How structural variants shape avian phenotypes: lessons from model systems

María Recuerda [1], Leonardo Campagna [1,2]

[1] Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA
[2] Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road, Ithaca, NY 14853, USA

*Correspondence: mariarecuerdacarrasco@gmail.com (M. Recuerda)

Keywords

Avian model systems, chromosomal rearrangements, genotype/phenotype associations, pangenomes, Structural variants.

Abstract

Despite receiving significant recent attention, the relevance of Structural Variation (SV) in driving phenotypic diversity remains understudied. Advances in long-read sequencing, bioinformatics and pangenomic approaches, enhance SV detection. We review the role of SVs in shaping phenotypes in avian model systems, and identify general patterns in SV type, length, and their associated traits. Notably, most of the identified SVs are short indels in chickens, frequently associated with changes in body weight and plumage coloration. This review highlights how SVs underlie phenotypes in avian model systems and sets expectations for when long-read technologies become commonly implemented in non-model birds. The growing interest in this subject suggests an increase in our understanding of the phenotypic effects of SVs in upcoming years.

Structural Variants shape phenotypes in avian model systems

Avian model species vary in plumage color and patterns, beak morphology, vocalizations and behaviors, as well as in economically relevant traits such as body size, immune response and egg production. Therefore, these systems can shed light on the underlying genomic mechanisms shaping such traits, both in model and non-model birds. Most research on the genetic basis of phenotypic traits has focused on single-nucleotide polymorphisms (SNPs, see Glossary) and smaller genetic rearrangements such as short insertion/deletion (indel) mutations. The impact of structural variants (SVs) on avian phenotypes, even in model systems, remains largely understudied despite their potential importance. SVs (Box 1), including insertions, deletions, inversions, and duplications, typically defined as longer than 50bp [1], can affect gene structure and function [2]. This knowledge gap may be due to methodological challenges in detecting SVs, the complex genetic basis of some traits and the lack of highly contiguous reference genomes. SVs are hard to detect and characterize, requiring third-generation sequencing techniques (i.e., long-read technologies such as Pacific Biosciences and Oxford nanopore sequencing), chromosome conformation capture techniques like Hi-C, and the implementation of robust analytical tools [3]. The more widely used short-read technologies are unable to cover many repetitive regions, leading to challenges in genome assembly and hindering SV identification. Additionally, when mapping population-level data against a reference genome, SVs might be overlooked if they are absent in the reference sequence. Moreover, SVs can interact with multiple genes to shape complex and polygenic traits, further complicating the ability to pinpoint the individual effects of each gene on a given phenotype. These challenges must be addressed to understand how SVs shape avian phenotypic evolution.

In this review, we explore how SVs shape different phenotypes in avian model systems, and compare their effect in relation to what is known for other types of genetic variation, such as SNPs. With the accessibility of third-generation sequencing, telomere-to-telomere genomes, and advances in bioinformatics, we expect a significant increase in studies uncovering the influence of SVs on avian phenotypes in the coming years, including in non-model species. We focus on avian model systems for various reasons. First, their commercial value attracts many

research resources, setting them apart from other avian species. Second, the longer-standing availability and superior quality of reference genomes, along with the availability of pedigrees, animal husbandry and genetic mapping techniques, facilitates precise phenotype/genotype associations. In consequence, the earlier studies utilizing long-read sequencing have been conducted on these species. Based on these criteria, we consider as avian model systems the chicken (*Gallus gallus*), Zebra finch (*Taeniopygia guttata*), Wild turkey (*Meleagris gallopavo*), Domestic Mallard duck (*Anas platyrhynchos*), Domestic pigeon (*Columba livia*) and the Common and Japanese quails (*Coturnix coturnix* and *Coturnix japonica*). The outline of the review is organized around four major issues: type and length of SVs related with phenotypic traits in avian model systems, the main traits associated with SVs, the current pangenome availability and relevance, and the relationship between Transposable Elements (TEs) and SVs and their impact on phenotypes.

Systematic analysis of SVs in avian model systems

We conducted a systematic search in Web of Science and Google Scholar using the terms shown in Table S1, which yielded 2,005 studies. From these, we identified 103 articles reporting SVs in avian model systems associated with phenotypic traits. We categorized the studies based on SV type, including small Indels, larger insertions, deletions, duplications, copy number variants (CNV) which encompasses both deletions and duplications, inversions, or complex rearrangements. SV lengths were subsequently classified into the following intervals: <50bp, 50bp to 1Kb, 1-10Kb, 10-100Kb, >100Kb, or Unknown. When a study reported multiple SVs, we used the mean length for our analysis, if it was provided. Similarly, if a study reported multiple SVs associated with different phenotypes, we treated them as distinct entities for analysis. Moreover, when a study documented multiple SVs of the same type that resulted in the same phenotype, we counted one occurrence for our analysis. Additionally, only the initial study from several that described the same SV was counted. Ten out of the eleven studies that focused on characterizing SVs among breeds or populations were excluded from the analysis, as they reported numerous SVs related to broad traits such as domestication or multiple inter-breed

differences. The retained study [4] met the criteria for inclusion. After this initial revision, 88 studies remained in our analysis (see a full list in Table S1).

Type and length of SVs associated with phenotypic traits in avian model systems

All types of SVs are implicated in shaping phenotype, ranging from less than 50 bp all the way to megabases. However, the most commonly detected SVs are indels and duplications, and they tend to be short (<1Kb) (Fig. 1A, B). We note that all the reviewed articles (except one, [5]) relied primarily on short-read sequencing, which introduces a bias towards short SV detection due to the challenges in identifying long SVs with short reads. As the use of long-read technologies becomes more prevalent, long SV identification may increase due to more accurate detection.

In total, 95 SVs were identified among the 88 articles that associated SVs with phenotypic changes. Most of the detected SVs (31.6%) were shorter than 50bp, 46.4% ranged between 50-bp and 100Kb, and only 13.7% of SVs were over 100Kb (Fig. 1A). Considering SVs exclusively as variants over 50bp [1] reduces the number of studies included in our analyses by over 30%. However, several studies show that variants shorter than 50bp can influence phenotype. For instance, short SVs underlie plumage coloration in Japanese quail [6,7] and chickens [8], as well as impacting egg production in both species [9-11]. As proposed by Mérot et al. [2], we agree that the SV concept should encompass the full size-range, from single nucleotide SVs to megabases, without an arbitrary size threshold.

The types of SVs reported in each model species is limited by the availability of studies. In the case of the Zebra finch, only inversions have been reported. Similarly, for the turkey, only deletions and duplications have been documented. Conversely, species such as the Japanese quail and, particularly, the chicken, which were the subjects of a higher percentage of studies included in the review (9.2% and 74.7%, respectively), show higher SV diversity (Fig. 1C, D). Although we included both Common and Japanese quails in our analyses, eight out of nine studies focused on the Japanese quail. Considering the relationship between SV length and

type, insertions are typically shorter, while deletions and duplications show the highest length variability (ranging from a few bases to over 100Kb). Inversions and complex SVs are longer, always exceeding 10Kb (Fig. 1D), yet there are relatively few examples of these SVs. This pattern is most likely a product of detectability and reduced discovery, rather than indicating that inversions and complex rearrangements are uncommon SVs, which is consistent with the limitations of detecting long SVs using the prevailing short-read sequencing methods.

