

Review

Coevolution of Genome Architecture and Social Behavior

Dustin R. Rubenstein,^{1,*,@} J. Arvid Ågren,^{2,@} Lucia Carbone,^{3,4,@} Nels C. Elde,^{5,@} Hopi E. Hoekstra,^{2,6,@} Karen M. Kapheim,^{7,@} Laurent Keller,⁸ Corrie S. Moreau,^{9,@} Amy L. Toth,^{10,@} Sam Yeaman,^{11,@} and Hans A. Hofmann^{12,*,@}

Although social behavior can have a strong genetic component, it can also result in selection on genome structure and function, thereby influencing the evolution of the genome itself. Here we explore the bidirectional links between social behavior and genome architecture by considering variation in social and/or mating behavior among populations (social polymorphisms) and across closely related species. We propose that social behavior can influence genome architecture via associated demographic changes due to social living. We establish guidelines to exploit emerging whole-genome sequences using analytical approaches that examine genome structure and function at different levels (regulatory vs structural variation) from the perspective of both molecular biology and population genetics in an ecological context.

Bidirectional Relationships between Social Behavior and Genome Architecture

Behavioral traits can be difficult to study because their development is often only poorly understood, they can change rapidly in response to the (social or ecological) environment, and they can be flexible in their expression. **Social behavior** (see [Glossary](#)), or the behavioral interactions between two or more individuals typically of the same species, can be particularly challenging to tackle mechanistically, not only because it encompasses a wide range of complex behavioral actions that emerge from a suite of less complex correlated displays or phenotypic traits, but also because the fitness of one individual depends on the interactions with others [1]. In addition, social traits often show high levels of behavioral flexibility [2], which we hypothesize can select for genomic characteristics that enhance **phenotypic plasticity**, such as increasing complexity of gene regulation or an increased reliance on **epigenetic mechanisms**. Intraspecific variation in social behavior is often associated with the evolution of alternative social states within a species, such as different mating strategies or forms of social organization. These **social polymorphisms** may themselves be plastic and related to differences in environmental factors [3,4] or fixed throughout an organism's lifetime and related to features of the genome such as **chromosomal inversions** [5,6] or **gene duplications** [7,8]. Social interactions can also generate emergent 'group-level' phenotypes, such as dominance hierarchies or caste systems, which in turn can favor gene duplication, **neofunctionalization**, or the evolution of novel genes [8,9].

Most studies examining the genomic underpinnings of social behavior [10–12] are typically unidirectional, in that they seek to identify genes that underlie behavior, but fail to consider how social behavior might affect genome structure and function and thereby influence the evolution of the genome itself (Figure 1). Importantly, the genome is not simply a sequence of nucleotides or a collection of genes. Rather, the genome has an intricate architecture, including a complex regulatory machinery, mobile elements, and chemical modifications, all of which potentially influence – and can be influenced by – behavioral phenotypes in underappreciated ways. For example, genome evolution (i.e., rates of molecular evolution) in ants has been shown to have been shaped by mu-

Highlights

Social behavior has a genetic basis but can also affect genome structure and function.

As a consequence, social behavior can influence the evolution of the genome itself.

We discuss bidirectional links between social behavior and genome architecture.

We argue that social living can influence genome architecture via demographic processes.

We propose guidelines for examining genome structure and function at different levels.

¹Columbia University, Department of Ecology, Evolution, and Environmental Biology and Center for Integrative Animal Behavior, New York, NY 10027, USA

²Harvard University, Department of Organismic and Evolutionary Biology, Cambridge, MA 02138, USA

³Oregon Health & Science University, Department of Medicine, KCVI, Portland, OR 97239, USA

⁴Oregon National Primate Research Center, Division of Genetics, Beaverton, OR 97006, USA

⁵University of Utah School of Medicine, Department of Human Genetics, Salt Lake City, UT 84112, USA

⁶Harvard University, Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Cambridge, MA 02138, USA

⁷Utah State University, Department of Biology, Logan, UT 84322, USA

⁸University of Lausanne, Department of Ecology and Evolution, Biophore, UNIL, 1015 Lausanne, Switzerland



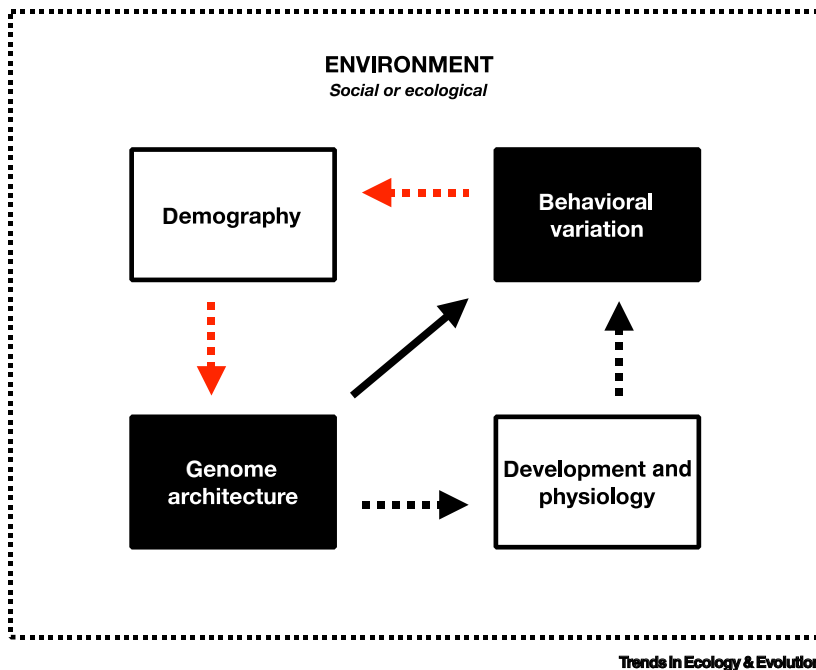


Figure 1. Bidirectional Links and Feedbacks between Social Behavior and the Genome. Most attempts to link behavioral variation to genome architecture do so directly by studying the genetic underpinnings of social polymorphisms (unbroken black arrow) or indirectly by identifying developmental and physiological pathways underlying social phenotypes (broken black arrows). However, variation in social or mating systems also influence demographic factors such as effective population size (N_e) through variation in operational sex ratio, age structure, and reproductive skew. These demographic differences can in turn influence the genome by altering genetic processes (e.g., recombination rates) or attributes of genome architecture (broken red arrows). Social behavior and demography are influenced by the environment, both social and ecological, as indicated by the dashed box.

