Resolving large-scale genome evolution in the high-throughput sequencing era: structural variants, genome rearrangement, and karyotype dynamics in animals

Celian Diblasi[1], Nicola Barson[1], Marie Saitou[1]*

Running Title: Structural and genomic variants evolution in animals

Keywords: Evolution, supergenes, recombination, karyotypic variants, genomic rearrangements

Author affiliations

[1]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Norway

Corresponding Author: marie.saitou@nmbu.no

Abstract

Genomic structural variation, genomic rearrangements and karyotype variants are important components on which evolution acts in addition to single nucleotide variants, which are the most common type of variant being studied since the rise of next generation sequencing. These variants have unique mutational mechanisms and evolutionary consequences compared to single nucleotide variants. Here we review the history and methods used to study genomic structural variants in animals, including technical challenges of current methods and promising new approaches. We then review case studies to illustrate how large genomic variants can drive interesting evolutionary phenomena. Meanwhile, most studies are limited to specific taxa and there remain many unanswered questions. We conclude that the time is ripe to expand study on genomic structural variants to various non-model species and across taxa by utilizing cutting edge technologies to open a new door in evolutionary genomics.

Preface: The value of studying genomic structural variants in evolution and ecology

One of the fundamental goals of ecology and evolution is to clarify the contribution of genetic mutations to phenotypic change within populations [1–3]. Recent technological breakthroughs in massively parallel sequencing have enabled researchers to investigate genetic variants in a genome-wide manner on a population scale, sometimes based on high-quality reference genomes [4–6]. At the same time, however, most of these genome-wide studies have focused only on single nucleotide variants, which are relatively easy to identify and analyze with established pipelines, however, this has led to other types of genetic variants with potentially high impacts being largely neglected [7–9].

Genomic structural variants and chromosomal rearrangements, including inversions encompassing multiple genes, differences in gene copy number, and polymorphic chromosome fission and fusion, are thought to have an important impact on phenotypes and genome evolution [10]. Recent findings that highlight the enormous potential importance of genomic structural variants in ecology and evolution [11–13]. However, in current standard pipelines for population and evolutionary genomics, such structural variants are often categorized as “complex exceptions” and as such tend to be excluded from the analysis [14]. In addition, the laborious and time-consuming investigation of polymorphic chromosome rearrangement that had been conducted intensively earlier has gradually shrunk in the high-throughput DNA sequencing era.

Bringing structural variants and genomic rearrangement back to the spotlight of the latest genome-wide methods will be greatly beneficial in the field of evolution and ecology, not only to find out the undiscovered genetic basis of phenotypic diversity but also to unveil the mechanism of large-scale genome evolution, hybridization, reproductive isolation, and speciation [14]. Here, we outline how the latest genomics and analytical algorithms can detect various types of genomic structural variants, and how these can unveil the genetic basis of organism ecology and evolution. These variants are present in all organisms, even though they differ in their frequency in different taxa, but we will here restrict ourselves to the animal kingdom.

1. Insights into genomic structural variants from early and recent technologies

1.1 The overview of genomic structural variants

Genomic mutation is the substrate of evolutionary dynamics and biological diversity. The majority of

genetics studies, both theoretical and empirical, have been focused on only one type of mutation, Single Nucleotide Polymorphisms (SNPs), which consist of a single base change (**Figure 1**). Because of their simple and short nature, detecting SNPs has become easy and cost-effective [15,16]. Numerous custom SNP arrays are available for extensively studied species, and sometimes for particular breeds and populations, to investigate their population genomics based on thousands or millions of SNPs [16]. However, other types of genomic variants, which we call Structural Variants (SV), have been shown to possess major influences on both evolution and phenotype [17,18]. While SNPs are known to account for 0.1% of differences between two humans, this number grows to 1.5% when taking SVs into account [19].

There are various types of structural variants, such as deletions, duplications, and insertions, which are generally defined as a change in the sequence of more than 50 bp (**Figure 1**) [19–21]. This size definition is simply for convenience and conventional, and there is no intrinsic significance in the 50-base cutoff, and certainly, there is no substantial problem with calling a 48-base deletion a structural variant. The size is defined to focus on those variants that would have been overlooked in short-read sequencing, however, a common practice is to filter out indels and focus on SNPs for analysis. Some SVs, such as duplications and deletions, can cause copy number variation of genes, meaning the number of copies of a gene can also vary among individuals or populations [21–23]. Translocations and inversions are more complex SVs. A translocation is a sequence that is moved from one location to another, often mediated by transposable elements or non-allelic non homologous recombination (NAHR) [21]. An inversion is a genetic segment that breaks off and is reinserted in the same location but in the reverse orientation [24]. By suppressing recombination, inversions can keep multiple genes within tight genetic linkage and can be inherited as a single locus, creating supergenes that have been associated with multiple ecologically important traits in various species [17,21,25].

There are even larger genomic variants, karyotypic variants. The most drastic case of karyotypic variants is whole genome duplication (WGD), leading to a doubling of the number of chromosomes and all the genetic content [26]. More frequently observed, smaller-scale karyotypic variants are chromosomal fusion and fission. A fusion between two chromosomes can happen, thereby reducing the number of chromosomes in the species and linking previously independent sequences (**Figure 2**). Chromosomal

fusion can be classified into different types, depending on the centromere position before and after the fusion (**Figure 2**) [27–29]. Chromosomal fission is the opposite phenomenon, where a single chromosome splits into two chromosomes [30,31].

1.2 A brief overview of historical discoveries of karyotype dynamics and structural variants

SVs were described more than a century ago, in the classic genetic mapping era in the 1920s [32], by Alfred H. Sturtevant, who established the foundations of modern biology by constructing the first genetic map in Drosophila. Sturtevant noticed that genes were not located in the same order among closely related flies using genetic mapping and he proposed that chromosomal inversion would explain the observed gene order differences [32]. Another example of SV presumption before the rise of observation methods is the case of supergenes, which were predicted by the observation of color pattern in recombinant and crosses of mimetic butterflies [33]. In early studies, SVs were detected by observing karyotypes in cell culture arrested at the metaphase of cell division, especially by using chromosome banding [34–37] (**Table 1**). This method has a low-resolution power and allows detection of only large SVs, but is less efficient than some methods developed later, such as fluorescent in situ hybridization or array CGH (**Table 1**) [10,35,36,38,39]. All of these techniques detect sufficiently large variants, however, have a low resolution for smaller variants.

Karyotype research flourished in the field of human medical genomics, but early findings were mainly uninheritable karyotypic variants, such as trisomy 21. Trisomy 21, commonly known as Down syndrome, is a genetic disorder caused by the presence of an extra chromosome 21 in humans that was first associated with the karyotype polymorphism in 1959 [40,41]. In the context of evolution, in 1970, Susumu Ohno proposed that gene duplication is the major driver of evolution [42]. Based on comparisons of genome size, he speculated that there were two rounds of complete genome duplication events in the early vertebrate lineage that would have allowed for vertebrate diversification [43]. However, at that time there was a little data to support or reject this hypothesis. This remained the case until the 1990s when multiple genetic analyses, including the discovery of paralogous regions among mammals, and the presence of 4 hox gene clusters in mammals (where only one is present in drosophila), brought support for Ohno's hypothesis [44,45]. The rise of sequencing data in the 2000s brought even more support to this

hypothesis, with discovery of multiple duplicated genes, reviewed in section 3 [43].

1.3 Exploiting short and long-read high-throughput sequencing for ecological and evolutionary genomics of structural variants

High-throughput sequencing technology has revolutionized genomics by enabling genome-wide scale analysis. Sequencing technologies are categorized as short- and long-read sequencing. Short-read sequencing, which is affordable and widely used for population-scale genomics, has limitations in detecting large or complex structural variants such as inversions, tandem repeats, insertions, and translocations [19,20]. Nevertheless, several methods have been developed to detect these SVs, principally by mapping reads to the reference sequence and detecting SV breakpoints [21] (**Table 1**).

After the emergence of short-read sequencing technologies, human geneticists immediately started to investigate the impact of genetic variants, including SVs, on human diseases, such as Alzaheimer's and Parkinson's diseases [39,46–49]. The 1000 genome project was launched as one of the first population-scale genomics projects and provided great advancement in cataloging SVs [50]. Building on such early-stage, short-read-based methods for SV detection and analysis, the field is now expanding to more evolutionary-related research questions on various species, such as linking recombination hotspots and SVs[51], the divergence between ecotypes[52], and the importance of SVs in domestication and selected phenotypes[53].