Phenotypic traits associated to SVs

SVs underlie traits primarily related to body size and weight [12-42] followed by plumage coloration and pigmentation [4,6-8,16,38,43-61] in avian model systems. There are examples in quails, where both traits are affected by the same SV [62,63]. There are many studies, mostly in chickens, on feathering phenotypes (45,64-72); comb, muff and beard traits [73-81], and egg production [9-11,82,83]. Although less common, there are also associations between SVs and behavior and domestication [4,5,64,84-90]. Other uncommon traits associated with SVs are craniofacial deformities [91,92], fertility [93], muscle glycogen content [94], number of vertebrae [95] and Aldehyde flavor [96]. The most detected traits are usually economically relevant, such as body size and egg production, and/or conspicuous like plumage coloration. This pattern could be due to a detection bias leaving traits, which are harder to study, like immune responses, underrepresented.

The same phenotype in different species can be achieved by modifying the same gene in various ways. For instance, the late feathering trait, that is a sex-linked phenotype used for sexing individual birds at an early age, in both chickens [72] and turkeys [65], involves SVs in the Prolactin receptor gene (PRLR). In chickens, the SV is a partial duplication of the PRLR and SPEF2 genes that affects gene expression through dosage effect, while in turkeys, a 5bp deletion in the PRLR terminal exon results in a truncated protein lacking 98 C-terminal amino acids (Fig 2A). Moreover, deletions in the Prolactin gene (PRL) are independently involved in egg production both in chickens and Japanese quails [9,82] (Fig 2B). Similarly, larger body size in commercial chicken [37] and Mallard duck breeds [38] has been associated with an SV in the promoter

region of the IGF2BP1 gene that results in increased expression. In chickens, the SV is a deletion, whereas in ducks, it involves a Gypsy long terminal repeat (LTR) TE insertion which leads to higher body mass (Fig 2C). More complex traits such as body size and growth are commonly linked to a wide variety of genes and SVs [e.g., 13,14,18,35,97]

The same phenotype can also be obtained through different types and lengths of SVs in different genes. For instance, the white phenotype in chickens [8] and domestic ducks [38] is attributed to a 4bp deletion in the RAI14 gene and a 6kb insertion in the MITF gene, respectively. In white chickens, the deletion is accompanied by variation in 3 SNPs, one of them affecting the TYR gene. In Pekin and Cherry Valley ducks, a Gypsy TE insertion generates a novel MITF transcript that lacks 39 amino acids, which in turn affects the expression of four downstream genes including the TYR gene, resulting in white plumage (Fig. 2D).

Moreover, the same trait can be modified by either the same or different genes. For instance, in Japanese quails, Fawn-2-beige and yellow plumage coloration arise from a tandem duplication and a deletion in the ASIP locus, respectively [52] (Fig. 1E). Notably, different chicken combs, such as the pea-comb [79], V-shape, buttercup [73], and Rose comb [78], are strongly linked to SVs in different genes. The pea and V-shape combs are associated with duplications in the SOX5 and EOMES genes, respectively; while the Rose comb is associated with an inversion that affects expression of the MNR2 gene, which is not within the inversion but located adjacent to its breakpoints. Interestingly, in all these cases, the SVs lead to the ectopic expression of the affected genes, likely impacting comb development and resulting in their phenotypic diversity (Fig. 2F). Moreover, the same genetic variant can have pleiotropic effects on several traits. For example, the inversion causing the rose comb phenotype also affects sperm mobility [78].

Additionally, the same phenotype can be achieved by similar types of SVs in different genes. In Zebra finches, sperm mobility is also influenced by an inversion, yet on a different chromosome than in the chicken example [78], clustering several genes into a supergene [98]. Supergenes

involve inversions which link genes by reducing the recombination rate, causing blocks of multiple genes to be transmitted as a unit, with the potential for co-adaptation. Because these supergenes include several genes, whereas SNPs are limited to more localized nucleotide differences generally affecting one gene, this type of SV may result in more complex phenotypic variation, such as changes in behaviors, compared to what may be generated by SNPs [99]. Among avian model systems, two such supergenes have been reported, one in Common quails [62] and the aforementioned one in Zebra finches [98], with different and pleiotropic phenotypic effects. In quails it is associated with geographically isolated populations that differ in several traits, including body size, throat color and wing shape; whereas the Zebra finch supergene affects sperm morphology and swimming speed (Fig. 2G).

Only four complex rearrangements have been reported in avian model systems, and due to their larger size, they typically impact multiple genes, potentially shaping various phenotypic traits. For example, in quails two inversions and a partial deletion that affect four genes result in changes in plumage coloration, body weight and temperature [63] (Fig. 2E). Two studies on hyperpigmentation [46,54] and muff and beard development [74,80] in chickens have reported SVs implicating the same genes. Interestingly, the initial set of studies for each trait showcased complex SVs, yet the second set, while trying to narrow down the genomic mechanism, reported only duplications. These studies illustrate the complexity of both characterizing SVs and understanding the genetic causes underlying a specific trait.

Some traits can have a complex genetic basis, and SVs are often associated to phenotypes in conjunction with other types of genetic variation, such as SNPs [e.g., 8,74,80]. Therefore, in non-model systems, certain traits that have been linked to SNPs, due to the current prevailing short-read methodologies, might actually have a more complex genetic basis also involving SVs. Overall, given the genetic complexity underlying phenotypic traits, it is important to account for multiple types of genetic variation when trying to find associations between phenotypes and genotypes.

Among the 103 reviewed articles, seven characterize SVs across different chicken breeds and populations [16,33,40,47,64,97,100], along with two each on turkey [101,102] and domestic ducks [3,4]. Exploring diverse breeds within a species offers an opportunity to examine whether similar phenotypes stem from comparable genetic mechanisms. For instance, the Creeper trait which involves abnormally short legs, is associated with the IHH gene in two chicken breeds. The IHH gene is completely deleted in Chinese Xingyi batam chickens [103], while a complex rearrangement involving deletions and an insertion, affects both the IHH and NHEJ1 genes in Japanese bantam chickens [104]. Additionally, there are instances where identical or nearly identical SVs in the same gene lead to the same phenotype. The frizzle feather trait is caused by a 15-bp deletion in the KRT75L4 gene in Kirin chickens [66] and Xiushui Yellow Chickens [105]. The same trait is observed in crosses between a heterozygous frizzle rooster and wild-type hens, generated by a 69-bp deletion with autosomal incomplete dominant inheritance in the same gene [71]. Another example, reviewed in [106], is blue egg coloration in Araucana, Chinese and European chicken breeds [57,58]. In these breeds, blue eggs are caused by the insertion of a ~4.2Kb retrovirus (EAV-HP) in the promoter region of the SLCO1B3 gene, leading to ectopic expression in the shell glands of the uterus. However, the integration site differs between the Asian breed and the Araucana and European breeds, suggesting two independent origins. Notably, similar SVs can also yield diverse phenotypic outcomes, exemplified by a SOX10 gene deletion generating both dark brown and yellow coloration in different chicken breeds [48,59]. Overall, these examples illustrate how SVs can generate phenotypic diversity in avian model systems.