tualistic associations with plants [13], illustrating how complex social behavior can influence genome sequence. The ecological conditions underlying social behavior, as well as the unique demographic attributes of social living, may also influence **genome architecture** via the interplay between natural selection and genetic drift. **Operational sex ratio**, age structure (e.g., overlapping generations), and **reproductive skew** (as a consequence of social organization) can lead to a reduced **effective population size (N_e)**, which in turn can affect genome architecture via changes in the efficacy of natural selection and the increased chance of genetic drift. For example, high reproductive skew in eusocial insect colonies results in a low N_e , despite often having large numbers of individuals [14]. Demography-related changes in genome architecture may include **recombination rate**, mutation rate, **transposable element (TE)**, or **simple sequence repeat (SSR)** accumulation, and/or replication error [15]. In addition to influencing sequence variation, these genomic changes can also affect **ploidy** levels, **gene copy number variation (CNV)**, chromosomal inversions, and **novel and orphan gene** formation. All of these processes have been suggested to potentially influence – or be influenced by – social evolution, although a comprehensive framework for the study of the coevolution of social behavior and genome architecture is still lacking.

Here we hypothesize that social behavior can profoundly affect the processes that drive the evolution of genome architecture, mediated through the effects of demographic changes on natural selection via the response to interactions with ecological factors. To fully understand the complex and bidirectional interrelationships between social behavior and genome architecture,

⁹Cornell University, Departments of Entomology and Ecology and Evolutionary Biology, Ithaca, NY 14850, USA

¹⁰Iowa State University, Department of Ecology, Evolution, and Organismal Biology and Department of Entomology, Ames, IA 50011, USA

¹¹University of Calgary, Department of Biological Sciences, Calgary, AB T2N 1N4, Canada

¹²The University of Texas at Austin, Department of Integrative Biology and Institute for Cellular and Molecular Biology, 2415 Speedway C-0990, Austin, TX 78712, USA

*Correspondence: dr2497@columbia.edu (D.R. Rubenstein) and hans@utexas.edu (H.A. Hofmann).
 @Twitter: @DustRubenstein (D.R. Rubenstein), @arvidagren (J.A. Ågren), @Carbone_Lab (L. Carbone), @EIEarlyBird (N.C. Elde), @hopihoekstra (H.E. Hoekstra), @KapheimLab (K.M. Kapheim), @CorrieMoreau (C.S. Moreau), @Amy_L_Toht (A.L. Toht), @sam_yeamann (S. Yeaman), and @HofmannLab (H.A. Hofmann).

researchers must examine genome structure and function at different levels (i.e., regulatory vs structural variation) and consider how socially and ecologically driven changes in demography shape genome architecture. The rise of ever more powerful genomic tools provides an opportunity to explore and generate new questions and hypotheses linking sociality and the genome. In what follows, we review the literature on these connections, develop a predictive framework for thinking about these bidirectional relationships, and outline avenues of future research about where we expect to see links between intra- and interspecific variation in social behavior and genome features.

Attributes of Genome Architecture

As more whole-genome data from individuals of the same species have become available, it is clear that there is an unexpected amount of individual and population variability in genome structure that is associated with variation in phenotypes. Mutations affecting single base pairs (SNPs) are the most studied form of genetic variation (in either the coding or regulatory regions), although small insertions or deletions (indels), gene CNVs and gene family expansions, and even large chromosomal rearrangements (e.g., inversions) can also commonly be found. Changes in genome architecture can occur through a variety of mechanisms, with different effects depending on the scale and frequency of these events, as numerous processes often work together to dynamically alter the genome over time.

At the most global level, **genome size** can be highly variable across eukaryotic species, ranging from 2.3 Mbp to 150 giga base pairs (Gbp), mostly driven by the accumulation of repetitive sequences such as TEs or SSRs [16]. Variation in genome size affects numerous cellular properties (e.g., size, proliferation, differentiation) and as a consequence can affect brain [17] and body [18] size. Increases in genome size also are correlated with genomic changes that can influence genome structure, such as decreases in recombination rate and increases in mutation rate [19, 20]. Whether changes in genome size affect the evolution of behavioral traits *per se* is less clear. However, recent evidence suggests that genome size variation underlies in part the extensive diversity of social structures in *Synalpheus* snapping shrimps [21].

Many other structural attributes of genome architecture can vary and change within as well as across species, often as a consequence of DNA replication or recombination (Figure 2 and Table 1). It has recently been shown that mutation rates can be elevated in specific genomic loci in ways that can substantially affect evolution under realistic demographic conditions. Specifically, repeated independent deletions of the enhancer of the *Pitx1* gene have been causally linked to the loss of pelvic hind fins in stickleback fish (*Gasterosteus aculeatus*), as different populations of this species adapted to freshwater habitats multiple times independently [22]. Notably, increased DNA fragility in this enhancer is the consequence of SSRs (TG dinucleotide repeats), which increase the frequency of double-strand breaks and deletions during DNA replication [23].

Studies linking genome architecture to behavioral variation directly or indirectly through changes in development and physiology have been successful at identifying the genetic underpinnings of a variety of social characteristics, ranging from **alternative reproductive tactics** [7] to differences in social organization in eusocial species [5,6] (Figure 1). However, the converse – where mating system and/or social behavior influence the structure of the genome – is also likely to occur (e.g. [24]). Thus, we argue (and discuss in detail below) that a species' mating and social system can alter its genome architecture through dynamic processes such as recombination, gene duplication, inversion, and TE accumulation as a consequence of effects on demography and subsequent shifts in the efficacy of natural selection relative to genetic drift (Figure 1). Below we highlight three attributes of genome architecture that have been shown to be related

Glossary

Alternative reproductive tactics:

differences in suites of life history, morphological, physiological, and behavioral traits that affect nearly every aspect of their reproductive biology.

Chromosomal inversion:

rearrangement in which a segment of a chromosome is reversed end to end.

Dosage overcompensation: adaptive outcome of gene duplication when both copies retain the original function but increase gene product output.

Effective population size (N_e):

number of breeding individuals in the population.

Epigenetic mechanisms: processes (e.g., DNA methylation, histone modification, noncoding RNA regulation) that determine chromatin structure, accessibility of genetic loci to transcriptional machinery, and gene expression levels.

Gene copy number variation (CNV): additional copies of DNA sequence and losses of genetic material at an intermediate scale (i.e., DNA segments 1 kb to 5 Mb in length). The number of repeated sections can vary between individuals.

Gene duplication: any duplication of a stretch of DNA that contains a gene.