Long-read sequencing provides an improvement in detecting SVs, particularly in complex or repeated regions where many SVs are present [54–56] (**Table 1**). Algorithms have been developed to detect SVs using short-read and/or long-read sequencing data sets [57–59]. Nevertheless, there is no consensus method, silver bullet, for all types of research questions, partly because most of the software was developed based on human data, and some are specifically developed for somatic variants detection in cancer cells. To apply the methods for ecological and evolutionary genomics, we need to carefully consider species-specific genomic architecture, repeat structure, study questions, and available

resources such as reference genome quality. Recently, machine-learning methods have been developed for SVs detection and confidence-based filtering [60]. Hi-C, which investigates the 3D organization of the genome, is also increasingly used to detect SVs. In particular, Hi-C has been used in repetitive regions, translocations, and large variants [10,61], where abnormal interactions between regions may signal the presence of SVs[61–63] (**Table 1**). The methods to analyze SVs evolution have also recently improved, with the possibility to simulate evolution of SVs [64,65] or investigate population genomics of SVs [66].

2. The mutational mechanism of structural variants and genome rearrangements

Mutation is the source of variation for evolution, providing the variation for selection to act upon to create adaptation and phenotypic diversity. One of the main characteristics of SVs compared with SNPs is that SVs can be more often generated recurrently than SNPs since SV formation depends on genome architecture such as recombination, segmental duplications and tandem repeats [67]. In cases when variation can be beneficial, for example, resistance to pathogens, multiple recurrent variants are reported and such variants can function as a source of genetic variation under balancing selection [68–71].

The various types of SVs can be generated by multiple mutational mechanisms. The principal SV formation mechanism is recombination, especially NAHR, but other formation mechanisms may occur during DNA repair including non-homologous end joining (NHEJ), microhomology-mediated break-induction replication (MMBIR) and Fork stalling and template switching (FoSTeS) [21,36,67]. Some SVs are more likely to be formed by specific mechanisms. such as tandem duplication, which is often caused by slipped-strand mispairing during replication [72]. Transposable elements (TEs) are also involved in SVs formation. Indeed, they have been shown to be associated with hotspots of recombination events, particularly NAHR events [73]. Alu, LINE and HERV elements are particularly known to be associated with SVs formation [36,67,73]. Genomic architecture also has an impact on SVs formation, with a greater probability to be formed in highly repetitive regions [67]. Low copy repeats (LCRs) , by increasing the probabilities of NAHR events because of frequent recombination, are also associated with more SV-rich

regions [67]. Sex chromosomes may also be more prone to SVs formation because they generally have no homolog to pair with during meiosis [74,75].

Chromosomal fusions occur when telomere ends of chromosomes get lost or shortened, even in the case of Robertsonian fusion involving the fusion of two acrocentric chromosomes (**Figure 2C**) [29]. NHEJ is the main mechanism responsible for this telomere deprotection and chromosome fusion in mammals, as it is the main pathway to repair unprotected DNA ends [76,77]. Telomeres can also lose their sheltering proteins which can lead to chromosome fusion [76]. Other mechanisms, such as the presence of inverted repeats or single-strand annealing (SSA), which is another DNA repairing mechanism, are known to be responsible of chromosome fusion in non-animal species, while we still are unsure if they play a similar role in animals [76]. For their stability fused chromosomes cannot maintain two centromeres. This issue can be solved in different ways, depending on the type of fusion. In end-to-end fusion, two telomeric chromosomes are fused together (**Figure 2**). This leads to a chromosome with two centromeres (called a dicentric chromosome), and a centromere has to be inactivated or eliminated for the fusion to stay stable [78]. In some cases, a small part of a telomeric chromosome containing the centromere is not fused, often lost through subsequent meiosis, avoiding the formation of a dicentric fused chromosome (**Figure 2**) [79]. Chromosome fission mechanisms rely on centromere modifications. The major mechanism is simple centric fission; while centric duplication-fission, where the centromere is pre-duplicated, or centric activation-fission, where a new centromere is activated, can cause similar consequences [31].

A whole genomic duplication is a rare event, sometimes, large-scale duplications lead to imperfect duplication and sterile offspring. For example, when diploid gametes may be produced, two gametes need to be diploid to form the whole genome duplicated offspring, which is unlikely to happen [80]. One realistic scenario is normal fertilization followed by reduplication during the first division of the zygote with all genetic material duplicated, as it can be observed in some mammals [80,81]. Hybridization of two closely related species can also lead to genome content duplication, without a subsequent proper meiosis with ploidy reduction, producing diploid gametes at a high rate that can then cross and form polyploid individuals, which is known to happen in fish, amphibians and reptiles [80]. Genomes can also undergo a reduction in size, by loss of genes and non-coding sequence, which sometimes can be a

consequence of accumulation of deleterious mutations [82]. This phenomenon is commonly known in bacteria species, but also in some worms and birds [83–86]. The genome size reduction is particularly investigated in multiple parasite taxa, because genome compaction is associated with the unique evolutionary pressures due to their ecology [82,83,87].

205

206 3. The evolutionary fate of genomic structural variants and chromosomal dynamics

207 3.1- The evolutionary trajectory and phenotypic effects of SVs

208 Similar to SNPs, the majority of newly arising SVs are expected to disappear without being passed on
209 from generation to generation. However, SVs can make unique evolutionary contributions because they
210 can cover larger genomic segments with different functional effects than SNPs, including gene fusion,
211 inversion/deletion/duplication of one or multiple genes, and mechanistically expanding tandem repeat
212 sequencing [88,89]. For the same reasons, and also because they differ from SNPs which can have no
213 effect on genes because of the redundancy of genetic code, they are also more likely to be highly
214 deleterious, especially in coding sequences. SVs can also affect gene dosage, by affecting gene copy
215 numbers and regulatory regions [35,62] and gene expression by altering genomic architecture through
216 "position effects" [35,62]. The new frontier of this field is the effect of SVs on the 3D organization of the
217 genome, including the disruption of boundaries between interacting regions[62].

218 Supergenes can be formed by SVs that keep beneficial alleles together by suppressing recombination.
219 They are typically maintained as polymorphism by balancing selection, which can occur in multiple forms
220 [17]. This is exemplified in *Heliconius numata* butterflies, where multiple successive inversions have led
221 to the emergence of a supergene controlling mimicry wing pattern, where alleles of different genes have
222 to be maintained together to ensure the mimetic pattern is produced [90]. This supergene is maintained
223 by disassortative mating, a type of frequency dependant selection,, where females prefer males with
224 different wing patterns that facilitates the generation of heterozygous offspring [91]. Supergenes and
225 inversions can also be maintained by overdominance or associative overdominance, the latter appearing
226 in the scenario where different recessive deleterious alleles or when dominant advantageous alleles are
227 fixed on each supergene haplotype [92]. Another important consideration for supergene and inversion
228 evolution is that any deleterious mutations generated will not have many opportunities of being
229 eliminated by recombination. Therefore, in the long term, inversions can be also deleterious because of
230 the accumulation of deleterious mutations [64,92–95]. Additional advantageous inversions may happen in
231 close proximity of an existing advantageous supergene or inversion, which may extend the
232 supergene/inversion non-recombining region, leading to "evolutionary strata" [93,96,97].

Duplicated genes increase the production of gene transcripts [98], which can often be deleterious [42,99] and require the restoration of well-balanced gene expression. This dosage modification can be achieved by the degeneration of one copy (pseudogenization) [100], or the duplicated copy can also acquire a new function (neofunctionalization) [42,100]. Sometimes both gene copies only assume a part of the original function (subfunctionalization) [100,101]. There has been a long debate over which process is most important in the evolution of duplicated genes, with recent observations accumulating in favor of neofunctionalization. [99,102]. Gene gain and loss can evolve across species and characterize lineage-specific traits. A famous example is the Amylase gene, which is responsible for starch digestion. In addition to marked copy number polymorphisms between human populations with high-starch diets and those with low-starch diets, Amylase underwent copy number gain in multiple domestic mammalian lineages, which consume more starch than their wild counterparts [103,104].

Genes losses can also facilitate adaptation [105]. Genes can be lost because some are dispensable, especially in the case of genes that are specific to some environment [106]. A common situation where this occurs is for animals living in perpetual dark environments such as caves, where skin pigmentation and eyes are no longer needed to survive, and loss of genes related to these functions are often observed in such animals [107–110]. It is interesting to notice that all genes are not equally likely to be lost, and factors such their function or position influence their probability of being lost [105,106,111].