The role of pangenomes in detecting SVs

The study of SVs is closely linked to the pangenome concept. Traditional reference-based genome studies have predominantly focused on a single reference genome, leading to the underrepresentation of SVs, as sequences from individuals which possess the SV may not map against reference genomes which lack them. Pangenomes integrate information from multiple genomes within a species or a group of related organisms, thus revealing a more

comprehensive landscape of genetic variation, including SVs [107]. Pangenomes aim to uncover the full spectrum of genetic variation, including both small and large-scale SVs, capturing the core genome shared among all individuals from that species and the dispensable genome containing non-reference sequences. The best resolution is achieved by generating pangenomes from high-quality reference genomes derived from long reads, ideally telomere-to-telomere, because short-read assemblies may not capture important variants, such as long repeats. Through pangenomic approaches, researchers have been able to detect and characterize previously unknown SVs that play a significant role in shaping phenotypic diversity [e.g., 108,109].

Pangenomes, which first emerged for bacteria, remain more prevalent in bacteria and plants, but there is a growing tendency and an increasing effort to generate pangenomes in other organisms [107]. Currently, the chicken [37,110] and the domestic duck [38] are the only avian model species with an available pangenome. This approach revealed new SVs associated with phenotypic traits, highlighting the power of using pangenomes to study the complex genomic basis of phenotypic diversity. Moreover, the emergence of the first pangenome in a non-model avian species, the barn swallow (*Hirundo rustica*) [111], demonstrates that advances in sequencing and bioinformatics are enabling the implementation of this approach in diverse organisms. The pangenomes themselves will also improve as larger numbers of individuals (and from different populations) are incorporated, leading to the increased detection of rare or population-specific variants.

How transposable elements (TEs) impact SVs and phenotype in avian model systems

Transposable elements (TEs) are mobile genetic elements that play a significant role in shaping genome structure, adaptation and the development of reproductive barriers [112]. TEs have the potential to act as building blocks for SV formation, as their insertion, deletion, duplication, or rearrangement can lead to gene modifications, altered recombination patterns, and changes in the genome's structural architecture. TEs can generate phenotypic variation through alterations in gene expression patterns due to the introduction of regulatory elements such as

promoters, silencers or enhancers or by modifying the spacing between these elements and promoters [113].

SVs are closely interrelated with TEs and therefore pangenomes are essential for their characterization, and to understand their relevance in evolutionary processes. Notably, the domestic duck pangenome revealed that the phenotypic impact of TE-related SVs can be important, exemplified by a Gipsy LTR element insertion in the promoter region of the IGF2BP1 gene, that accounts for a large proportion (27.61%) of the variation in body mass [38]. Moreover, the domestic duck pangenome [38] and Zebra finch inversions [114] have shown an accumulation of TEs at the breakpoints of SVs, suggesting a potential correlation between TEs and the generation of SVs. Specifically, the presence of endogenous retrovirus LTR retrotransposons is relatively common among avian model systems, and associates with different phenotypic traits such as blue eggshell in chickens [115] and domestic duck body size and plumage coloration [38]. Boman et al. [114] reported 4.5Mb of LTR in the Zebra finch genome, likely associated with the numerous inversions present in this species. However, the causality between the presence of LTR at the breakpoints and the generation of the inversion that affect sperm motility [19,116] remains to be established. Overall, more effort is needed to annotate and characterize the TE diversity and abundance in avian genomes, a challenging process that has likely led to their underreporting [117]. Investigating the impact of TEs in avian model systems, as well as their interactions with other genetic elements and environmental factors, will provide valuable insights on how they shape phenotypic diversity.

Concluding remarks and future perspectives

There are still relatively few studies associating SVs with phenotypic traits in avian model systems, and most examples detect small variants (<50 bp) in chickens (Fig. 1A and B). This knowledge gap is likely due to the requirements of third-generation sequencing and robust analytical methods for long SV detection, rather than indicating that their impact on the phenotype is insignificant. With the increased affordability of long-read sequencing methods,

the continuous improvement of bioinformatics to detect and characterize SVs, and the emergence of pangenomic approaches, we anticipate a shift in focus in the coming years. The almost exclusive emphasis on SNPs will give way to a more integrative approach that includes different types of genetic variants and their interactions, incorporating the detection of SVs and evaluating their role in shaping phenotypic traits. Many studies have initially associated certain phenotypes to specific SNPs, yet the underlying reality might be more complex. SVs may be at play, and uncovering these associations will provide a deeper understanding of such traits. Moreover, understanding the intricate relationship between SVs and TEs is crucial for comprehending the genetic basis underlying evolutionary processes. Further research is needed to elucidate the specific mechanisms by which TEs and SVs interact, including the impact of TEs on SV formation and the influence of SVs on TE behavior. This will provide a better understanding of the functional significance of SV-TE interactions and their contributions to phenotypic diversity in various organisms, including avian model and non-model systems.

The adoption of an integrative approach that studies multiple forms of genetic variation holds great potential to clarify how different types of variants interact to generate the wide diversity of phenotypic traits observed in avian species. Avian model systems provide an opportunity to understand the relative roles of SVs and their interrelationships with for example SNPs and TEs (see Outstanding questions). Avian model systems can serve as a valuable resource in disentangling the complex genetic mechanisms underlying phenotypic diversity, ultimately leading to a better understanding of gene regulation and expression. As the different techniques discussed in this review become more widely available, we expect to see associations between SVs and phenotypes in non-model avian systems become more common.

Box 1. Structural Variants (SVs) and their phenotypic effect

Structural variants (SVs) encompass a wide range of genomic alterations, ranging in size from small changes (~50 bp) to large-scale modifications spanning megabases. These mutations classified into two categories. Unbalanced changes lead to alterations in DNA content. These changes include insertions and deletions (indels), which are short-scale genetic changes involving the insertion or deletion of one or more nucleotides, Copy Number Variants (CNV) involving both deletions and duplications, and presence/absence variants (PAV) that represent changes related to the presence or absence of large genomic segments. Such mutations result in the loss or gain of DNA information. Secondly, balanced changes, such as inversions and inter or intra-chromosomal translocations, impact the orientation or location of DNA without altering the overall genetic content. Additionally, in a broader sense, SVs include insertions of transposable elements, tandem and segmental duplications, as well as complex rearrangements involving combinations of all these mutations, for example, inverted duplications (Figure I).