Gene order: organization of genes within a genome.

Genome architecture: nonrandom structural and spatial attributes of the genome.

Genome size: total amount of DNA contained in one copy of a single genome.

Genomic imprinting: parent-of-origin effect on a gene's expression during embryo or endosperm formation.

Neofunctionalization: outcome of gene duplication when one paralog assumes a new, and adaptive, function.

Novel and orphan genes: genes lacking detectable orthologs due to *de novo* origination, duplication and divergence, or horizontal gene transfer.

Odorant receptor (OR): proteins expressed in olfactory receptor neurons that bind odorants and are critical for the sense of smell.

Operational sex ratio: the ratio of sexually competing males to sexually competing females that are ready to mate.

Phenotypic plasticity: the ability of a genotype to produce more than one phenotype when exposed to different environments.

Ploidy: number of complete sets of chromosomes in a cell.

bidirectionally to social behavior within (gene CNV, **supergenes**) or across (gene duplications/gene family expansions) vertebrate and invertebrate species.

Supergenes

Supergenes are clusters of tightly linked loci with related phenotypic effects, characterized by reduced recombination, often also involving a chromosomal inversion event [25]. TEs have been suggested to play a role in generating inversions as well as relocating genes into supergenes [26–28]. One of the best examples linking supergenes to social behavior comes from variation in the number of queens per colony in eusocial insects, which is frequent in ants and is typically associated with differences in a suite of reproductive and life history traits including queen phenotype, queen lifespan, breeding strategy, mode of dispersal, and mode of colony founding [3]. A recent study revealed that variation in social organization in the fire ant *Solenopsis invicta* is controlled by a supergene comprising a large nonrecombining chromosomal region with more than 500 genes in tight linkage [5]. The lack of recombination between the two existing haplotypes allows the two social forms to express fixed differences in several behavioral and life history traits that are essential to social evolution [29], including the mode of dispersal, body size, and chemical odor of individuals. Interestingly, variation in social organization in another ant species, the alpine silver ant *Formica selysi*, also was found to be under the control of a single supergene or ‘social chromosome’ [6], which evolved independently from the supergene in *S. invicta*. In addition to differences in social structure, other social polymorphisms in which multiple traits are linked together also are controlled by supergenes in plants [30], cichlid fishes [31], birds [32–34], and insects [35, 36].

Because supergenes allow the cosegregation of adaptive variation within species, they may facilitate the spread of complex phenotypes across species boundaries and be more common in social species than currently realized [25]. Theoretical analyses suggest that supergenes should readily emerge when there is coevolution between dispersal and social traits (e.g., mediated by variation in physiological or locomotor traits) because there will be selection for more benevolent individuals to preferentially interact with relatives (i.e., disperse less) and self-serving individuals to interact with nonrelatives (i.e., disperse more) [37]. As a result, linkage between loci responsible for dispersal and social behavior should be favored under a wide spectrum of conditions. In line with this prediction, dispersers across a diversity of organisms have been reported to differ in their social tendencies from nondispersers [37].

Gene Copy Number

CNVs appear to account for much of the genomic variation within many eukaryotic species [38]. Recent research on a plumage color polymorphism in the seabird the common murre (*Uria aalge*) demonstrated how a much more complex CNV (potentially mediated by TEs) of a 60-kb genomic region acts as a ‘supergene’ that maintains the association between three candidate genes affecting two traits – plumage color and thermal physiology – despite random mating [39]. Although it is not yet known whether this CNV also affects social behavior, it underscores how different structural attributes of the genome can be related to each other. There is, nevertheless, accumulating evidence that CNV may play a role in social behavior, including for humans in terms of social disorders [40] and for species with alternative reproductive tactics, where several individual traits can be functionally integrated to produce divergent reproductive phenotypes (reviewed in [4]). For example, in freshwater *Xiphophorus* swordtail fish, large males are deeper bodied, have higher dorsal fins, exhibit longer swords (extended rays of the caudal fin), and perform stereotyped courtship displays to entice females to mate, whereas small males are slender, have relatively smaller dorsal fins, lack prominent swords, and chase females to force copulation (reviewed in [41]). Lampert *et al.* [7] showed that variation in male size – and thus in reproductive tactic – in

Pseudogenization: neutral or adaptive outcome of gene duplication, when duplicated genes lose at least some functionality and may disappear entirely.

Recombination rate: frequency at which chromosomal segments recombine during meiosis.

Reproductive skew: the unequal partitioning of reproductive success within a population or social group.

Simple sequence repeat (SSR): repetitive DNA motifs (e.g., microsatellites); also known as short tandem repeat (STR).

Social behavior: interactions among two or more organisms where one individual affects the other, usually within the same species.