Transposable elements (TEs) contribute to SV formation. TEs have their own evolutionary phases, where they are initially invasive in the host genome, but after some time, they reach an equilibrium of invasion and purging [112,113]. This can impact the SV formation dynamics in the host genome, where more SVs are generated in the initial invasion period, with regression in generation rate as the equilibrium is reached. TEs can also bring new functions in the host genome [114,115], starting from the emergence of the TE-mediated phenotype, followed by selection of this TE and the immobilization of the TE in the genome [116]. Some notable examples include the RAG genes in mammals, which play a role in recombination of immune genes, and Drosophila telomeres, which are composed of inactivated TEs [117–119]. TEs also can have various impacts on genome structure, such as modification of recombination rate or genome size [120,121]. By increasing genome size, TEs can increase energy consumption during genome replication [121,122]. TEs are, thus, a major component target for genome size reduction [123].

It has recently become increasingly clear that repetitive sequences, previously thought to have no function and eliminated from analysis due to their complexity, contribute to phenotypic and adaptive evolution in a variety of species. In shrimps, simple repetitive sequence expansions play a role in genomic plasticity and osmoregulatory capacity through modulating gene expression [124]. Another case suggests that a large repeat-rich block with multiple TEs which is not yet assigned to a chromosome, is associated with songbird subspecies with different migrator behaviour [125]. These cases suggest that a considerable amount of evolutionarily important genetic factors are located in repeat-rich regions, which have been overlooked by conventional studies due to technological limitations. Application of the latest methods, such as targeted long-read sequencing, promise to unveil the evolutionary significance of such repeat sequences.

3.2 The evolution of larger genomic rearrangements: Karyotype evolution, whole genome duplication, sex chromosomes and B chromosomes

After a chromosome fusion or fission appears, the karyotype first has to be stabilized to be maintained in the population over generations. It is important to ensure that there is only one centromere on the chromosome, since chromosome fusions may lead to chromosome segregation errors during meiosis including aneuploidy [75]. Another consequence of fusions is that they create longer chromosomes, which may have an impact on the cell biomechanisms, for example, the centromere may need to become stronger to ensure proper mobility of the chromosome. As a consequence, the number of kinetochores microtubules and the size of the kinetochore is likely to increase [78]. Chromosome fusions are also known to reduce recombination rate [126,127], which can lead to various evolutionary consequences. For instance, chromosome fusion can bring a significant selective advantage when linkage disequilibrium is favored, such as by linking locally adapted loci[128].

De novo karyotype mutations are often observed in human embryos, but such individuals do not survive until birth, and many chromosome number mutations are considered lethal. Chimpanzees with equivalent trisomy have been found and they share some symptoms with those of Down syndrome in humans [129]. This report is insightful in whether specific *de novo* chromosomal trisomy with low lethality is common across closely related species, and how the chromosome dosage affects phenotype in closely related species. However, *de novo* karyotype variation at the population level has rarely been examined in non-human animal species, leaving many open evolutionary questions, such as their effects on phenotype or reproductive success.

WGD doubles the entire set of chromosomes and thus has important biological and evolutionary consequences, including facilitating speciation[130]. It has the immediate consequence of doubling genome size, which as mentioned earlier, increases the cost of cell multiplication [131,132]. Moreover, cell size and cell volume often increase after a genome duplication with various metabolic consequences [132]. This cell size increase also can lead to the “gigantism” effect, with polyploid individuals having larger

300 body or organ size [131]. WGD is also a duplication of all genes at a single time, and similar to the
301 previously mentioned case for gene duplication, but more intensified dosage compensation happens
302 after WGD [133]. Species that undergo a whole genome duplication are thus subject to a rediploidization,
303 where most duplicate gene copies are lost after some time [133–135]. Sex chromosomes and B
304 chromosomes are other karyotypic variations that should be noted. Sex chromosomes are more extreme
305 cases of supergenes, and some supergenes even display some sex chromosome-like behaviour [136].
306 Differentiated sex chromosomes often have a phase of degeneration (loss of functional genes) over
307 time [137–139] and are characterized by a non-recombining region with its homolog, as in supergenes or
308 inversions [93,138]. B chromosomes are supernumerary chromosomes, which are mainly composed of
309 repeated sequences, although they can contain some genes [140,141]. B chromosomes are usually
310 considered to reduce the fitness of the host, due to the resources needed to be replicated, however,
311 they are sometimes preserved, when they contain important genetic elements [140–142]. Some sex and B
312 chromosomes also undergo non-mendelian inheritance, which is known to have important evolutionary
313 impacts in similar circumstances as with cytoplasmically inherited elements or segregation distorters [143–
314 145].

315

316 4. Concluding remarks - Unanswered questions and future directions

317 As we have reviewed, SVs and karyotypic variants are an important part of genomic variation, which
318 can have various phenotypic and evolutionary consequences. They are still overlooked compared to
319 SNPs, partially because of the difficulty to sequence them in a cost effective way. Recent advances in
320 sequencing technology with long reads and detection algorithms will certainly make SVs studies more
321 reliable and affordable, and are paving the way for the new era for evolutionary genomics research [146].
322 One of the main challenges is to improve integration of SVs and SNPs data, which is currently often
323 avoided because of the different informatic coding used for these variants [14]. More methods to analyze
324 SVs evolution are also needed, in order to take into account the full ranges of genomic variation and
325 have multiple indicators for evolutionary analysis as well as more reliable results [14]. SVs are also often
326 considered on their own, but in reality they are often interacting between each other. Sex chromosomes
327 are for instance often encountered with inversions, duplication or deletion [137,147,148]. Evolution of this type

of structure might lead to specific dynamics which need to be investigated and considered. SVs have not been studied in a broad taxonomic range, therefore, deeper and wider investigations on different organisms will significantly improve our knowledge on SVs content in the genome and their maintenance. We also still lack information on how different types of SVs play a role in various evolutionary processes such as speciation or divergence between populations or closely related species. The impact of these variants on gene expression is still an active field, and the evolution of gene expression after gene/genome duplication is not fully understood yet. The recents advancement in 3D genome conformation is particularly relevant to investigate how SVs can alter genome conformation creating different expression patterns [149]. Overall, the study of SVs and karyotypic variants is a fascinating area of research that holds great potential for advancing our understanding of genome evolution. With continued advancements in theory, technology and new analytical tools, we can look forward to many unseen discoveries.

Acknowledgements

We thank Dr. Simen Sandve at Faculty of Biosciences, Norwegian University of Life Sciences for valuable discussion. The study was supported by The Research Council of Norway (grant nos. 275310 and 221734).

Glossary

Genomic variant: Variation in the DNA sequence compared to a reference sequence.

Structural variant (SV): Genomic variant that covers multiple base pairs.

Non-allelic homologous recombination (NAHR): Recombination events between similar sequences (Homologous recombination) but at different genomic positions.

Karyotypic variant: Variant that impacts the number or conformation of the karyotype.

Polymorphic variant: variant that is present (in a population, or in a group of species considered) in several forms (for instance, some individuals have a deletion, others have not).

Non-homologous end joining (NHEJ): Broken DNA is repaired with ligation without homologous template.

Microhomology-mediated break-induction replication (MMBIR): During replication, if DNA is broken and has to be repaired, a region of microhomology can invade a nearby replication fork.

Fork stalling and template switching (FoSTeS): During replication, if a fork is stalling, the lagging strand might disengage and invade a nearby replicating fork.

Alu: Most abundant family of TE in human genome, of around 300bp long, classified as short interspersed nuclear element.

Long interspersed nuclear element (LINE):A family of retrotransposons.

Human endogenous retroviral element (HERV): Sequences derived from ancient retroviruses.

Single-strand annealing (SSA): DNA repair mechanism in which a long 5' single strand is paired with another single stranded DNA on the another chromosome.

Polyploidy: When the chromosomes are in more than 2 copies (diploidy) in the genome.

Position effect: Change in expression of a gene without any direct change in the coding or regulatory sequences, but often because of a change in the gene environment.

Overdominance: When the heterozygous genotype is advantageous compared to both homozygous genotypes (in the case a a bi-allelic locus).

Aneuploidy: When the chromosomes are present in an abnormal number, which is often the case when all chromosome pairs do not have the same number of copies.