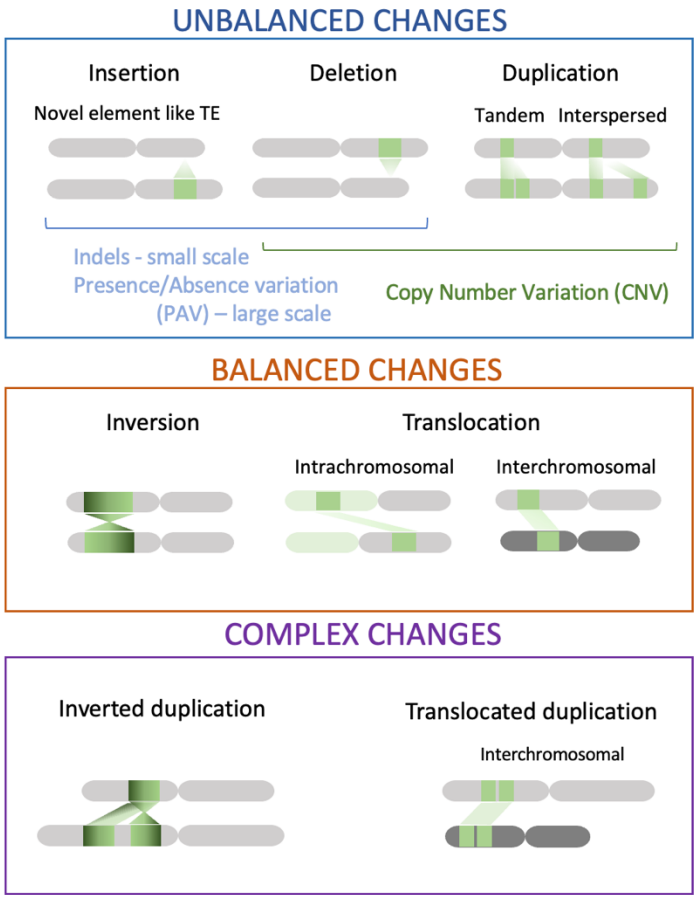

Figure I. Graphical representation of Structural Variants (SVs). SVs are categorized into unbalanced changes, which include insertions, deletions, and duplications; balanced changes, such as inversions and translocations; and complex changes that are a combination of the previous types.

In chickens, structural variation occurs in both coding and noncoding regions of the genome and the presence of these variants is positively correlated with chromosome size [118]. Furthermore, due to structural variants involving larger stretches of the genome compared to

SNPs, they have the potential to significantly impact phenotype [119,120]. SVs can affect gene expression through many mechanisms, including gene disruption, alteration of gene dosage, position effects, and disruption of gene expression at breakpoints [78]. SVs can also directly affect genes leading to the production of non-functional proteins or causing failures/modifications in mRNA translation or expression. Gene dosage alterations occur due to CNVs which cause changes in the number of gene copies, subsequently leading to modifications in gene expression. Gene expression could also be modified through position effects due to shifts in a gene's genomic location or changes in its surrounding chromatin environment that affect gene accessibility and expression. For instance, SVs are likely to alter the position of Cis-Regulatory Elements (CREs), such as promoters and enhancers. Not only can the SVs impact gene expression, but also their breakpoints (the edges at the 5' and 3' ends of the SV) can affect the expression of nearby genes [2,119,121].

Outstanding Questions Box

- How do structural variants contribute to the remarkable diversity of phenotypes observed in avian species, and what are the specific genetic mechanisms underlying this variation? Model species suggest SVs have a strong effect on phenotype and we expect the same to be true in non-model avian systems, once detection becomes more prevalent.

- What is the extent of structural variation in the avian genome, and how does it compare to other forms of genetic variation (e.g., SNPs), in terms of frequency and phenotypic impact? Additionally, how do structural variants interact with other sources of genetic variation, such as SNPs, TEs or regulatory elements, to shape complex phenotypic traits in avian model systems?

- What is the impact of TEs on SV formation and how do SVs influence TE behavior?

- To what extent do structural variants play a role in complex avian phenotypes, such as mating displays, vocalizations, or migratory patterns, and how do they influence social interactions and reproductive success?

- What are the evolutionary forces driving the maintenance or elimination of structural variants in avian populations, and how do they contribute to the generation of genetic diversity?

- How can the insights gained from studying SVs in avian model systems be translated to improve conservation initiatives, breeding programs and our understanding of the genetic basis of phenotypic traits in other avian species? Furthermore, what is the contribution of SVs to adaptations in avian populations, particularly in response to environmental changes such as habitat fragmentation and climate change?

Highlights

- Most SVs associated with phenotypic traits reported in avian model systems are short indels and predominantly detected in the chicken, which limits our understanding of their relevance in shaping the phenotype in model and non-model systems alike.

- Pangenomes, exemplified by the chicken and the domestic duck, improve SV detection and, in combination with long-read sequencing technologies, are crucial for characterizing SVs and exploring their impact on phenotypic traits.

- Moving towards an integrative approach that characterizes different forms of genetic variation, such as SNPs, SVs, TEs and their interactions, is crucial to improve our understanding of the mechanisms underlying phenotypic traits.

- There is a significant gap in our understanding regarding the complex interactions between TEs and SVs.

Glossary

Cis-regulatory elements (CREs) are non-coding DNA regions, including promoters, enhancers and silencers, that regulate the transcription of genes located in the same chromosome or neighboring genomic region.

C-terminal amino acid (AA): last amino acid in a protein sequence.

Ectopic Expression: atypical expression of a gene in a cell type, tissue, or developmental stage where it is normally inactive. This results from genetic or regulatory changes activating the gene in a novel context.

Enhancers: sequences that can increase transcription by interacting with the transcription machinery and can be located either upstream, downstream or within the intronic regions of the gene.

Exon: coding region of a gene that contains the instructions for producing a part of the final protein or functional RNA. Exons are interspersed with introns within a gene, and they are retained and joined together in the mature mRNA after splicing.

Gene expression: involves the processes of transcription, where the gene's DNA sequence is copied into mRNA, and translation, where mRNA directs the assembly of amino acids into proteins. This dual process carries the genetic information necessary for protein synthesis.

Gypsy Long Terminal Repeat (LTR) Transposable Element: type of TE that belongs to the class of retrotransposons, possesses long terminal repeats (LTR) at both ends and can transpose within a genome via an RNA intermediate.

Intron: non-coding regions of a gene between exons. During gene expression, introns are removed from the RNA transcript through RNA splicing.

Pangenome: Collection of representative DNA sequences from a species, including both the sequences shared among all individuals (core genome) and specific sequence information unique to subsets of individuals (variable genome).

Polygenic traits: phenotypes that are influenced by multiple genes, each contributing a small effect, in combination with environmental factors.

Promoters: sequences that provide a binding site for transcription factors and RNA polymerase, which initiate gene transcription and are usually located upstream of the gene's coding region.

Silencers: sequences that can modulate the transcription process by binding to repressors, effectively preventing transcription and leading to lower gene expression.

Single Nucleotide Polymorphism (SNP): genetic variation that occurs at a single position in the DNA sequence, where only one nucleotide differs among individuals.

Supergene: closely linked genes on a chromosome, inherited as a unit due to reduced recombination that results from being captured within an inversion. These genes often evolve together to control complex traits facilitating local adaptation.

Transcription factor (TF): protein that regulates gene expression by binding to specific DNA sequences, such as promoters, enhancers or silencers and recruiting the transcription machinery.

Transposable Element (TE): also known as "jumping genes", are DNA segments that can move within a genome. They can contribute to genetic variability by causing mutations, influence gene regulation, and have significant evolutionary implications.