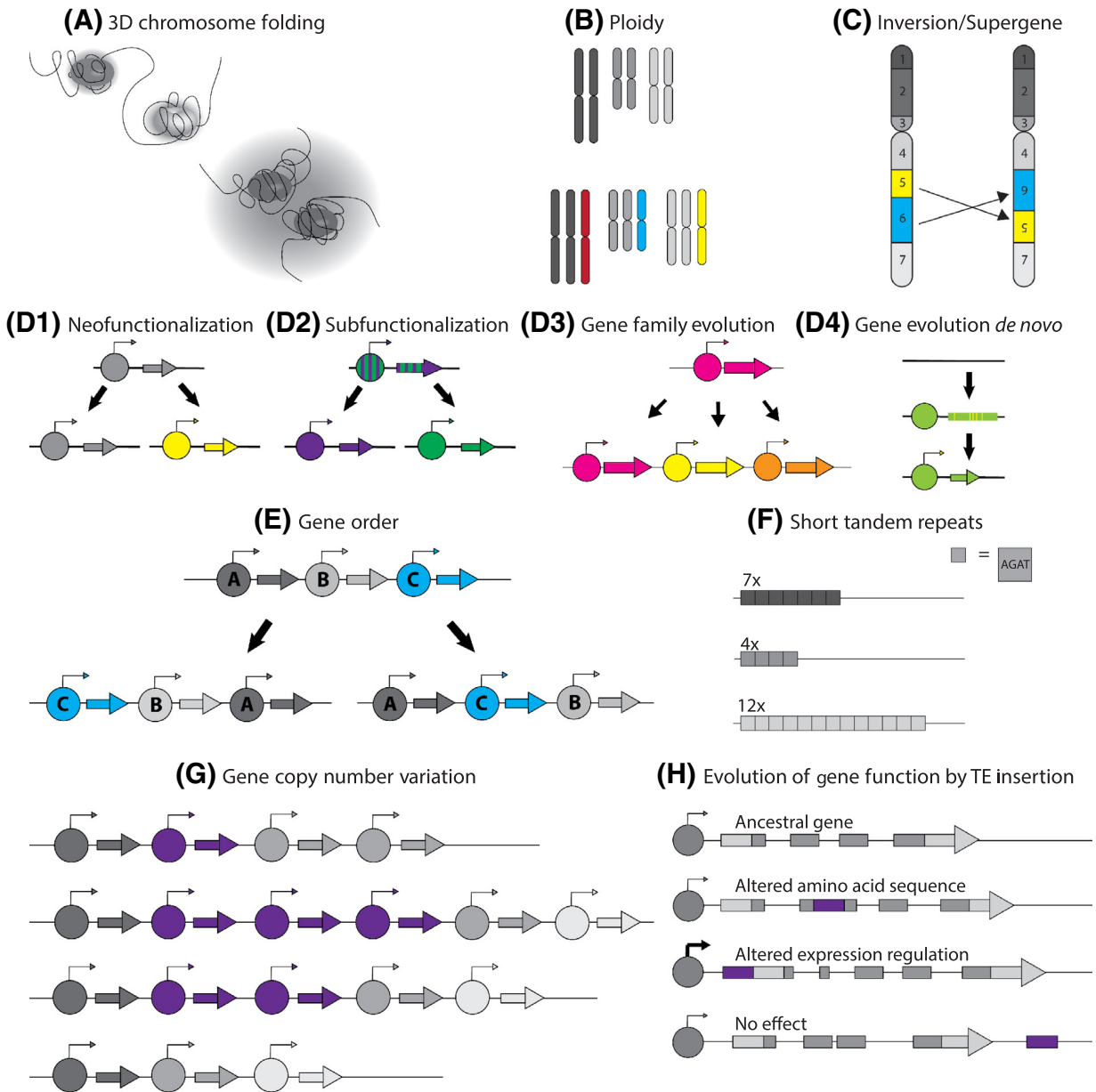
Social polymorphism: intraspecific variation in social or mating behavior.

Structural variation in the 3D genome: folding of chromatin in the 3D space of the nucleus.

Subfunctionalization: outcome of gene duplication when both paralogs retain different aspects of the ancestral gene’s function (neutral process).

Supergenes: clusters of tightly linked loci with related phenotypic effects, characterized by reduced recombination, often also involving a chromosomal inversion event.

Transposable elements (TEs): sequences of DNA that move from one location in the genome to another.



Trends in Ecology & Evolution

Figure 2. Attributes of Genome Architecture That Can Generate Phenotypic Variation across Individuals and Species. How these attributes affect social behavior is described in Table 1. (A) **Structural variation in the 3D genome** and (B) variation in ploidy are likely to affect all aspects of organismal function broadly. Variation in genome structure limited to specific chromosomes, such as (C) chromosomal inversions (which can result in supergenes) and (E) gene order, may have intermediate effects on trait variability. More local changes as a consequence of gene duplication events – such as (D1) neofunctionalization, (D2) subfunctionalization, (D3) evolution of gene families, and (G) gene copy number variations – or (D4) *de novo* evolution of genes are likely to have more specific phenotypic effects, depending on the extent of pleiotropy. Altered gene function due to (F) variation in the length of short tandem repeats or (H) random insertion of transposable elements (TEs) may affect phenotypes in more subtle ways. Importantly, all of these attributes can be sources of or be affected by variation in social behavior. Note that the two most common forms of genetic variation, SNPs and small insertions or deletions (indels), are not shown here, as their role in creating phenotypic (including behavioral) variation has been well studied.

Table 1. Attributes of Genome Architecture and Their Relation to Social Behavior

Structural attribute ^a	Definition	Behavioral example	Refs
Genome size (variation across species)	Total amount of DNA contained in one copy of a single genome	Social structure in <i>Synalpheus</i> shrimps Brain size in salamanders	[17,18,21]
Structural variation in the 3D genome	Folding of chromatin in the 3D space of the nucleus	Unknown	[88,89]
Ploidy (variation across species)	Number of complete sets of chromosomes in a cell	Sexual behavior in parthenogenetic whiptail lizards Calling behavior and call preference in frogs	[90–93]
Novel or orphan genes	Genes lacking detectable orthologs due to: (i) <i>de novo</i> origination; (ii) duplication and divergence; or (iii) horizontal gene transfer	Insertion of TEs in noncoding DNA resulted in <i>de novo</i> formation of <i>FLJ33706</i> gene involved in human cognitive function	[30,94]
Gene duplication: Neofunctionalization (adaptive) Subfunctionalization (neutral) Pseudogenization (adaptive or neutral) Dosage overcompensation (adaptive)	Divergence of paralogs after a gene duplication event	Neofunctionalization of Na ⁺ channel paralogs <i>Scn4aa</i> and <i>Scn4ab</i> in weakly electric fish	[45,95–97]
Gene family expansion (variation across species)	Proliferation of a set of several similar genes by duplication of a single original gene	Odorant receptor gene family in ants	[52,98]
CNV	Repeated sections of the genome, the number of which varies between individuals	CNV of <i>mc4r</i> gene in <i>Xiphophorus</i> swordtail fish affects age at maturity, body size, and reproductive tactic	[7,38]
Gene order	Organization of genes within a genome	Unknown	[99]
Supergenes	Clusters of tightly linked loci (with reduced recombination), often as a consequence of a chromosomal inversion	Sex chromosomes Social polymorphisms in ants Alternative mating forms in ruffs and white-crowned sparrows Subspecies variation in foraging behavior in honeybees	[5,6,25,32,33,100]
TEs	Sequences of DNA that move from one location in the genome to another	Move genes and accumulate on supergenes Exercise-dependent L1 retrotransposition in mouse hippocampus	[60,101]
STRs, SSRs (e.g., microsatellites)	Repetitive DNA motifs	Microsatellite length variation in prairie vole <i>V1aR</i> gene promoter affects stress-coping behavior and brain region-specific <i>V1aR</i> expression	[102,103]

^aUnless otherwise noted in parentheses, structural genomic attributes vary within a species.

X. nigrensis and *X. multilineatus* is correlated with Y-linked CNV of the 'B' allele of the *mc4r* gene, which encodes a melanocortin receptor subtype that binds α -melanocyte-stimulating hormone (α -MSH). Importantly, while *mc4r* CNV directly regulates growth and size at maturity, it controls behavioral tactics only indirectly [42].