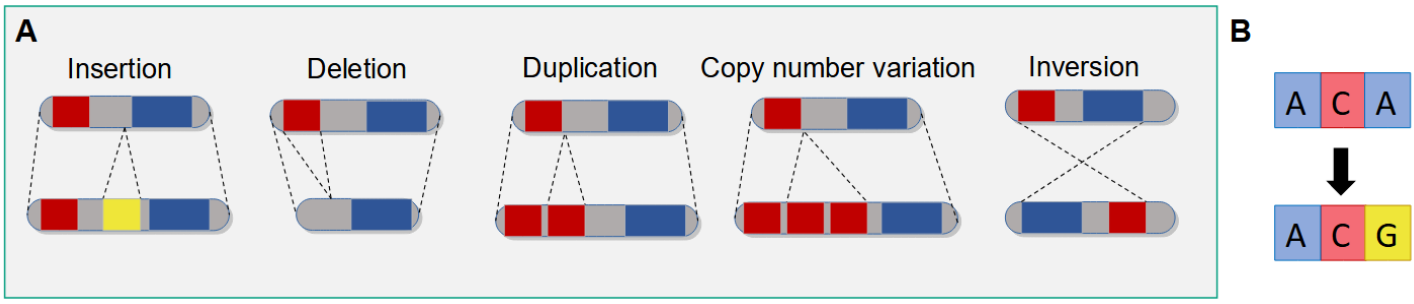

Figure 1: Different types of genomics variants

A: Structural variants. **B:** Single nucleotide Polymorphism

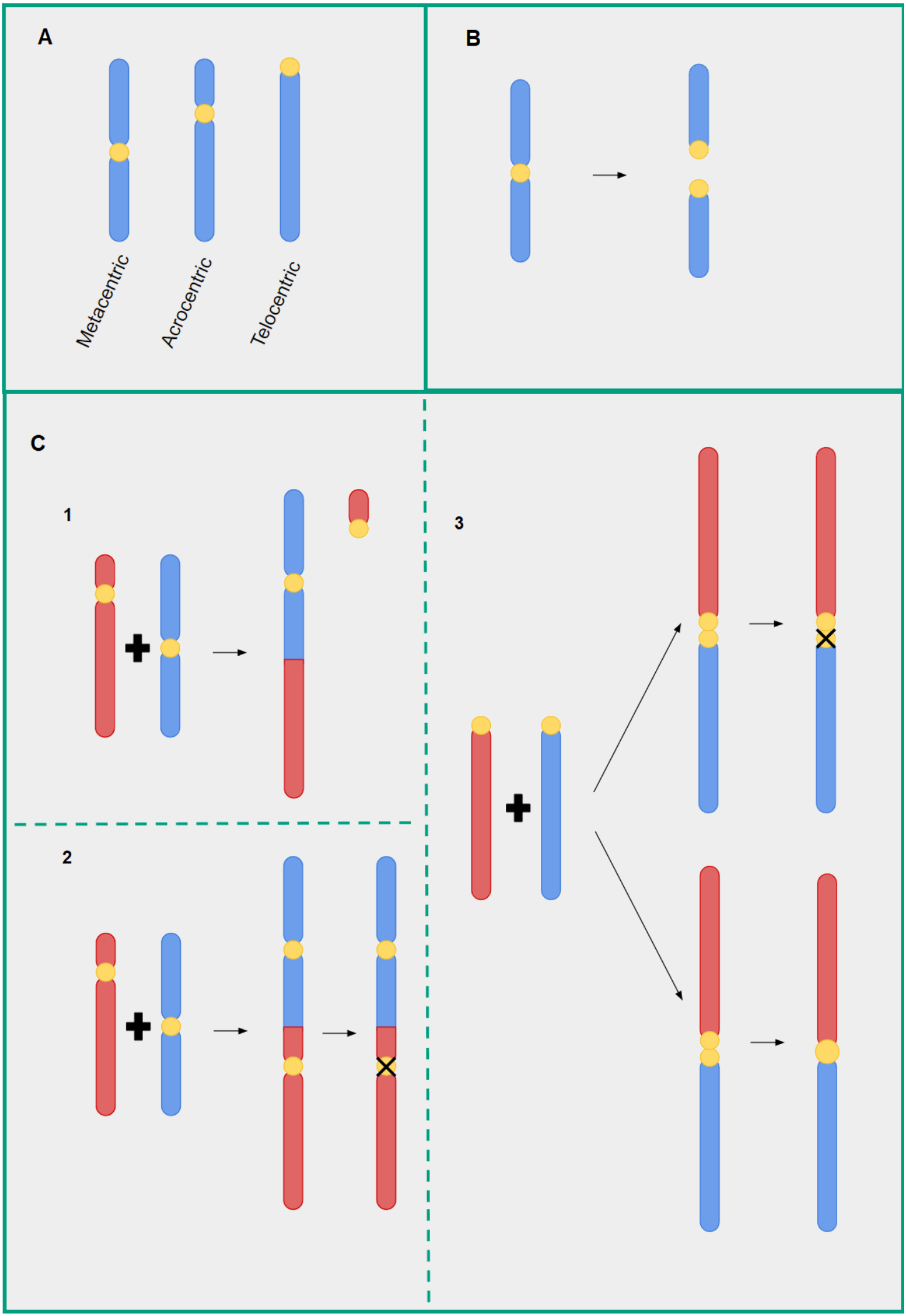
A
Metacentric
Acrocentric
Telocentric
B
C
1
2
3

Figure 2: Chromosome fusion and fission

A: Type of chromosomes depending on the position of the centromere. **B:** Chromosome fission, here a metacentric chromosome is split into two acrocentric chromosomes. **C:** Chromosome fusion. **1:** telomere to telomere fusion with elimination of a centromere bearing arm: Fusion of a metacentric chromosome and a telocentric (or acrocentric) chromosome results in a fused chromosome and a small free chromosome issued from the telocentric (or acrocentric) initial chromosome. **2:** Telomere to telomere fusion with elimination of one centromere: Fusion of a metacentric (or any other type) with a telocentric (or any other type) chromosomes, which leads the two chromosomes to completely merge, resulting in a chromosome with two centromeres, where one of them will be quickly lost. **3:** Centromere to centromere fusion: Fusion of two telocentric (or acrocentric) chromosomes by their centromere. In this case, the two centromeres will be adjacent, which can lead to one of them being lost (top - as in C2), or they can form a single, possibly stronger centromere (bottom).

Table 1: Chronology of the main methods used to identify SVs

Technology name	Year	Description	References
Chromosome banding	1968	Stain the chromosomes allowing their observation with a microscope	[150]Caspersson, T., Farber, S., Foley, G. E., Kudynowski, J., Modest, E. J., Simonsson, E., ... & Zech, L. (1968). Chemical differentiation along metaphase chromosomes. *Experimental cell research*, *49*(1), 219-222.
Fluorescent in situ hybridization (FISH)	1982	Uses fluorescent probes to bind to specific nucleic acids.	[151]Langer-Safer, P. R., Levine, M., & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences*, *79*(14), 4381-4385.
Array Comparative Genomic Hybridization (Array CGH)	1992	Hybridize two different DNAs fragments to the same probes (oligonucleotides), to observe differences in intensities.	[152]Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., & Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, *258*(5083), 818-821.
SNP-array	1998	Similar as array CGH but use allele specific oligonucleotides	[153]Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., ... & Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, *280*(5366), 1077-1082.
Solexa (sequencing by synthesis)	2006	Short read sequencing method relying on the synthesis of complementary strand nucleotide by nucleotide	[154]Pickrell, W. O., Rees, M. I., & Chung, S. K. (2012). Next generation sequencing methodologies-an overview. *Advances in protein chemistry and structural biology*, *89*, 1-26.

Solid (sequencing by ligation)	2008	Short read sequencing method making use of a ligase enzyme	[154]Pickrell, W. O., Rees, M. I., & Chung, S. K. (2012). Next generation sequencing methodologies-an overview. *Advances in protein chemistry and structural biology*, *89*, 1-26.
High throughput chromosome conformation capture (Hi-C)	2009	Capture chromatin conformation to investigate 3D organization of the genome	[155]Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., ... & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, *326*(5950), 289-293.
Pacbio SMRT sequencing	2011	Long reads method using fluorescent nucleotide to sequence the DNA fragments in real time	[154]Pickrell, W. O., Rees, M. I., & Chung, S. K. (2012). Next generation sequencing methodologies-an overview. *Advances in protein chemistry and structural biology*, *89*, 1-26.
Nanopore sequencing	2012	Long reads method where DNA fragments pass through pores and emits differents electric signal depending on the base passing by the pore	[154]Pickrell, W. O., Rees, M. I., & Chung, S. K. (2012). Next generation sequencing methodologies-an overview. *Advances in protein chemistry and structural biology*, *89*, 1-26.