References

1. Bickhart, D.M. and Liu, G.E. (2014) The challenges and importance of structural variation detection in livestock. *Front. Genet.* 5, 37.
2. Mérot, C. et al. (2020) A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol. Evol.* 35, 561–572.
3. van Dijk, E.L. et al. (2023) Genomics in the long-read sequencing era. *Trends in Genet*.
4. Zhou, Z. et al. (2018) An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat. Commun*. 9,2648.
5. Zhu, F. et al. (2021) Three chromosome-level duck genome assemblies provide insights into genomic variation during domestication. *Nat. Commun*. 12, 5932.
6. Hiragaki, T. et al. (2008) Recessive black is allelic to the yellow plumage locus in Japanese quail and associated with a frameshift deletion in the ASIP gene. *Genetics*. 178, 771–775.
7. Minvielle F. et al. (2010) The" silver" Japanese quail and the MITF gene: causal mutation, associated traits and homology with the" blue" chicken plumage. *BMC Genetics*. 11, 1–7.
8. Adetula, A.A. et al. (2020) RAI14 in the blood feather regulates chicken pigmentation. Archives *Animal Breeding*. 63, 231–239.
9. Lan, L.T.T. et al. (2021) Association of polymorphisms in prolactin receptor and melatonin receptor 1c genes on egg production and egg quality traits of japanese quails (*Coturnix coturnix japonica*). *JAPS.* 31.
10. Manoharan, A. et al. (2021) Identification of 24bp indel (s) Polymorphism in the Promoter Region of prolactin Gene and its Association with Broodiness in Tellicherry Native Chicken. *Indian J. Anim. Res*. 55, 1137–1140.
11. Vinh, N.T. et al. (2021) Single nucleotide polymorphisms of candidate genes related to egg production traits in Vietnamese indigenous chickens. *Braz. J. Poultry Sci*. 23, eRBCA–2020.
12. Cao, Z.P. (2007) Association of Spot14α gene polymorphisms with body weight in the chicken. *Poultry Sci*. 86, 1873–1880.
13. Fernandes, A.C. et al. (2021) Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics*. 22, 1–18.
14. Fu, R. et al. (2020) A novel 65-bp Indel in the GOLGB1 gene is associated with chicken growth and carcass traits. *Animals*. 10, 475.
15. Han, R. et al. (2019) Chicken ZNF764L gene: mRNA expression profile, alternative splicing analysis and association analysis between first exon indel mutation and economic traits. *Gene*. 695, 92–98.
16. Han, R. et al. (2014) Identification and functional characterization of copy number variations in diverse chicken breeds. *BMC Genomics*. 15, 1–10.
17. Hirwa, C.D. et al. (2010) Effects of the thyroid hormone responsive spot 14α gene on chicken growth and fat traits. *Poultry Sci*. 89, 1981–1991.
18. Jing, Z. et al. (2020) Detection of CNV in the SH3RF2 gene and its effects on growth and carcass traits in chickens. *BMC Genetic*s. 21, 1–7.

19. Knief, U. et al. (2016) Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biol.* 17, 1–22.
20. Lei, M. et al. Polymorphism of growth-correlated genes associated with fatness and muscle fiber traits in chickens. *Poultry Sci*. 86, 835–842.
21. Li, T. et al. (2021) A 104-bp Structural Variation of the ADPRHL1 Gene Is Associated With Growth Traits in Chickens. *Front. Genet*. 12, 691272.
22. Li, T. et al. (2022) A novel 27-bp indel in the intron region of the YBX3 gene is associated with growth traits in chickens. *Br. Poult. Sci.* 63, 590–596.
23. Liang, K. et al. (2019) Molecular characterization and an 80-bp indel polymorphism within the prolactin receptor (PRLR) gene and its associations with chicken growth and carcass traits. *3 Biotech*. 9, 1–10.
24. Lin, S. et al. (2018) Copy number variation in SOX6 contributes to chicken muscle development. *Genes*. 9, 42.
25. Lin, S. et al. (2022) Mining of chicken muscle growth genes and the function of important candidate gene RPL3L in muscle development. *Front. Physiol.* 2295.
26. Lin, W. et al. (2021) Novel 61-bp indel of RIN2 is associated with fat and hatching weight traits in chickens. *Front. Genet*. 12, 672888.
27. Lin, Z.T. et al. (2023) A 2-bp deletion in intron 1 of TMEM182 is associated with TMEM182 mRNA expression and chicken body weight. *Br. Poult. Sci*. 64, 11–18.
28. Liu, D. et al. (2019) A novel 86-bp indel of the motilin receptor gene is significantly associated with growth and carcass traits in Gushi-Anka F2 reciprocal cross chickens. *Br. Poult. Sci.* 60, 649–658.
29. Qin, P. et al. (2023) Molecular Characterization, Expression Profile, and A 21-bp Indel within the ASB9 Gene and Its Associations with Chicken Production Traits. *Genes*. 14, 339.
30. Ren, T. et al. (2019) Two insertion/deletion variants in the promoter region of the QPCTL gene are significantly associated with body weight and carcass traits in chickens. *Anim. Genet*. 50, 279–282.
31. Ren, T. et al. (2020) A 51 bp indel polymorphism within the PTH1R gene is significantly associated with chicken growth and carcass traits. *Anim. Genet*. 51, 568–578.
32. Shimogiri, T. et al. (2012) Detection of a Polymorphism Associated with Shank Length and Body Weight in Japanese Quail (Coturnix japonica) by AFLP. *Poult. Sci. J.* 49, 5–11.
33. Sohrabi, S.S. et al. (2018) A. Detection of breed-specific copy number variations in domestic chicken genome. *Genome*. 61, 7–14.
34. Wang, Y.X. (2012) Correlation Analysis between Single Nucleotide Polymorphism of Fatty Acid Synthase and Fatness Traits in Chickens. *Adv. Mat. Res*. 535, 2283–2286.
35. Wang, X. et al. (2020) Association of growth traits with a structural variation downstream of the KCNJ11 gene: a large population-based study in chickens. *Br. Poult. Sci.* 61, 320–327.
36. Wang, X. et al. (2020) Association of a new 99-bp indel of the CEL gene promoter region with phenotypic traits in chickens. *Sci. Rep.* 10, 3215.