Gene Duplication and Gene Family Expansion

Functional divergence of paralogs after gene duplication events has long been suggested to generate phenotypic variation and novel traits [43]. Possible outcomes of gene duplication include neofunctionalization, **subfunctionalization**, **pseudogenization**, and **dosage overcompensation**. Duplications of genes that encode ion channels have been critical to the evolution of phenotypic diversity in many taxa. A compelling example illustrating how gene duplication can influence the evolution of different social behaviors is the convergent evolution of two sodium

channel paralogs (*Scn4aa* and *Scn4ab*) in two divergent lineages of weakly electric fishes that have independently evolved electric organs, whose discharges are used for active sensing and communication [44]. Although *Scn4aa* is downregulated in the ancestors of both electric lineages, expression of this paralog has shifted from skeletal muscle to a novel communication function in the muscle-derived electric organ in a remarkable example of convergent evolution. Intriguingly, a similar shift in expression away from ancestral muscle cells to an evolutionarily novel muscle-derived sonic organ occurred for the same paralog in the sonic fish *Porichthys notatus*, in which males advertise their availability to females by characteristic sounds [45].

Another example of how gene duplication can influence social behavior is the *Malvolio* (*Mvl*) gene, a member of the natural resistance-associated macrophage protein (Nramp) family, which has been duplicated in *Nicrophorus vespilloides*, a subsocial beetle that exhibits advanced parental care behavior [46]. Although the two gene copies show differing expression profiles across many tissues, *Mvl1* increases in expression during resource preparation and feeding offspring whereas *Mvl2* decreases in the same states. This result suggests that following its duplication, *Mvl* has experienced sub- or neofunctionalization in *N. vespilloides* in ways relevant to the regulation of parental care behavior.

Mechanisms for recognition, the ability to distinguish friend from foe or kin from nonkin, are critical to most social interactions [47]. In many eusocial insects, for example, recognition is mediated via chemical signals that are detected and discriminated among by chemosensory receptors. The **odorant receptor (OR)** gene family, which arose soon after the evolution of life on land, encodes proteins that detect odorants with high specificity and sensitivity [48]. In Hymenoptera, eusocial species have several-fold more ORs than solitary ones [49]. In both bees [50] and ants [51], OR gene evolution is the result of lineage-specific gene duplications that underwent independent expansions. In ants in particular, where all of the more than 13 000 species are eusocial, a single lineage of the OR gene family underwent a massive expansion in the ancestor of all extant species [52], resulting in up to nearly 400 putative functional OR genes in some ants [51,53], largely due to tandem array expansion [54]. Similar dynamic expansions and contractions are seen among olfactory receptor gene families in vertebrates [55], although they have not yet been examined in relation to social behavior *per se*.

Avenues for Future Research in an Age of Comparative Genomics

As our ability to sequence large numbers of individuals rapidly increases, researchers will inevitably move beyond studies of transcriptomic variation to those of genomic variation both within and across species. In particular, complete genome sequencing of multiple individuals within a single species, coupled with population-genetic analyses, will enable researchers to identify the genomic architectural polymorphisms that give rise to intraspecific variation in social behavior. At the interspecific level, newly developed phylogenetic comparative methods, along with emerging long-read sequencing technologies, will also allow researchers to identify key architectural features of the genome that are shared among species with similar social or mating systems and to examine not just cooperation but also conflict in the genome (Box 1). Below we discuss three areas ripe for future study.

Linking Transposable Elements and Regulatory Variation in Genome Architecture in Social Species

One potentially powerful source of variation in regulatory relationships is the movement of TEs. First described by Barbara McClintock [56] as 'controlling elements', TEs are DNA sequences that move from one location in the genome to another and in this way provide a continuous source of genetic innovation. By inserting within genes or their regulatory elements, for example, TEs can directly disrupt gene transcription [57,58]. TEs may become exapted into novel

regulatory elements, thus shaping the evolution of gene regulatory networks [59,60], or spread sequences readily modified by epigenetic marks to neighboring protein-coding genes and repress gene expression [61]. There is also evidence in support of the hypothesis that regulatory features can be derived from TEs (e.g., miRNAs [62,63], transcription factor-binding sites [64, 65]), although to date few of these have been explicitly linked to novel behavioral phenotypes, possibly due to an ascertainment bias. One exception comes from a study of ORs in ants. Although TE islands were depauperate in ORs of the clonal raider ant *Ooceraea biroi*, these ORs were enriched for TEs relative to the rest of the genome, potentially facilitating gene duplication by increasing unequal crossing over, promoting ectopic recombination, or via retrotransposition [54]. Ultimately, understanding how TEs might affect behavioral phenotypes will require high-quality genomes, which are often still lacking for nontraditional model organisms, and studies of their possible effects on gene expression rather than simple descriptions of their abundance, type, and distribution.

Linking Demography and Structural Variation in Genome Architecture in Social Species

Intraspecific variation in social behavior may be caused by plasticity and/or genetic variation maintained by balancing selection. In the case of complex (polygenic) behavioral traits, we can hypothesize that selection acts through gene linkage relationships or gene regulation. The tension between recombination and balancing selection on linkage relationships favors spatially integrated architectures, which can arise through chromosomal rearrangements that move behaviorally relevant loci into close proximity on a chromosome or inversions that reduce the rate of recombination [66]. As noted previously, chromosomal inversions can contain numerous genes underlying behavioral, morphological, and physiological traits that are all associated with social polymorphisms. Similarly, we hypothesize that rearrangements that move coexpressed loci into close proximity on a chromosome can be adaptive, if gene expression can be more efficiently or precisely coordinated this way.

If selection favors such rearrangement mutations, the rate at which they accumulate should vary as a function of the effective population size N_e , the frequency μ at which such mutations occur,

Box 1. From Social to Genomic Conflict

Social behavior is hierarchical in its organization: genes cooperate in genomes, genomes in eukaryotic cells, cells in multicellular organisms, and organisms in societies. The major transitions framework [79,80] envisions this hierarchy to be a product of independently replicating units giving up their own reproduction to form a new level of individuality, while overcoming individual selfish interests [81]. However, conflict still exists, not just at the level of society but also at the level of the genome, where the proliferation of selfish genetic elements could be either a cause or a consequence of social behavior [82].