Bibliography:

1. Carroll, S.P., Hendry, A.P., Reznick, D.N., and Fox, C.W. (2007). Evolution on ecological time-scales. Funct. Ecol. *21*, 387–393. 10.1111/j.1365-2435.2007.01289.x.
2. Mitchell-Olds, T., Willis, J.H., and Goldstein, D.B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. *8*, 845–856. 10.1038/nrg2207.
3. Hendry, A.P., Kinnison, M.T., Heino, M., Day, T., Smith, T.B., Fitt, G., Bergstrom, C.T., Oakeshott, J., Jørgensen, P.S., Zalucki, M.P., et al. (2011). Evolutionary principles and their practical application: Evolutionary principles and applications. Evol. Appl. *4*, 159–183. 10.1111/j.1752-4571.2010.00165.x.
4. Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., et al. (2013). Genome-Wide Association Studies Identify Two Novel BMP15 Mutations Responsible for an Atypical Hyperprolificacy Phenotype in Sheep. PLoS Genet. *9*, e1003482. 10.1371/journal.pgen.1003482.
5. Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. *15*, 749–763. 10.1038/nrg3803.
6. Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., and Lohmueller, K.E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc. Natl. Acad. Sci. *113*, 152–157. 10.1073/pnas.1512501113.
7. Sanchis-Juan, A., Stephens, J., French, C.E., Gleadall, N., Mégy, K., Penkett, C., Shamardina, O., Stirrups, K., Delon, I., Dewhurst, E., et al. (2018). Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. Genome Med. *10*, 95. 10.1186/s13073-018-0606-6.
8. Schikora-Tamarit, M.À., and Gabaldón, T. (2022). PerSVade: personalized structural variant detection in any species of interest. Genome Biol. *23*, 175. 10.1186/s13059-022-02737-4.
9. de Bruijn, S.E., Rodenburg, K., Corominas, J., Ben-Yosef, T., Reurink, J., Kremer, H., Whelan, L., Plomp, A.S., Berger, W., Farrar, G.J., et al. (2023). Optical genome mapping and revisiting short-read genome sequencing data reveal previously overlooked structural variants disrupting retinal disease−associated genes. Genet. Med. *25*, 100345. 10.1016/j.gim.2022.11.013.
10. Ho, S.S., Urban, A.E., and Mills, R.E. (2020). Structural variation in the sequencing era. Nat. Rev. Genet. *21*, 171–189. 10.1038/s41576-019-0180-9.
11. Kirubakaran, T.G., Grove, H., Kent, M.P., Sandve, S.R., Baranski, M., Nome, T., De Rosa, M.C., Righino, B., Johansen, T., Otterå, H., et al. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. Mol. Ecol. *25*, 2130–2143. 10.1111/mec.13592.
12. Weissensteiner, M.H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., Peona, V., Pophaly, S.D., Sedlazeck, F.J., Suh, A., et al. (2020). Discovery and population genomics of structural variation in a songbird genus. Nat. Commun. *11*, 3403. 10.1038/s41467-020-17195-4.
13. Matschiner, M., Barth, J.M.I., Tørresen, O.K., Star, B., Baalsrud, H.T., Brieuc, M.S.O., Pampoulie, C., Bradbury, I., Jakobsen, K.S., and Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. Nat. Ecol. Evol. *6*, 469–481. 10.1038/s41559-022-01661-x.
14. Mérot, C., Oomen, R.A., Tigano, A., and Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. Trends Ecol. Evol. *35*, 561–572. 10.1016/j.tree.2020.03.002.
15. Kumar, S., Banks, T.W., and Cloutier, S. (2012). SNP Discovery through Next-Generation Sequencing and Its Applications. Int. J. Plant Genomics *2012*, 1–15. 10.1155/2012/831460.
16. Balagué-Dobón, L., Cáceres, A., and González, J.R. (2022). Fully exploiting SNP arrays: a systematic review on the tools to extract underlying genomic structure. Brief. Bioinform. *23*,

bbac043. 10.1093/bib/bbac043.
17. Thompson, M.J., and Jiggins, C.D. (2014). Supergenes and their role in evolution. Heredity *113*, 1–8. 10.1038/hdy.2014.20.
18. Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., et al. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. Nature *606*, 527–534. 10.1038/s41586-022-04808-9.
19. Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. Genome Biol. *20*, 246. 10.1186/s13059-019-1828-7.
20. Lappalainen, T., Scott, A.J., Brandt, M., and Hall, I.M. (2019). Genomic Analysis in the Age of Human Genome Sequencing. Cell *177*, 70–84. 10.1016/j.cell.2019.02.032.
21. Escaramís, G., Docampo, E., and Rabionet, R. (2015). A decade of structural variants: description, history and methods to detect structural variation. Brief. Funct. Genomics *14*, 305–314. 10.1093/bfgp/elv014.
22. del Gaudio, D., Fang, P., Scaglia, F., Ward, P.A., Craigen, W.J., Glaze, D.G., Neul, J.L., Patel, A., Lee, J.A., Irons, M., et al. (2006). Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. Genet. Med. *8*, 784–792. 10.1097/01.gim.0000250502.28516.3c.
23. Rees, E., Walters, J.T.R., Georgieva, L., Isles, A.R., Chambert, K.D., Richards, A.L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., et al. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. Br. J. Psychiatry *204*, 108–114. 10.1192/bjp.bp.113.131052.
24. Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. PLoS Biol. *8*, e1000501. 10.1371/journal.pbio.1000501.
25. Lam, H.Y.K., Mu, X.J., Stütz, A.M., Tanzer, A., Cayting, P.D., Snyder, M., Kim, P.M., Korbel, J.O., and Gerstein, M.B. (2010). Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. Nat. Biotechnol. *28*, 47–55. 10.1038/nbt.1600.
26. Glasauer, S.M.K., and Neuhauss, S.C.F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics *289*, 1045–1060. 10.1007/s00438-014-0889-2.
27. Robertson, Wm.R.B. (1916). Chromosome studies. I. Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. J. Morphol. *27*, 179–331. 10.1002/jmor.1050270202.
28. Schubert, I., Schriever-Schwemmer, G., Werner, T., and Adler, I.-D. (1992). Telomeric signals in Robertsonian fusion and fission chromosomes: implications for the origin of pseudoaneuploidy. Cytogenet. Genome Res. *59*, 6–9. 10.1159/000133186.
29. Slijepcevic, P. (1998). Telomeres and mechanisms of Robertsonian fusion. Chromosoma *107*, 136–140. 10.1007/s004120050289.
30. de Vos, J.M., Augustijnen, H., Bätscher, L., and Lucek, K. (2020). Speciation through chromosomal fusion and fission in Lepidoptera. Philos. Trans. R. Soc. B Biol. Sci. *375*, 20190539. 10.1098/rstb.2019.0539.
31. Perry, J., Slater, H.R., and Choo, K.H.A. (2004). Centric fission – simple and complex mechanisms. Chromosome Res. *12*, 627–640. 10.1023/B:CHRO.0000036594.38997.59.
32. Sturtevant, A.H. (1921). A Case of Rearrangement of Genes in Drosophila. Proc. Natl. Acad. Sci. *7*, 235–237. 10.1073/pnas.7.8.235.
33. Clarke, C.A., and Sheppard, P.M. (1960). Super-genes and mimicry. Heredity *14*, 175–185. 10.1038/hdy.1960.15.
34. Warburton, D. (1991). De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. Am. J. Hum. Genet. *49*, 995–1013.
35. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97. 10.1038/nrg1767.
36. Weckselblatt, B., and Rudd, M.K. (2015). Human Structural Variation: Mechanisms of Chromosome Rearrangements. Trends Genet. *31*, 587–599. 10.1016/j.tig.2015.05.010.