37. Wang, K. et al. (2021) The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. *Mol. Biol. Evol.* 38, 5066–5081.
38. Wang, K. et al. (2023) Duck pan-genome reveals two transposon-derived structural variations caused bodyweight enlarging and white plumage phenotype formation during evolution. *BioRxiv*. 2023–01.
39. Wei, C. et al. (2020) Molecular characterization and a duplicated 31-bp indel within the LDB2 gene and its associations with production performance in chickens. *Gene*. 761, 145046.
40. Zhang, H. et al. (2014) Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. *BMC Genomics*. 15, 1–12.
41. Zhao, N.N. et al. (2015) VLDLR gene polymorphism associated with abdominal fat in Gaoyou domestic duck breed. *Czech J. Anim. Sci*. 60, 178–184.
42. Zhao, X.H. et al. (2015) Single nucleotide polymorphisms in IGFBP-2 gene and their associations with body weight traits on Jinghai Yellow chicken. *Braz. J. Poult. Sci*. 17, 497–502.
43. Bruders, R. et al. (2020) A copy number variant is associated with a spectrum of pigmentation patterns in the rock pigeon (*Columba livia*). *PLoS Genet*. 16, e1008274.
44. Chang, C.M. et al. (2006) Complete association between a retroviral insertion in the tyrosinase gene and the recessive white mutation in chickens. *BMC Genomics*. 7, 1–15.
45. Domyan, E.T. et al. (2016) Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. *Elife*. 5, e12115.
46. Dorshorst, B. et al. (2011) A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. *PLoS Genet*. 7, e1002412.
47. Drobik-Czwarno, W. et al. (2018) Detection of copy number variations in brown and white layers based on genotyping panels with different densities. *Gen. Sel. Evol*. 50, 1–13.
48. Gunnarsson, U. et al. (2011) The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. *PCMR*. 24, 268–274.
49. Krishnan, S. and Cryberg, R.L. (2019) Effects of mutations in pigeon Mc1r implicate an expanded plumage color patterning regulatory network. *BioRxiv*. 792945.
50. Krishnan, S. (2019) Nested tandem duplications of the gene Melanoma antigen recognized by T-cells (Mlana) underlie the sexual dimorphism locus in domestic pigeons. *BioRxiv*. 754986.
51. Maclary, E.T. et al. (2023) An allelic series at the EDNRB2 locus controls diverse piebalding patterns in the domestic pigeon. *BioRxiv*. 2023–07.
52. Robic, A. et al. (2019) Two new structural mutations in the 5’ region of the ASIP gene cause diluted feather color phenotypes in Japanese quail. *Gen. Sel. Evol*. 51, 1–10.
53. Shen, Q. et al. (2022) Genome-Wide Association Study Identifies Candidate Genes for Stripe Pattern Feather Color of Rhode Island Red Chicks. *Genes*. 13, 1511.

54. Shinomiya, A. et al. (2012) Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (Fibromelanosis) in silky chickens. *Genetics*. 190, 627–638.
55. Tobita-Teramoto, T.J.G.K. et al. (2000). Autosomal albino chicken mutation (ca/ca) deletes hexanucleotide (-ΔGACTGG817) at a copper-binding site of the tyrosinase gene. *Poult. Sci*. 79, 46-50.
56. Vickrey, A.I. et al. (2018) Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon. *Elife*. 7, e34803.
57. Wang, Z. et al. (2013) An EAV-HP insertion in 5′ flanking region of SLCO1B3 causes blue eggshell in the chicken. *PLoS Genet.* 9, e1003183.
58. Wragg, D. et al. (2013) Endogenous retrovirus EAV-HP linked to blue egg phenotype in Mapuche fowl. *PloS One*. 8, e71393.
59. Zhu, T. et al. (2022) A deletion upstream of SOX10 causes light yellow plumage colour in chicken. *Genes*. 13, 327.
60. Han, R.L. et al. (2011) Novel 9-bp indel in visfatin gene and its associations with chicken growth. *British Poult. Sci*. 52, 52–57.
61. Domyan, E.T. et al. (2014) Epistatic and combinatorial effects of pigmentary gene mutations in the domestic pigeon. *Curr. Biol*. 24, 459–464.
62. Sanchez-Donoso, I. et al. (2022) Massive genome inversion drives coexistence of divergent morphs in common quails. *Curr. Biol*. 32, 462–469.
63. Bed'hom, B. et al. (2012) The lavender plumage colour in Japanese quail is associated with a complex mutation in the region of MLPH that is related to differences in growth, feed consumption and body temperature. *BMC Genomics*. 13, 1–10.
64. Chen, X. et al. (2022) Population Genomic Sequencing Delineates Global Landscape of Copy Number Variations that Drive Domestication and Breed Formation of in Chicken. *Front. Genet*. 13, 830393.
65. Derks, M.F.L. et al. (2018) Early and late feathering in turkey and chicken: same gene but different mutations. *Gen. Sel. Evol.* 50, 1–7.
66. Dong, J. et al. (2018) A novel deletion in KRT75L4 mediates the frizzle trait in a Chinese indigenous chicken. *Gen. Sel. Evol*. 50, 1–9.
67. Elferink, M.G. et al. (2008) Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC Genomics*. 9, 1–9.
68. Li, J. et al. (2021) The crest phenotype in domestic chicken is caused by a 195 bp duplication in the intron of HOXC10. *G3*. 11, jkaa048.
69. Li, J. et al. (2020) Mutations upstream of the TBX5 and PITX1 transcription factor genes are associated with feathered legs in the domestic chicken. *Mol. Biol. Evol*. 37, 2477–2486.
70. Mou, C. et al. (2011) Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. *PLoS Biology*. 9, e1001028.

71. Ng, C.S. et al. (2012) The chicken frizzle feather is due to an α-keratin (KRT75) mutation that causes a defective rachis. *PLoS Genet.* 8, e1002748.
72. Shen, Q. et al. (2023) Identification of Duplication Genotypes of the Feathering Rate Gene in Chicken by a Multiplex PCR Following Electrophoresis and/or Sanger Sequencing. *Animals*. 13, 1091.
73. Dorshorst, B. et al. (2015) A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet.* 11, e1004947.
74. Guo, Y. et al. (2016) A complex structural variation on chromosome 27 leads to the ectopic expression of HOXB8 and the muffs and beard phenotype in chickens. *PLoS Genet*. 12, e1006071.
75. Imsland, F. et al. (2012) The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet*. 8, e1002775.
76. Moro, C. et al. (2015) Quantitative effect of a CNV on a morphological trait in chickens. *PLoS One*. 10, e0118706.
77. Sato, S. et al. (2010) Sequence analysis of a pea comb locus on chicken chromosome 1. *Anim. Genet*. 41, 659–661.
78. Wang, Y. et al. (2017) Transcriptome analysis of comb and testis from Rose-comb Silky chicken (R1/R1) and Beijing Fatty wild type chicken (r/r). *Poult. Sci*. 96, 1866–1873.
79. Wright, D. et al. (2009) Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. *PLoS Genet*. 5, e1000512.
80. Yang, K.X. et al. (2020) Copy number variation in HOXB7 and HOXB8 involves in the formation of beard trait in chickens. *Anim. Genet*. 51, 958–963.
81. Yang, Z. et al. (2021) Genome-wide association study using whole-genome sequencing identifies a genomic region on chromosome 6 associated with comb traits in Nandan-Yao chicken. *Front. Genet*. 12, 682501.
82. Cui, J.X. et al. (2006) Association of polymorphisms in the promoter region of chicken prolactin with egg production. *Poult. Sci*. 85, 26–31.
83. Huang, T. et al. (2018) A copy number variation generated by complicated organization of PCDHA gene cluster is associated with egg performance traits in Xinhua E-strain. *Poult. Sci*. 97, 3435–3445.
84. Abe, H. , et al. (2013). Short copy number variations potentially associated with tonic immobility responses in newly hatched chicks. *PLoS One*. 8, e80205.
85. Falker-Gieske, C. et al. (2023) Structural variation and eQTL analysis in two experimental populations of chickens divergently selected for feather-pecking behavior. *Neurogenetics*. 24, 29–41.
86. Khatri, B. et al. (2019) Copy number variation study in Japanese quail associated with stress related traits using whole genome re-sequencing data. *PLoS One*. 14, e0214543.