Since the evolutionary transition to higher levels of biological organization often requires the suppression of conflict at lower levels, we might expect less genomic conflict in lineages that have undergone more transitions. Consistent with this hypothesis, TE abundance and diversity are negatively correlated with the evolution of eusociality in bees [74]. Whether eusociality can generally be expected to provide protection against genomic conflicts, however, is less clear. For example, increases in social complexity leading to a reproductive division of labor and skew may also be associated with a reduction in N_e , which population genetic theory predicts should lead to the proliferation of TEs [15]. In inbred species of ants and wasps, similar patterns of TEs suggest that shared life history strategies in these social species affect N_e and consequently genome architecture [83]. Furthermore, several other aspects of social behavior, including the evolution of self-fertilization, may reduce N_e and affect both the transmission of TEs and the efficacy of selection acting against them [84].

Shifts in mating behavior may also create the circumstances for new kinds of genomic conflict. For example, since kin selection theory predicts that lifetime monogamy prevents the evolution of **genomic imprinting** [85], shifts from monogamy to a situation where females mate with multiple males (i.e., polyandry) may therefore lead to the evolution of genomic imprinting. The observation of higher levels of imprinting in outcrossing compared with self-fertilizing *Arabidopsis* species [86] supports this prediction. Finally, the spread of selfish genetic elements may also play a causal role in changing social behavior at the organismal level. For example, the spread of a deleterious sex ratio-distorting gene in *Drosophila pseudoobscura* has resulted in the evolution of higher female re-mating rates [87].

and the realized strength s of selection on the rearrangement [15]. Although N_e can be readily estimated from genomic data, the variables μ and s are often unknown. In the case of selection on linkage, the benefit that accrues from a given rearrangement will be proportional to the reduction in recombination and the alleviation of fitness costs that would have resulted in recombinant offspring [67]. In the case of selection on gene regulation, fitness benefits would depend on the regulatory interactions of the genes involved and resulting changes in energetic efficiency. Any benefits would also be offset by costs associated with the rearrangement due to disruption of existing genes or regulatory machinery. While accumulation rates have been measured for different kinds of rearrangements (e.g., [68]), their rates of occurrence are largely unknown, although they may correlate with overall TE abundance [69]. Thus, we hypothesize that species with either a suitable type of TE or a high abundance of TEs (e.g., as a consequence of reduced N_e in social species, in species where N_e fluctuates over time) will have more flexible genomes and can therefore evolve more readily to optimize **gene order** in response to selection on linkage or gene expression regulation. Such increased evolvability would, in turn, allow species to more readily adapt to novel or changing environments [70,71]. More generally, we predict tightly linked gene clusters (e.g., supergenes) to occur in species with: (i) a long history of sociality, where polymorphisms are maintained by strong frequency-dependent selection or migration–selection balance; (ii) abundant and/or mobile TEs; (iii) and traits determined by tandemly duplicated (and thus already tightly linked) genes.

Identifying Mechanisms of Behavioral Plasticity in the Genomes of Social Species

The opportunities and constraints imposed by social structure are likely to influence the degree of intraspecific variation observed among individuals within populations and the genomic mechanisms underlying this plasticity. Social living can lead to increases in behavioral plasticity by releasing individual group members from certain behavioral and ecological constraints, while also adding a social component to genotype-by-environment ($G \times E$) interactions. For example, reproductive development in many animals is influenced by nutritional feedback on endocrine pathways, which in social insects is typically modulated by pheromones and dominance interactions [72]. In species showing rudimentary forms of eusociality (e.g., many paper wasps and sweat bees), workers are not only totipotent and maintain reproductive capabilities, but also more plastic in that they can dramatically change their roles in society (e.g., switching castes) in response to cues from the social environment [73]. Recent studies suggest that the early stages of eusociality, and the associated increase in individual plasticity, are shaped by genetic variants [11], a high diversity and abundance of TEs, and changes in the regulatory architecture of key gene networks [74,75]. As social complexity increases further, reproductive skew intensifies as fewer and fewer individuals reproduce (due to division of labor). We hypothesize that the resulting small N_e will cause changes in TE content, which in turn will facilitate further evolution of genome structure. It is intriguing in this context that in bumblebees (*Bombus* spp.), which are highly socially plastic, genes experiencing positive selection are predominately expressed by reproductive foundresses, whereas in honeybees, *Apis mellifera*, worker-biased genes are more likely to experience adaptive evolution, suggesting that the loss of queen totipotency may change how selection acts on the social insect genome [76].

Concluding Remarks

Social behavior can affect selection regimes and demography, which in turn influence genome evolution in important and underappreciated ways. We have illustrated some of the bidirectional links and feedbacks between social behavior and genome architecture, emphasizing the role that population genetics can play in understanding social evolution. Because social traits often emerge from a suite of less complex correlated behavioral patterns or phenotypic traits [1], genetic relationships underlying behaviors may emerge less frequently than other, simpler traits.

Outstanding Questions

Do genomic polymorphisms that give rise to intraspecific variation in genome architecture have properties that are shared across disparate taxa?

What are the key architectural features of the genome that are shared among species with similar social or mating systems?

How does demography affect structural variation in genome architecture in social species?

What is the null expectation for the evolutionary dynamics associated with genome architecture in social species?

How do transposable elements contribute to regulatory variation in genome architecture in social species?

What are the genomic mechanisms of behavioral plasticity in social animals?

Does the dynamic nature of the genome differ between social and nonsocial species?

Conversely, because numerous genes contribute to complex social behavior, they may have profound impacts on genome architecture. For example, the accumulation of traits related to social polymorphisms on chromosomal inversions in the form of supergenes may be a common, but underappreciated, mechanism underlying alternative reproductive strategies and driving differences in social structures.

As genomic tools and resources become more readily accessible, opportunities to study the co-evolution of social behavior and genome architecture abound. Moreover, advances in digital evolution approaches provide novel avenues for *in silico* testing of hypotheses about the causal relationship between sociality and genome structure [77]. However, we caution that a better general understanding of the evolution of genome architecture is necessary. Vertebrate genomes, for example, have been dynamic over evolutionary time, often the result of TE expansion and deletion [78]. Moreover, it remains to be seen what the null expectation is for the evolutionary dynamics associated with genome architecture in social species. If demographic changes related to social behavior influence genome architecture, comparing social and nonsocial species with similar demographic histories might help to isolate the role of social behavior on genome evolution and further define null expectations for future studies. Ultimately, by simultaneously considering both the genomic causes and the consequences of behavioral variation within and across species, we will generate a richer and more integrated view of social evolution (see Outstanding Questions).