37. Balachandran, P., Walawalkar, I.A., Flores, J.I., Dayton, J.N., Audano, P.A., and Beck, C.R. (2022). Transposable element-mediated rearrangements are prevalent in human genomes. Nat. Commun. *13*, 7115. 10.1038/s41467-022-34810-8.
38. Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., and Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc. Natl. Acad. Sci. *85*, 9138–9142. 10.1073/pnas.85.23.9138.
39. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951. 10.1038/ng1416.
40. Lejeune, J., Gautier, M., and Turpin, R. (1959). [Study of somatic chromosomes from 9 mongoloid children]. Comptes Rendus Hebd. Seances Acad. Sci. *248*, 1721–1722.
41. Mégarbané, A., Ravel, A., Mircher, C., Sturtz, F., Grattau, Y., Rethoré, M.-O., Delabar, J.-M., and Mobley, W.C. (2009). The 50th anniversary of the discovery of trisomy 21: The past, present, and future of research and treatment of Down syndrome. Genet. Med. *11*, 611–616. 10.1097/GIM.0b013e3181b2e34c.
42. Ohno, S. (1970). Evolution by Gene Duplication (Springer Berlin Heidelberg) 10.1007/978-3-642-86659-3.
43. Hokamp, K., McLysaght, A., and Wolfe, K.H. (2003). The 2R hypothesis and the human genome sequence. J. Struct. Funct. Genomics *3*, 95–110.
44. Lundin, L.G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics *16*, 1–19. 10.1006/geno.1993.1133.
45. Garcia-Fernández, J., and Holland, P.W. (1994). Archetypal organization of the amphioxus Hox gene cluster. Nature *370*, 563–566. 10.1038/370563a0.
46. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-Scale Copy Number Polymorphism in the Human Genome. Science *305*, 525–528. 10.1126/science.1098918.
47. Graubert, T.A., Cahan, P., Edwin, D., Selzer, R.R., Richmond, T.A., Eis, P.S., Shannon, W.D., Li, X., McLeod, H.L., Cheverud, J.M., et al. (2007). A High-Resolution Map of Segmental DNA Copy Number Variation in the Mouse Genome. PLoS Genet. *3*.
48. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Meur, N.L., Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., et al. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat. Genet. *38*, 24–26. 10.1038/ng1718.
49. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. Science *302*, 841. 10.1126/science.1090278.
50. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al. (2011). Mapping copy number variation by population-scale genome sequencing. Nature *470*, 59–65. 10.1038/nature09708.
51. Morgan, A.P., Gatti, D.M., Najarian, M.L., Keane, T.M., Galante, R.J., Pack, A.I., Mott, R., Churchill, G.A., and de Villena, F.P.-M. (2017). Structural Variation Shapes the Landscape of Recombination in Mouse. Genetics *206*, 603–619. 10.1534/genetics.116.197988.
52. Hager, E.R., Harringmeyer, O.S., Wooldridge, T.B., Theingi, S., Gable, J.T., McFadden, S., Neugeboren, B., Turner, K.M., Jensen, J.D., and Hoekstra, H.E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. Science *377*, 399–405. 10.1126/science.abg0718.
53. Rubin, C.-J., Megens, H.-J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., et al. (2012). Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. *109*, 19529–19536. 10.1073/pnas.1217149109.
54. Low, W.Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D.M., Rosen, B.D., Kronenberg, Z.N., Kingan, S.B., Tseng, E., et al. (2020). Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. Nat. Commun. *11*, 2071. 10.1038/s41467-020-15848-y.

55. Beyter, D., Ingimundardottir, H., Oddsson, A., Eggertsson, H.P., Bjornsson, E., Jonsson, H., Atlason, B.A., Kristmundsdottir, S., Mehringer, S., Hardarson, M.T., et al. (2021). Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. Nat. Genet. *53*, 779–786. 10.1038/s41588-021-00865-4.
56. Marx, V. (2023). Method of the year: long-read sequencing. Nat. Methods *20*, 6–11. 10.1038/s41592-022-01730-w.
57. Cameron, D.L., Di Stefano, L., and Papenfuss, A.T. (2019). Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat. Commun. *10*, 3240. 10.1038/s41467-019-11146-4.
58. Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y. (2019). Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. *20*, 117. 10.1186/s13059-019-1720-5.
59. Kuzniar, A. (2020). sv-callers: a highly portable parallel workflow for structural variant detection in whole-genome sequence data. 12.
60. Liu, Y., Zhang, M., Sun, J., Chang, W., Sun, M., Zhang, S., and Wu, J. (2020). Comparison of multiple algorithms to reliably detect structural variants in pears. BMC Genomics *21*, 61. 10.1186/s12864-020-6455-x.
61. Wang, S., Lee, S., Chu, C., Jain, D., Kerpedjiev, P., Nelson, G.M., Walsh, J.M., Alver, B.H., and Park, P.J. (2020). HiNT: a computational method for detecting copy number variations and translocations from Hi-C data. Genome Biol. *21*, 73. 10.1186/s13059-020-01986-5.
62. Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467. 10.1038/s41576-018-0007-0.
63. Li, Y., Wang, S., Zhang, Z., Luo, J., Lin, G.L., Deng, W.-D., Guo, Z., Han, F.M., Wang, L.-L., Li, J., et al. (2023). Large-Scale Chromosomal Changes Lead to Genome-Level Expression Alterations, Environmental Adaptation, and Speciation in the Gayal (*Bos frontalis*). Mol. Biol. Evol. *40*, msad006. 10.1093/molbev/msad006.
64. Berdan, E.L., Blanckaert, A., Butlin, R.K., and Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. PLOS Genet. *17*, e1009411. 10.1371/journal.pgen.1009411.
65. Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Mol. Biol. Evol. *36*, 632–637. 10.1093/molbev/msy228.
66. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature *444*, 444–454. 10.1038/nature05329.
67. Carvalho, C.M.B., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. *17*, 224–238. 10.1038/nrg.2015.25.
68. Leffler, E.M., Band, G., Busby, G.B.J., Kivinen, K., Le, Q.S., Clarke, G.M., Bojang, K.A., Conway, D.J., Jallow, M., Sisay-Joof, F., et al. (2017). Resistance to malaria through structural variation of red blood cell invasion receptors. Science *356*, eaam6393. 10.1126/science.aam6393.
69. Lin, Y.-L., and Gokcumen, O. (2019). Fine-Scale Characterization of Genomic Structural Variation in the Human Genome Reveals Adaptive and Biomedically Relevant Hotspots. Genome Biol. Evol. *11*, 1136–1151. 10.1093/gbe/evz058.
70. Louzada, S., Algady, W., Weyell, E., Zuccherato, L.W., Brajer, P., Almalki, F., Scliar, M.O., Naslavsky, M.S., Yamamoto, G.L., Duarte, Y.A.O., et al. (2020). Structural variation of the malaria-associated human glycophorin A-B-E region. BMC Genomics *21*, 446. 10.1186/s12864-020-06849-8.
71. Algady, W., Weyell, E., Mateja, D., Garcia, A., Courtin, D., and Hollox, E.J. (2021). Genotyping complex structural variation at the malaria-associated human glycophorin locus using a PCR-based strategy. Ann. Hum. Genet. *85*, 7–17. 10.1111/ahg.12405.
72. Levinson, G., and Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. *4*, 203–221. 10.1093/oxfordjournals.molbev.a040442.
73. Kidd, J.M., Graves, T., Newman, T.L., Fulton, R., Hayden, H.S., Malig, M., Kallicki, J., Kaul, R., Wilson, R.K., and Eichler, E.E. (2010). A Human Genome Structural Variation Sequencing