87. Komiyama, T. et al. (2014) Dopamine receptor genes and evolutionary differentiation in the domestication of fighting cocks and long-crowing chickens. *PLoS One*. 9, e101778.
88. Krause, E.T. et al. (2019) Fear but not social behaviour is affected by a polymorphism in the 5'-flanking region of the serotonin transporter (5-HTT) gene in adult hens. *Behav. Brain Res*. 361, 50–53.
89. Rubin, C.J. et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature.* 464, 587–591.
90. Seol, D. et al. (2019) Identification of copy number variation in domestic chicken using whole-genome sequencing reveals evidence of selection in the genome. *Animals*. 9, 809.
91. Bai, H. et al. (2018) Genome-wide detection of CNV s associated with beak deformity in chickens using high-density 600K SNP arrays. *Anim. Genet*. 49, 226–236.
92. Chang, C.F. et al. (2014) The cellular and molecular etiology of the craniofacial defects in the avian ciliopathic mutant talpid2. *Development.* 141, 3003–3012.
93. Gu, L. et al. (2017) Novel copy number variation of the TGFβ3 gene is associated with TGFβ3 gene expression and duration of fertility traits in hens. *PloS One*. 12, e0173696.
94. Liu, X. et al. (2020) Genome-wide association study of muscle glycogen in jingxing yellow chicken. *Genes*. 11, 497.
95. Xu, Y. et al. (2022) Genome-wide association analysis reveals 6 copy number variations associated with the number of cervical vertebrae in Pekin ducks. *Front. Cell Dev. Biol*. 10, 1041088.
96. Yuan, X. et al. (2022) Fatty acid metabolism-related genes are associated with flavor-presenting aldehydes in Chinese local chicken. *Front. Genet*. 13, 902180.
97. Rao, Y.S. et al. (2016) Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. *Poult. Sci.* 95, 1750–1756.
98. Kim, K.W. et al. (2017) A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nat. Ecol. Evol*. 1, 1168–1176.
99. Taylor, S. and Campagna, L. (2016) Avian supergenes. Science. 351, 446-447.
100. Fan, W.L. et al. (2013) Genome-wide patterns of genetic variation in two domestic chickens. *Gen. Biol. Evol*. 5, 1376–1392.
101. Strillacci, M.G. et al. (2019) Copy number variation mapping and genomic variation of autochthonous and commercial turkey populations. *Front. Genet*. 10, 982.
102. Strillacci, M.G. et al. (2021) Copy number variants in four Italian turkey breeds. *Animals*. 11, 391.
103. Jin, S. et al. (2016) Deletion of Indian hedgehog gene causes dominant semi-lethal Creeper trait in chicken. *Sci. Rep*. 6, 30172.
104. Kinoshita, K. et al. (2020) Combined deletions of IHH and NHEJ1 cause chondrodystrophy and embryonic lethality in the Creeper chicken. *Comm.s Biol*. 3, 144.
105. Chen, B. et al. (2022) Deletion in KRT75L4 linked to frizzle feather in Xiushui Yellow Chickens. Anim. Genet. 53, 101–107.

106. Campagna, L. and Toews, D.P.L. (2022) The genomics of adaptation in birds. *Curr. Biol*. 32, R1173–R1186.
107. Gong, Y. et al. (2023) A review of the pangenome: how it affects our understanding of genomic variation, selection and breeding in domestic animals? *J. Anim. Sci. Biotech.* 14, 1–19.
108. Li, N. et al. (2023) Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet*. 55, 852–860.
109. Liao, W.W. et al. (2023) A draft human pangenome reference. *Nature*. 617, 312–324.
110. Rice, E.S. et al. (2023) A pangenome graph reference of 30 chicken genomes allows genotyping of large and complex structural variants. *Research Square*.
111. Secomandi, S. et al. (2023) A chromosome-level reference genome and pangenome for barn swallow population genomics. *Cell Rep*. 42.
112. Bourque, G. et al. (2018) Ten things you should know about transposable elements. *Genome Biol*. 19, 1–12.
113. Bourgeois, Y. and Boissinot, S. (2019) On the population dynamics of junk: a review on the population genomics of transposable elements. *Genes*. 10, 419.
114. Boman, J. et al. (2019) The genome of blue-capped cordon-bleu uncovers hidden diversity of LTR retrotransposons in zebra finch. *Genes*. 10, 301.
115. Altgilbers, S. et al. (2022) Quantitative analysis of CRISPR/Cas9-mediated provirus deletion in blue egg layer chicken PGCs by digital PCR. *Sci. Rep*. 12, 15587.
116. Knief, U. et al. (2017) A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. *Nat. Ecol. Evol*. 1, 1177–1184.
117. Kapusta, A. and Suh, A. (2017) Evolution of bird genomes—a transposon's-eye view. *Ann. NY Acad. Sci*. 1389, 164–185.
118. Zhang, J., et al. (2022) A 4.1 kb deletion in IRX1 gene upstream is completely associated with rumplessness in Piao chicken. *Genomics*. 114, 110515.
119. Zhang. L. et al. (2021) How important are structural variants for speciation? *Genes*. 12, 1084.
120. Chiang, C. et al. (2017) The impact of structural variation on human gene expression. *Nat. Genet*. 49, 692–699.
121. Spielmann, M. et al. (2018). Structural variation in the 3D genome. *Nat. Rev. Genet*. 19, 453–467.

