. (2025). Pangenome graphs and their applications in biodiversity genomics. *Nature Genetics*, *57*(1), 13–26. https://doi.org/10.1038/s41588-024-02029-6

Secomandi, S., Gallo, G. R., Sozzoni, M., Iannucci, A., Galati, E., Abueg, L., Balacco, J., Caprioli, M., Chow, W., Ciofi, C., Collins, J., Fedrigo, O., Ferretti, L., Fungtammasan, A., Haase, B., Howe, K., Kwak, W., Lombardo, G., Masterson, P., … Formenti, G. (2023). A chromosome-level reference genome and pangenome for barn swallow population genomics. *Cell Reports*, *42*(1), 111992. https://doi.org/10.1016/j.celrep.2023.111992

Serrato-Capuchina, A., & Matute, D. R. (2018). The Role of Transposable Elements in Speciation. *Genes*, *9*(5), 254. https://doi.org/10.3390/genes9050254

Shipilina, D., Pal, A., Stankowski, S., Chan, Y. F., & Barton, N. H. (2023). On the origin and structure of haplotype blocks. *Molecular Ecology*, *32*(6), 1441–1457. https://doi.org/10.1111/mec.16793

Smeds, L., Huson, L. S. A., & Ellegren, H. (2024). Structural genomic variation in the inbred Scandinavian wolf population contributes to the realized genetic load but is positively affected by immigration. *Evolutionary Applications*, *17*(2), e13652. https://doi.org/10.1111/eva.13652

Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker Open-4.0* (Version http://www.repeatmasker.org) [Http://www.repeatmasker.org]. http://www.repeatmasker.org

Sopniewski, J., & Catullo, R. A. (2024). Estimates of heterozygosity from single nucleotide polymorphism markers are context-dependent and often wrong. *Molecular Ecology Resources*, *24*(4), e13947. https://doi.org/10.1111/1755-0998.13947

Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, *51*(9), 1321–1329. https://doi.org/10.1038/s41588-019-0484-x

Spielmann, M., Lupiáñez, D. G., & Mundlos, S. (2018). Structural variation in the 3D genome. *Nature Reviews Genetics*, *19*(7), 453–467. https://doi.org/10.1038/s41576-018-0007-0

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, *24*(9), 2241–2252. https://doi.org/10.1111/mec.13089

Stern, D. B., & Lee, C. E. (2020). Evolutionary origins of genomic adaptations in an invasive copepod. *Nature Ecology & Evolution*, *4*(8), 1084–1094. https://doi.org/10.1038/s41559-020-1201-y

Stewart, N. B., & Rogers, R. L. (2019). Chromosomal rearrangements as a source of new gene formation in Drosophila yakuba. *PLoS Genetics*, *15*(9), e1008314. https://doi.org/10.1371/journal.pgen.1008314

Stuart, K. C., Edwards, R. J., Sherwin, W. B., & Rollins, L. A. (2023). Contrasting Patterns of Single Nucleotide Polymorphisms and Structural Variation Across Multiple Invasions. *Molecular Biology and Evolution*, *40*(3), msad046. https://doi.org/10.1093/molbev/msad046

Stuart, K. C., Santure, A. W., & Rollins, L. A. (2025). Structural variants and transposable elements as facilitators of rapid evolutionary change in invasive populations. In *Invasion Genomics*.

Stuart, K. C., Tan, H. Z., Whibley, A., Bailey, S., Brekke, P., Ewen, J. G., Patel, S., & Santure, A. W. (2025). Both structural variant and single nucleotide polymorphism load impact lifetime fitness in a threatened bird species. *Molecular Ecology*. https://doi.org/10.1111/mec.17631

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—Or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279. https://doi.org/10.4300/JGME-D-12-00156.1

Tepolt, C. K., Grosholz, E. D., de Rivera, C. E., & Ruiz, G. M. (2022). Balanced polymorphism fuels rapid selection in an invasive crab despite high gene flow and low genetic diversity. *Molecular Ecology*, *31*(1), 55–69. https://doi.org/10.1111/mec.16143

Therkildsen, N. O., Wilder, A. P., Conover, D. O., Munch, S. B., Baumann, H., & Palumbi, S. R. (2019). Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science*, *365*(6452), 487–490. https://doi.org/10.1126/science.aaw7271

Tigano, A., Jacobs, A., Wilder, A. P., Nand, A., Zhan, Y., Dekker, J., & Therkildsen, N. O. (2021). Chromosome-Level Assembly of the Atlantic Silverside Genome Reveals Extreme Levels of Sequence Diversity and Structural Genetic Variation. *Genome Biology and Evolution*, *13*(6), evab098. https://doi.org/10.1093/gbe/evab098

Tigano, A., Reiertsen, T. K., Walters, J. R., & Friesen, V. L. (2018). *A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird* (p. 507384). bioRxiv. https://doi.org/10.1101/507384

Tigano, A., & Russello, M. A. (2022). The genomic basis of reproductive and migratory behaviour in a polymorphic salmonid. *Molecular Ecology*, *31*(24), 6588–6604. https://doi.org/10.1111/mec.16724

Tigano, A., Weir, T., Ward, H. G. M., Gale, M. K., Wong, C. M., Eliason, E. J., Miller, K. M., Hinch, S. G., & Russello, M. A. (2024). Genomic vulnerability of a freshwater salmonid under climate change. *Evolutionary Applications*, *17*(2), e13602. https://doi.org/10.1111/eva.13602