Resource Reveals Insights into Mutational Mechanisms. Cell *143*, 837–847. 10.1016/j.cell.2010.10.027.
74. Brashear, W.A., Raudsepp, T., and Murphy, W.J. (2018). Evolutionary conservation of Y Chromosome ampliconic gene families despite extensive structural variation. Genome Res. *28*, 1841–1851. 10.1101/gr.237586.118.
75. Kumar, P., Jain, M., Kalsi, A.K., and Halder, A. (2018). Molecular characterisation of a case of dicentric Y presented as nonobstructive azoospermia with testicular early maturation arrest. Andrologia *50*, e12886. 10.1111/and.12886.
76. Murnane, J.P. (2012). Telomere dysfunction and chromosome instability. Mutat. Res. Mol. Mech. Mutagen. *730*, 28–36. 10.1016/j.mrfmmm.2011.04.008.
77. Ohno, Y., Ogiyama, Y., Kubota, Y., Kubo, T., and Ishii, K. (2016). Acentric chromosome ends are prone to fusion with functional chromosome ends through a homology-directed rearrangement. Nucleic Acids Res. *44*, 232–244. 10.1093/nar/gkv997.
78. Lysak, M.A. (2022). Celebrating Mendel, McClintock, and Darlington: On end-to-end chromosome fusions and nested chromosome fusions. Plant Cell *34*, 2475–2491. 10.1093/plcell/koac116.
79. Schubert, I., and Lysak, M.A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. Trends Genet. *27*, 207–216. 10.1016/j.tig.2011.03.004.
80. Vasil'ev, V.P. (2009). Mechanisms of Polyploid Evolution in Fish: Polyploidy in Sturgeons. In Biology, Conservation and Sustainable Development of Sturgeons, R. Carmona, A. Domezain, M. García-Gallego, J. A. Hernando, F. Rodríguez, and M. Ruiz-Rejón, eds. (Springer Netherlands), pp. 97–117. 10.1007/978-1-4020-8437-9_6.
81. Dyban, A.P., and Baranov, V.S. (1987). Cytogenetics of mammalian embryonic development (Clarendon Press ; Oxford University Press).
82. Sundberg, L.-R., and Pulkkinen, K. (2015). Genome size evolution in macroparasites. Int. J. Parasitol. *45*, 285–288. 10.1016/j.ijpara.2014.12.007.
83. Andersson, S.G.E., and Kurland, C.G. (1998). Reductive evolution of resident genomes. Trends Microbiol. *6*, 263–268. 10.1016/S0966-842X(98)01312-2.
84. Yin, D., Schwarz, E.M., Thomas, C.G., Felde, R.L., Korf, I.F., Cutter, A.D., Schartner, C.M., Ralston, E.J., Meyer, B.J., and Haag, E.S. (2018). Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science *359*, 55–61. 10.1126/science.aao0827.
85. Moran, N.A., and Mira, A. (2011). The process of genome shrinkage in the obligate symbiont.
86. Wright, N.A., Gregory, T.R., and Witt, C.C. (2014). Metabolic 'engines' of flight drive genome size reduction in birds. Proc. R. Soc. B Biol. Sci. *281*, 20132780. 10.1098/rspb.2013.2780.
87. Poulin, R., and Randhawa, H.S. (2015). Evolution of parasitism along convergent lines: from ecology to genomics. Parasitology *142*, S6–S15. 10.1017/S0031182013001674.
88. Tankard, R.M., Bennett, M.F., Degorski, P., Delatycki, M.B., Lockhart, P.J., and Bahlo, M. (2018). Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data. Am. J. Hum. Genet. *103*, 858–873. 10.1016/j.ajhg.2018.10.015.
89. Abel, H.J., Larson, D.E., Regier, A.A., Chiang, C., Das, I., Kanchi, K.L., Layer, R.M., Neale, B.M., Salerno, W.J., Reeves, C., et al. (2020). Mapping and characterization of structural variation in 17,795 human genomes. 46.
90. Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., and Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. Nat. Genet. *53*, 288–293. 10.1038/s41588-020-00771-1.
91. Chouteau, M., Llaurens, V., Piron-Prunier, F., and Joron, M. (2017). Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. Proc. Natl. Acad. Sci. *114*, 8325–8329. 10.1073/pnas.1702482114.
92. Faria, R., Johannesson, K., Butlin, R.K., and Westram, A.M. (2019). Evolving Inversions. Trends Ecol. Evol. *34*, 239–248. 10.1016/j.tree.2018.12.005.
93. Jay, P., Tezenas, E., Véber, A., and Giraud, T. (2022). Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLOS Biol. *20*, e3001698. 10.1371/journal.pbio.3001698.
94. Charlesworth, B., Betancourt, A.J., Kaiser, V.B., and Gordo, I. (2009). Genetic Recombination and Molecular Evolution. Cold Spring Harb. Symp. Quant. Biol. *74*, 177–186.

10.1101/sqb.2009.74.015.
95. Campos, J.L., Charlesworth, B., and Haddrill, P.R. (2012). Molecular Evolution in Nonrecombining Regions of the Drosophila melanogaster Genome. Genome Biol. Evol. *4*, 278–288. 10.1093/gbe/evs010.
96. Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. *24*, 94–102. 10.1016/j.tree.2008.09.010.
97. Yan, Z., Martin, S.H., Gotzek, D., Arsenault, S.V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B.G., Salamin, N., Shoemaker, D., et al. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. Nat. Ecol. Evol. *4*, 240–249. 10.1038/s41559-019-1081-1.
98. Edger, P.P., and Pires, J.C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res. *17*, 699–717. 10.1007/s10577-009-9055-9.
99. Assis, R., and Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in *Drosophila*. Proc. Natl. Acad. Sci. *110*, 17409–17414. 10.1073/pnas.1313759110.
100. Lynch, M., and Conery, J.S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. Science *290*, 1151–1155. 10.1126/science.290.5494.1151.
101. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531–1545. 10.1093/genetics/151.4.1531.
102. Sandve, S.R., Rohlfs, R.V., and Hvidsten, T.R. (2018). Subfunctionalization versus neofunctionalization after whole-genome duplication. Nat. Genet. *50*, 908–909. 10.1038/s41588-018-0162-4.
103. Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., et al. (2007). Diet and the evolution of human amylase gene copy number variation. Nat. Genet. *39*, 1256–1260. 10.1038/ng2123.
104. Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R.-A., Garneau, D., Daugherity, E., Globig, A., Ruhl, S., and Gokcumen, O. (2019). Independent amylase gene copy number bursts correlate with dietary preferences in mammals. eLife *8*, e44628. 10.7554/eLife.44628.
105. Helsen, J., Voordeckers, K., Vanderwaeren, L., Santermans, T., Tsontaki, M., Verstrepen, K.J., and Jelier, R. (2020). Gene Loss Predictably Drives Evolutionary Adaptation. Mol. Biol. Evol. *37*, 2989–3002. 10.1093/molbev/msaa172.
106. Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. Nat. Rev. Genet. *17*, 379–391. 10.1038/nrg.2016.39.
107. Protas, M.E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W.R., Zon, L.I., Borowsky, R., and Tabin, C.J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat. Genet. *38*, 107–111. 10.1038/ng1700.
108. Protas, M.E., Trontelj, P., and Patel, N.H. (2011). Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. Proc. Natl. Acad. Sci. *108*, 5702–5707. 10.1073/pnas.1013850108.
109. Rétaux, S., and Casane, D. (2013). Evolution of eye development in the darkness of caves: adaptation, drift, or both? EvoDevo *4*, 26. 10.1186/2041-9139-4-26.
110. Emerling, C.A., and Springer, M.S. (2014). Eyes underground: Regression of visual protein networks in subterranean mammals. Mol. Phylogenet. Evol. *78*, 260–270. 10.1016/j.ympev.2014.05.016.
111. Makino, T., and McLysaght, A. (2012). Positionally biased gene loss after whole genome duplication: Evidence from human, yeast, and plant. Genome Res. *22*, 2427–2435. 10.1101/gr.131953.111.
112. Rouzic, A.L., and Deceliere, G. (2005). Models of the population genetics of transposable elements. Genet. Res. *85*, 171–181. 10.1017/S0016672305007585.
113. Le Rouzic, A., Boutin, T.S., and Capy, P. (2007). Long-term evolution of transposable elements. Proc. Natl. Acad. Sci. *104*, 19375–19380. 10.1073/pnas.0705238104.
114. Volff, J.-N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays *28*, 913–922. 10.1002/bies.20452.
115. Jangam, D., Feschotte, C., and Betrán, E. (2017). Transposable Element Domestication As an