Figures

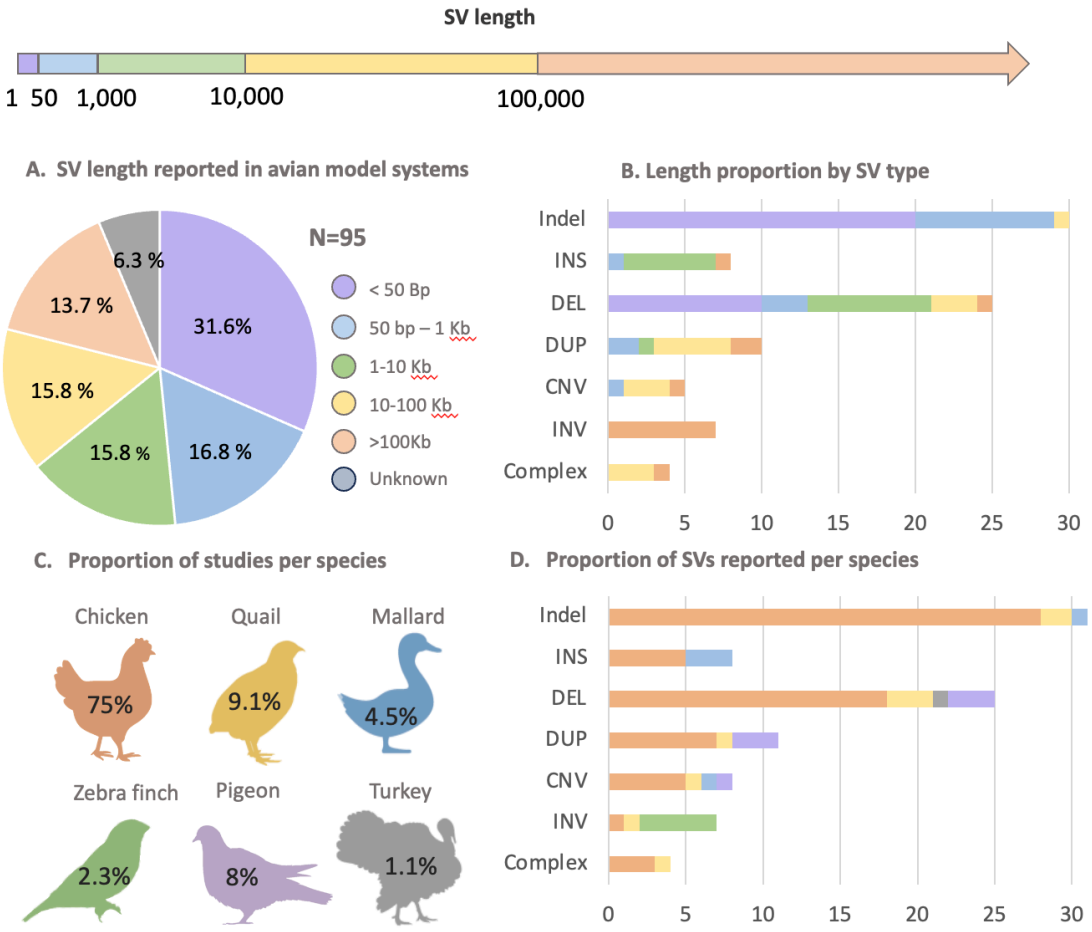

Figure 1. Summary of Structural Variants (SVs) reported in avian model systems. A) Length distribution of SVs associated with phenotypic traits categorized in the following intervals: <50 bp, 50-100 bp, 1-10 Kb, 10-100 Kb, >100 Kb and Unknown length. B) Length proportion by SV type, including indels, insertions (INS), deletions (DEL), duplications (DUP), Copy Number Variation (CNV) that include both deletions and duplications, inversions (INV) and complex rearrangements. The total number of studies reporting each type of SV is also indicated. C) Proportion of studies reporting SVs associated with phenotypic traits per species, including the chicken, the Common and Japanese quails, the domestic Mallard duck, the Zebra finch and the Wild turkey. D) Proportion of SV type reported per species, including the same SV types than those shown in B.

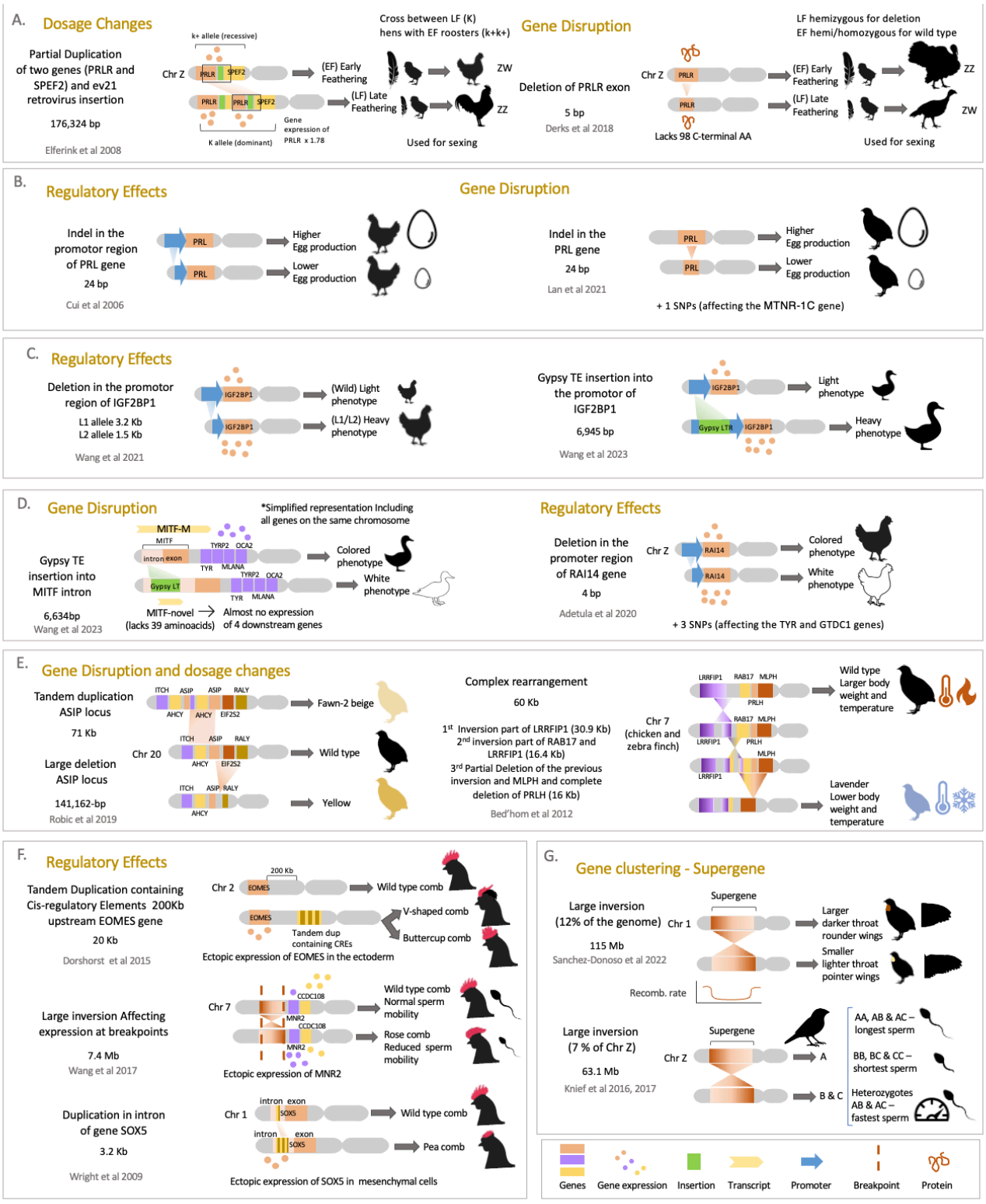

Figure 2. Examples of SVs affecting phenotypic traits in different avian model systems. A) Different SVs affecting the PRLR gene in chickens [67] and turkeys [65] that lead to changes in feathering time. This trait is linked to the Z sex chromosome and can be used for sexing in specific breads because females are heterogametic (ZW) and males

homogametic (ZZ). B) Indel in the PRL gene or its promoter in chickens [82] and Japanese quails [9] that affects egg production. C) Different SVs affecting the IGF2BP1 promotor in chicken [37] and domestic ducks [38] modulate body weight in both species. D) Different SVs in different genes generate the white phenotype in domestic ducks [38] and chickens [8], but in both cases the TYR gene is implicated. In the duck example the representation is simplified, including all the genes on the same chromosome, yet in reality some genes are found on different chromosomes. E) Different SVs in the ASIP gene generate variation in quail plumage coloration [52] and a large complex rearrangement affecting several genes modify several traits in quail, including plumage coloration, body weight and temperature [63]. F) Different SVs affect many genes and lead to their ectopic expression generating chicken comb diversity [73,78,79]. G) Large inversions in quail [62] and the Zebra finch [19,116] result in supergenes affecting different traits in each species.