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Tom, J. A., Reeder, J., Forrest, W. F., Graham, R. R., Hunkapiller, J., Behrens, T. W., & Bhangale, T. R. (2017). Identifying and mitigating batch effects in whole genome sequencing data. *BMC Bioinformatics*, *18*(1), 351. https://doi.org/10.1186/s12859-017-1756-z

Trickett, A. J., & Butlin, R. K. (1994). Recombination suppressors and the evolution of new species. *Heredity*, *73 (Pt 4)*, 339–345. https://doi.org/10.1038/hdy.1994.180

Vara, C., Paytuví-Gallart, A., Cuartero, Y., Álvarez-González, L., Marín-Gual, L., Garcia, F., Florit-Sabater, B., Capilla, L., Sanchéz-Guillén, R. A., Sarrate, Z., Aiese Cigliano, R., Sanseverino, W., Searle, J. B., Ventura, J., Marti-Renom, M. A., Le Dily, F., & Ruiz-Herrera, A. (2021). The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. *Nature Communications*, *12*(1), 2981. https://doi.org/10.1038/s41467-021-23270-1

Vigouroux, Y., Jaqueth, J. S., Matsuoka, Y., Smith, O. S., Beavis, W. D., Smith, J. S. C., & Doebley, J. (2002). Rate and Pattern of Mutation at Microsatellite Loci in Maize. *Molecular Biology and Evolution*, *19*(8), 1251–1260. https://doi.org/10.1093/oxfordjournals.molbev.a004186

Weetman, D., Djogbenou, L. S., & Lucas, E. (2018). Copy number variation (CNV) and insecticide resistance in mosquitoes: Evolving knowledge or an evolving problem? *Current Opinion in Insect Science*, *27*, 82–88. https://doi.org/10.1016/j.cois.2018.04.005

Wellband, K., Mérot, C., Linnansaari, T., Elliott, J. a. K., Curry, R. A., & Bernatchez, L. (2019). Chromosomal fusion and life history-associated genomic variation contribute to within-river local adaptation of Atlantic salmon. *Molecular Ecology*, *28*(6), 1439–1459. https://doi.org/10.1111/mec.14965

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

Westram, A. M., Faria, R., Johannesson, K., Butlin, R., & Barton, N. (2022). Inversions and parallel evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1856), 20210203. https://doi.org/10.1098/rstb.2021.0203

Wohns, A. W., Wong, Y., Jeffery, B., Akbari, A., Mallick, S., Pinhasi, R., Patterson, N., Reich, D., Kelleher, J., & McVean, G. (2022). A unified genealogy of modern and ancient genomes. *Science*, *375*(6583), eabi8264. https://doi.org/10.1126/science.abi8264

Wold, J. R., Guhlin, J. G., Dearden, P. K., Santure, A. W., & Steeves, T. E. (2023). The promise and challenges of characterizing genome-wide structural variants: A case study in a critically endangered parrot. *Molecular Ecology Resources*, *n/a*(n/a). https://doi.org/10.1111/1755-0998.13783

Wold, J. R., Koepfli, K.-P., Galla, S. J., Eccles, D., Hogg, C. J., Le Lec, M. F., Guhlin, J., Santure, A. W., & Steeves, T. E. (2021). Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. *Molecular Ecology*, *30*(23), 5949–5965. https://doi.org/10.1111/mec.16141

Wu, Z., Wu, Y., & Gao, J. (2020). InvBFM: Finding genomic inversions from high-throughput sequence data based on feature mining. *BMC Genomics*, *21*(1), 173. https://doi.org/10.1186/s12864-020-6585-1

Xi, R., Kim, T.-M., & Park, P. J. (2010). Detecting structural variations in the human genome using next generation sequencing. *Briefings in Functional Genomics*, *9*(5–6), 405–415. https://doi.org/10.1093/bfgp/elq025

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, *110*(19), E1743–E1751. https://doi.org/10.1073/pnas.1219381110

Zarate, S., Carroll, A., Mahmoud, M., Krasheninina, O., Jun, G., Salerno, W. J., Schatz, M. C., Boerwinkle, E., Gibbs, R. A., & Sedlazeck, F. J. (2020). Parliament2: Accurate structural variant calling at scale. *GigaScience*, *9*(12), giaa145. https://doi.org/10.1093/gigascience/giaa145

Zhang, L., Chaturvedi, S., Nice, C. C., Lucas, L. K., & Gompert, Z. (2023). Population genomic evidence of selection on structural variants in a natural hybrid zone. *Molecular Ecology*, *32*(6), 1497–1514. https://doi.org/10.1111/mec.16469

Zhang, L., Reifová, R., Halenková, Z., & Gompert, Z. (2021). How Important Are Structural Variants for Speciation? *Genes*, *12*(7), 1084. https://doi.org/10.3390/genes12071084

Zhao, X., Collins, R. L., Lee, W.-P., Weber, A. M., Jun, Y., Zhu, Q., Weisburd, B., Huang, Y., Audano, P. A., Wang, H., Walker, M., Lowther, C., Fu, J., Gerstein, M. B., Devine, S. E., Marschall, T., Korbel, J. O., Eichler, E. E., Chaisson, M. J. P., ... Talkowski, M. E. (2021). Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. *The American Journal of Human Genetics*, *108*(5), 919–928. https://doi.org/10.1016/j.ajhg.2021.03.014

Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., Zhang, J., Lyu, H., Lin, T., Gao, Q., Saha, S., Mueller, L., Fei, Z., Städler, T., Xu, S., ... Huang, S. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. *Nature*, *606*(7914), 527–534. https://doi.org/10.1038/s41586-022-04808-9