Adaptation to Evolutionary Conflicts. Trends Genet. *33*, 817–831. 10.1016/j.tig.2017.07.011.
116. Joly-Lopez, Z., and Bureau, T.E. (2018). Exaptation of transposable element coding sequences. Curr. Opin. Genet. Dev. *49*, 34–42. 10.1016/j.gde.2018.02.011.
117. Abad, J.P., de Pablos, B., Osoegawa, K., de Jong, P.J., Martín-Gallardo, A., and Villasante, A. (2004). TAHRE, a Novel Telomeric Retrotransposon from Drosophila melanogaster, Reveals the Origin of Drosophila Telomeres. Mol. Biol. Evol. *21*, 1620–1624. 10.1093/molbev/msh180.
118. Kapitonov, V.V., and Jurka, J. (2005). RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. PLoS Biol. *3*, e181. 10.1371/journal.pbio.0030181.
119. Shpiz, S., Kwon, D., Uneva, A., Kim, M., Klenov, M., Rozovsky, Y., Georgiev, P., Savitsky, M., and Kalmykova, A. (2007). Characterization of Drosophila Telomeric Retroelement TAHRE: Transcription, Transpositions, and RNAi-based Regulation of Expression. Mol. Biol. Evol. *24*, 2535–2545. 10.1093/molbev/msm205.
120. Choudhury, R.R., Rogivue, A., Gugerli, F., and Parisod, C. (2019). Impact of polymorphic transposable elements on linkage disequilibrium along chromosomes. Mol. Ecol. *28*, 1550–1562. 10.1111/mec.15014.
121. Catlin, N.S., and Josephs, E.B. (2022). The important contribution of transposable elements to phenotypic variation and evolution. Curr. Opin. Plant Biol. *65*, 102140. 10.1016/j.pbi.2021.102140.
122. Lertzman-Lepofsky, G., Mooers, A.Ø., and Greenberg, D.A. (2019). Ecological constraints associated with genome size across salamander lineages. Proc. R. Soc. B Biol. Sci. *286*, 20191780. 10.1098/rspb.2019.1780.
123. Lamichhaney, S., Catullo, R., Keogh, J.S., Clulow, S., Edwards, S.V., and Ezaz, T. (2021). A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. Proc. Natl. Acad. Sci. *118*, e2011649118. 10.1073/pnas.2011649118.
124. Yuan, J., Zhang, X., Wang, M., Sun, Y., Liu, C., Li, S., Yu, Y., Gao, Y., Liu, F., Zhang, X., et al. (2021). Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp. Commun. Biol. *4*, 186. 10.1038/s42003-021-01716-y.
125. Sokolovskis, K., Lundberg, M., Åkesson, S., Willemoes, M., Zhao, T., Caballero-Lopez, V., and Bensch, S. (2023). Migration direction in a songbird explained by two loci. Nat. Commun. *14*, 165. 10.1038/s41467-023-35788-7.
126. Dumas, D., and 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*, 1355–1366. 10.1093/genetics/162.3.1355.
127. 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., et al. (2021). The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. Nat. Commun. *12*, 2981. 10.1038/s41467-021-23270-1.
128. Guerrero, R.F., and Kirkpatrick, M. (2014). LOCAL ADAPTATION AND THE EVOLUTION OF CHROMOSOME FUSIONS: CHROMOSOME FUSIONS AND LOCAL ADAPTATION. Evolution *68*, 2747–2756. 10.1111/evo.12481.
129. Hirata, S., Hirai, H., Nogami, E., Morimura, N., and Udono, T. (2017). Chimpanzee Down syndrome: a case study of trisomy 22 in a captive chimpanzee. Primates *58*, 267–273. 10.1007/s10329-017-0597-8.
130. Novikova, P.Yu., Brennan, I.G., Booker, W., Mahony, M., Doughty, P., Lemmon, A.R., Moriarty Lemmon, E., Roberts, J.D., Yant, L., Van De Peer, Y., et al. (2020). Polyploidy breaks speciation barriers in Australian burrowing frogs Neobatrachus. PLOS Genet. *16*, e1008769. 10.1371/journal.pgen.1008769.
131. Mable, B.K., Alexandrou, M.A., and Taylor, M.I. (2011). Genome duplication in amphibians and fish: an extended synthesis. J. Zool. *284*, 151–182. 10.1111/j.1469-7998.2011.00829.x.
132. Doyle, J.J., and Coate, J.E. (2019). Polyploidy, the Nucleotype, and Novelty: The Impact of Genome Doubling on the Biology of the Cell. Int. J. Plant Sci. *180*, 1–52. 10.1086/700636.
133. Sémon, M., and Wolfe, K.H. (2007). Consequences of genome duplication. Curr. Opin. Genet. Dev. *17*, 505–512. 10.1016/j.gde.2007.09.007.

134. Robertson, F.M., Gundappa, M.K., Grammes, F., Hvidsten, T.R., Redmond, A.K., Lien, S., Martin, S.A.M., Holland, P.W.H., Sandve, S.R., and Macqueen, D.J. (2017). Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. Genome Biol. *18*, 111. 10.1186/s13059-017-1241-z.
135. Parey, E., Louis, A., Montfort, J., Guiguen, Y., Crollius, H.R., and Berthelot, C. (2022). An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication. Genome Res. *32*, 1685–1697. 10.1101/gr.276953.122.
136. Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., Stager, M., Betuel, A., Cheviron, Z.A., Warren, W.C., et al. (2016). Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Curr. Biol. *26*, 344–350. 10.1016/j.cub.2015.11.069.
137. Graves, J.A.M. (2006). Sex Chromosome Specialization and Degeneration in Mammals. Cell *124*, 901–914. 10.1016/j.cell.2006.02.024.
138. Peichel, C.L., McCann, S.R., Ross, J.A., Naftaly, A.F.S., Urton, J.R., Cech, J.N., Grimwood, J., Schmutz, J., Myers, R.M., Kingsley, D.M., et al. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol. *21*, 177. 10.1186/s13059-020-02097-x.
139. Charlesworth, D. (2021). The timing of genetic degeneration of sex chromosomes. Philos. Trans. R. Soc. B Biol. Sci. *376*, 20200093. 10.1098/rstb.2020.0093.
140. Camacho, J.P.M., Sharbel, T.F., and Beukeboom, L.W. (2000). B-chromosome evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *355*, 163–178. 10.1098/rstb.2000.0556.
141. Ahmad, S.F., Jehangir, M., Cardoso, A.L., Wolf, I.R., Margarido, V.P., Cabral-de-Mello, D.C., O'Neill, R., Valente, G.T., and Martins, C. (2020). B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. BMC Genomics *21*, 656. 10.1186/s12864-020-07072-1.
142. Silva, D.M.Z.D.A., Ruiz-Ruano, F.J., Utsunomia, R., Martín-Peciña, M., Castro, J.P., Freire, P.P., Carvalho, R.F., Hashimoto, D.T., Suh, A., Oliveira, C., et al. (2021). Long-term persistence of supernumerary B chromosomes in multiple species of Astyanax fish. BMC Biol. *19*, 52. 10.1186/s12915-021-00991-9.
143. Borodin, P., Chen, A., Forstmeier, W., Fouché, S., Malinovskaya, L., Pei, Y., Reifová, R., Ruiz-Ruano, F.J., Schlebusch, S.A., Sotelo-Muñoz, M., et al. (2022). Mendelian nightmares: the germline-restricted chromosome of songbirds. Chromosome Res. *30*, 255–272. 10.1007/s10577-022-09688-3.
144. Camus, M.F., Alexander-Lawrie, B., Sharbrough, J., and Hurst, G.D.D. (2022). Inheritance through the cytoplasm. Heredity *129*, 31–43. 10.1038/s41437-022-00540-2.
145. Hurst, G.D.D., and Werren, J.H. (2001). The role of selfish genetic elements in eukaryotic evolution. Nat. Rev. Genet. *2*, 597–606. 10.1038/35084545.
146. Guiglielmoni, N., Houtain, A., Derzelle, A., Van Doninck, K., and Flot, J.-F. (2021). Overcoming uncollapsed haplotypes in long-read assemblies of non-model organisms. BMC Bioinformatics *22*, 303. 10.1186/s12859-021-04118-3.
147. Bachtrog, D., Mahajan, S., and Bracewell, R. (2019). Massive gene amplification on a recently formed Drosophila Y chromosome. Nat. Ecol. Evol. *3*, 1587–1597. 10.1038/s41559-019-1009-9.
148. Connallon, T., Olito, C., Dutoit, L., Papoli, H., Ruzicka, F., and Yong, L. (2018). Local adaptation and the evolution of inversions on sex chromosomes and autosomes. Philos. Trans. R. Soc. B Biol. Sci. *373*, 20170423. 10.1098/rstb.2017.0423.
149. Schöpflin, R., Melo, U.S., Moeinzadeh, H., Heller, D., Laupert, V., Hertzberg, J., Holtgrewe, M., Alavi, N., Klever, M.-K., Jungnitsch, J., et al. (2022). Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes. Nat. Commun. *13*, 6470. 10.1038/s41467-022-34053-7.
150. Caspersson, T., Farber, S., Foley, G.E., Kudynowski, J., Modest, E.J., Simonsson, E., Wagh, U., and Zech, L. (1968). Chemical differentiation along metaphase chromosomes. Exp. Cell Res. *49*, 219–222. 10.1016/0014-4827(68)90538-7.
151. Langer-Safer, P.R., Levine, M., and Ward, D.C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. Proc. Natl. Acad. Sci. *79*, 4381–4385. 10.1073/pnas.79.14.4381.

152. Kallioniemi, A., Kallioniemi, O.-P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., and Pinkel, D. (1992). Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors. Science *258*, 818–821. 10.1126/science.1359641.
153. Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science *280*, 1077–1082. 10.1126/science.280.5366.1077.
154. Pickrell, W.O., Rees, M.I., and Chung, S.-K. (2012). Next generation sequencing methodologies--an overview. Adv. Protein Chem. Struct. Biol. *89*, 1–26. 10.1016/B978-0-12-394287-6.00001-X.
155. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293. 10.1126/science.1181369.