Acknowledgments

This work is the result of a symposium and workshop on *Social Evolution and Genome Complexity* held at Columbia University in March 2017. The event was organized by D.R.R. and H.A.H. with support from Columbia University Seminars and the Research Coordination Network in Sociogenomics, which was funded by the US National Science Foundation (IOS-1256839). We thank Dr W. Wilczynski for his generous support and Nicole Elmer for assistance with the preparation of Figure 2.

References

- Hofmann, H.A. *et al.* (2014) An evolutionary framework for studying mechanisms of social behavior. *Trends Ecol. Evol.* 29, 581–589
- Foster, S.A. (2013) Evolution of behavioural phenotypes: influences of ancestry and expression. *Anim. Behav.* 85, 1061–1075
- Ross, K.G. and Keller, L. (1995) Ecology and evolution of social organization: insights from fire ants and other highly eusocial insects. *Annu. Rev. Ecol. Syst.* 26, 631–656
- Oliveira, R.F. *et al.* (2008) *Alternative Reproductive Tactics: An Integrative Approach*, Cambridge University Press
- Wang, J. *et al.* (2013) A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493, 664–668
- Purcell, J. *et al.* (2014) Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* 24, 2728–2732
- Lampert, K.P. *et al.* (2010) Determination of onset of sexual maturation and mating behavior by melanocortin receptor 4 polymorphisms. *Curr. Biol.* 20, 1729–1734
- Chau, L.M. and Goodisman, M.A.D. (2017) Gene duplication and the evolution of phenotypic diversity in insect societies. *Evolution* 71, 2871–2884
- Samaco, R.C. *et al.* (2012) Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nat. Genet.* 44, 206–211
- Bendesky, A. *et al.* (2017) The genetic basis of parental care evolution in monogamous mice. *Nature* 544, 434–439
- Kocher, S.D. *et al.* (2018) The genetic basis of a social polymorphism in halictid bees. *Nat. Commun.* 9, 4338
- Young, R.L. *et al.* (2019) Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1331–1336
- Rubin, B.E.R. and Moreau, C.S. (2016) Comparative genomics reveals convergent rates of evolution in ant–plant mutualisms. *Nat. Commun.* 7, 12679
- Romiguier, J. *et al.* (2014) Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J. Evol. Biol.* 27, 593–603
- Lynch, M. (2007) *The Origins of Genome Architecture*, Sinauer Associates
- Elliott, T.A. and Gregory, T.R. (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 370, 20140331
- Roth, G. *et al.* (1997) Genome size, secondary simplification, and the evolution of the brain in salamanders. *Brain Behav. Evol.* 50, 50–59
- Sun, C. *et al.* (2012) Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol. Evol.* 4, 1340–1348
- Petrov, D.A. *et al.* (2000) Evidence for DNA loss as a determinant of genome size. *Science* 287, 1060–1062
- Lynch, M. (2006) The origins of eukaryotic gene structure. *Mol. Biol. Evol.* 23, 450–468
- Jeffery, N.W. *et al.* (2016) Patterns of genome size variation in snapping shrimp. *Genome* 59, 393–402
- Chan, Y.F. *et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327, 302–305
- Xie, K.T. *et al.* (2019) DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* 363, 81–84
- Corona, M. *et al.* (2013) Vitellogenin underwent subfunctionalization to acquire caste and behavioral specific expression in the harvester ant *Pogonomyrmex barbatus*. *PLoS Genet.* 9, e1003730
- Schwander, T. *et al.* (2014) Supergenes and complex phenotypes. *Curr. Biol.* 24, R288–R294
- Mathiopoulos, K.D. *et al.* (1999) Are chromosomal inversions induced by transposable elements? A paradigm from the malaria mosquito *Anopheles gambiae*. *Parassitologia* 41, 119–123

27. Gray, Y.H. (2000) It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* 16, 461–468
28. Branco, S. *et al.* (2018) Multiple convergent supergene evolution events in mating-type chromosomes. *Nat. Commun.* 9, 2000
29. Rubenstein, D.R., Abbot, P., eds (2017) *Comparative Social Evolution*, Cambridge University Press
30. Li, C.-Y. *et al.* (2010) A human-specific *de novo* protein-coding gene associated with human brain functions. *PLoS Comput. Biol.* 6, e1000734
31. Roberts, R.B. *et al.* (2009) Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* 326, 998–1001
32. Küpper, C. *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* 48, 79–83
33. Lamichhaney, S. *et al.* (2016) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* 48, 84–88
34. Tuttle, E.M. *et al.* (2016) Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* 26, 344–350
35. Day, T.H. *et al.* (1983) A chromosomal inversion polymorphism in Scandinavian populations of the seaweed fly, *Coelopa frigida*. *Hereditas* 99, 135–145
36. Kunte, K. *et al.* (2014) *doublesex* is a mimicry supergene. *Nature* 507, 229–232
37. Mullon, C. *et al.* (2018) Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nat. Ecol. Evol.* 2, 132–140
38. Sudmant, P.H. *et al.* (2013) Evolution and diversity of copy number variation in the great ape lineage. *Genome Res.* 23, 1373–1382
39. Tigano, A. *et al.* (2018) A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird. *bioRxiv* Published online December 28, 2018. <https://doi.org/10.1101/507384>
40. Rao, S. *et al.* (2019) Genome-wide copy number variation-, validation- and screening study implicates a novel copy number polymorphism associated with suicide attempts in major depressive disorder. *bioRxiv* Published online January 30, 2019. <https://doi.org/10.1101/534909>
41. Ryan, M.J. and Rosenthal, G. (2011) Variation and selection in swordtails. In *Model Systems in Behavioral Ecology: Integrating Conceptual, Theoretical and Empirical Approaches* (Dugatkin, L., ed.), pp. 133–148, Princeton University Press
42. Smith, C.C. *et al.* (2015) Copy number variation in the melanocortin 4 receptor gene and alternative reproductive tactics of the swordtail *Xiphophorus multilineatus*. *Environ. Biol. Fish* 98, 23–33
43. Ohno, S. (1970) *Evolution by Gene Duplication*, Allen & Unwin
44. Zakon, H.H. (2012) Adaptive evolution of voltage-gated sodium channels: the first 800 million years. *Proc. Natl. Acad. Sci. U. S. A.* 109, 10619–10625
45. Thompson, A. *et al.* (2014) Expression evolution facilitated the convergent neofunctionalization of a sodium channel gene. *Mol. Biol. Evol.* 31, 1941–1955
46. Mehlferber, E.C. *et al.* (2017) Duplication and sub/neofunctionalization of *Malvolio*, an insect homolog of *Nramp*, in the subsocial beetle *Nicrophorus vespilloides*. *G3 (Bethesda)* 7, 3393–3403
47. Waldman, B. (1988) The ecology of kin recognition. *Annu. Rev. Ecol. Syst.* 19, 543–571
48. Brand, P. *et al.* (2018) The origin of the odorant receptor gene family in insects. *eLife* 7, e38340
49. LeBoeuf, A.C. *et al.* (2013) The molecular basis of social behavior: models, methods and advances. *Curr. Opin. Neurobiol.* 23, 3–10
50. Brand, P. and Ramirez, S.R. (2017) The evolutionary dynamics of the odorant receptor gene family in corbiculate bees. *Genome Biol. Evol.* 9, 2023–2036
51. Engsontia, P. *et al.* (2015) Diversification of the ant odorant receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. *BMC Res. Notes* 8, 380
52. McKenzie, S.K. *et al.* (2016) Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14091–14096
53. Tribble, W. *et al.* (2017) *orco* mutagenesis causes loss of antennal lobe glomeruli and impaired social behavior in ants. *Cell* 170, 727–735.e10
54. McKenzie, S.K. and Kronauer, D.J.C. (2018) The genomic architecture and molecular evolution of ant odorant receptors. *Genome Res.* 28, 1757–1765
55. Bear, D.M. *et al.* (2016) The evolving neural and genetic architecture of vertebrate olfaction. *Curr. Biol.* 26, R1039–R1049
56. McClintock, B. (1956) Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216
57. Han, J.S. *et al.* (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 429, 268–274
58. Wheelan, S.J. *et al.* (2005) Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* 15, 1073–1078
59. Warnefors, M. *et al.* (2010) Transposable elements: insertion pattern and impact on gene expression evolution in hominids. *Mol. Biol. Evol.* 27, 1955–1962
60. Chuong, E.B. *et al.* (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86
61. Rebollo, R. *et al.* (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* 7, e1002301
62. Roberts, J.T. *et al.* (2014) Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences. *Mob. Genet. Elem.* 4, e29255
63. Piriyaopongsa, J. *et al.* (2007) Origin and evolution of human microRNAs from transposable elements. *Genetics* 176, 1323–1337
64. Rebollo, R. *et al.* (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* 46, 21–42
65. Hoen, D.R. and Bureau, T.E. (2015) Discovery of novel genes derived from transposable elements using integrative genomic analysis. *Mol. Biol. Evol.* 32, 1487–1506
66. Yeaman, S. (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci. U. S. A.* 110, E1743–E1751
67. Lenormand, T. and Otto, S.P. (2000) The evolution of recombination in a heterogeneous environment. *Genetics* 156, 423–438
68. Ranz, J.M. *et al.* (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Res.* 11, 230–239
69. Raffaele, S. and Kamoun, S. (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430
70. Sun, S.-J. *et al.* (2014) Climate-mediated cooperation promotes niche expansion in burying beetles. *eLife* 3, e02440
71. Cornwallis, C.K. *et al.* (2017) Cooperation facilitates the colonization of harsh environments. *Nat. Ecol. Evol.* 1, 57
72. Amsalem, E. *et al.* (2015) *The Physiological and Genomic Bases of Bumble Bee Social Behaviour*, pp. 37–93, Academic Press
73. Patalano, S. *et al.* (2015) Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci. U. S. A.* 112, 13970–13975
74. Kapheim, K.M. *et al.* (2015) Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348, 1139–1143
75. Simola, D.F. *et al.* (2013) Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23, 1235–1247
76. Harpur, B.A. *et al.* (2017) Queens and workers contribute differently to adaptive evolution in bumble bees and honey bees. *Genome Biol. Evol.* 9, 2395–2402
77. Adami, C. (2006) Digital genetics: unravelling the genetic basis of evolution. *Nat. Rev. Genet.* 7, 109–118

78. Kapusta, A. *et al.* (2017) Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U. S. A.* 114, E1460–E1469
79. Bourke, A. (2011) *Principles of Social Evolution*, Oxford University Press
80. Maynard Smith, J. and Szathmáry, E. (1997) *The Major Transitions in Evolution*, Oxford University Press
81. West, S.A. *et al.* (2015) Major evolutionary transitions in individuality. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10112–10119
82. Koonin, E.V. (2016) Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 371, 20150442
83. Schrader, L. *et al.* (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Commun.* 5, 5495
84. Ågren, J.A. *et al.* (2014) Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* 15, 602
85. Patten, M.M. *et al.* (2014) The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity (Edinb.)* 113, 119–128
86. Klosinska, M. *et al.* (2016) Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat. Plants* 2, 16145
87. Price, T.A.R. *et al.* (2008) Selfish genetic elements promote polyandry in a fly. *Science* 322, 1241–1243
88. Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293
89. Spielmann, M. *et al.* (2018) Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453–467
90. Van de Peer, Y. *et al.* (2017) The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424
91. Woolley, S.C. *et al.* (2004) Tracing the evolution of brain and behavior using two related species of whiptail lizards: *Cnemidophorus uniparens* and *Cnemidophorus inornatus*. *ILAR J.* 45, 46–53
92. Tucker, M.A. and Gerhardt, H.C. (2012) Parallel changes in mate-attracting calls and female preferences in autotriploid tree frogs. *Proc. Biol. Sci.* 279, 1583–1587
93. Evans, B.J. *et al.* (2015) Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (*Xenopus*, Pipidae) from West and Central Africa. *PLoS One* 10, e0142823
94. Tautz, D. and Domazet-Lošo, T. (2011) The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12, 692–702
95. Grant, S.G.N. (2016) The molecular evolution of the vertebrate behavioural repertoire. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 371, 20150051
96. Lee, H. *et al.* (2016) Effects of gene dose, chromatin, and network topology on expression in *Drosophila melanogaster*. *PLoS Genet.* 12, e1006295
97. Sun, D. *et al.* (2018) Rapid regulatory evolution of a nonrecombining autosome linked to divergent behavioral phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2794–2799
98. Demuth, J.P. and Hahn, M.W. (2009) The life and death of gene families. *Bioessays* 31, 29–39
99. Hurst, L.D. *et al.* (2004) The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310
100. Wallberg, A. *et al.* (2017) Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genet.* 13, e1006792
101. Muotri, A.R. *et al.* (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* 19, 1002–1007
102. Richard, G.-F. *et al.* (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72, 686–727
103. Donaldson, Z.R. and Young, L.J. (2013) The relative contribution of proximal 5' flanking sequence and microsatellite variation on brain vasopressin 1a receptor (*Avpr1a*) gene expression and behavior. *PLoS Genet.* 9, e1003729