Introduction

Author for correspondence:
Emma L. Berdan
e-mail: emma.berdan@gmail.com

Genomic architecture of supergenes: connecting form and function

Emma L. Berdan[1,2,3], Thomas Flatt[4], Genevieve M. Kozak[5], Katie E. Lotterhos[6] and Ben Wielstra[1,2]

[1]Institute of Biology Leiden, Leiden University, PO Box 9505, 2300 RA, Leiden, The Netherlands
[2]Naturalis Biodiversity Center, PO Box 9517, 2300 RA Leiden, The Netherlands
[3]Tjärnö Marine Laboratory, Department of Marine Sciences, University of Gothenburg, 45296 Strömstad, Sweden
[4]Department of Biology, University of Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland
[5]Department of Biology, University of Massachusetts Dartmouth, 285 Old Westport Road, MA 02747, USA
[6]Department of Marine and Environmental Sciences, Northeastern University, 430 Nahant Road, Nahant, MA 01908, USA

ELB, 0000-0002-6435-4604; TF, 0000-0002-5990-1503; GMK, 0000-0001-6413-1403; KEL, 0000-0001-7529-2771; BW, 0000-0002-7112-5965

Supergenes are tightly linked sets of loci that are inherited together and control complex phenotypes. While classical supergenes—governing traits such as wing patterns in *Heliconius* butterflies or heterostyly in *Primula*—have been studied since the Modern Synthesis, we still understand very little about how they evolve and persist in nature. The genetic architecture of supergenes is a critical factor affecting their evolutionary fate, as it can change key parameters such as recombination rate and effective population size, potentially redirecting molecular evolution of the supergene in addition to the surrounding genomic region. To understand supergene evolution, we must link genomic architecture with evolutionary patterns and processes. This is now becoming possible with recent advances in sequencing technology and powerful forward computer simulations. The present theme issue brings together theoretical and empirical papers, as well as opinion and synthesis papers, which showcase the architectural diversity of supergenes and connect this to critical processes in supergene evolution, such as polymorphism maintenance and mutation accumulation. Here, we summarize those insights to highlight new ideas and methods that illuminate the path forward for the study of supergenes in nature.

This article is part of the theme issue 'Genomic architecture of supergenes: causes and evolutionary consequences'.

1. The evolutionary puzzle of supergenes

A supergene is a set of tightly linked loci affecting a complex phenotype, with recombination between at least two different combinations of alleles (referred to as haplotypes hereafter) being reduced or absent [1–4]. As a consequence, these sets of alleles are inherited together, as if they were a single Mendelian locus (thus explaining the term 'supergene') [5]. Such an architecture can be beneficial when complex multi-trait phenotypes are under divergent selection in the presence of gene flow so that recombination opposes adaptation and speciation by homogenizing critical allele combinations [6–9]. The concept of a supergene is closely related to Dobzhansky's concept of 'coadapted gene complexes', i.e. epistatic fitness interactions among loci maintained by reduced recombination, which he applied to chromosomal inversion polymorphisms [10–13]. The reduced recombination in the genomic region of the supergene splits the evolutionary trajectory of the region into at least two semi-independent branches [2,14] allowing for distinct multi-trait phenotypes ('polymorphisms': [15,16]) to segregate within a single population or to be easily transmitted across

species. Although the potential evolutionary benefit of such a scenario was noted nearly a century ago [12,17], disentangling the complexities of supergene origin and evolution has remained a challenge.

New genomic, theoretical and bioinformatic methods now allow us to begin to understand the complex evolutionary history and diversity of supergenes [1,3,18,19]. In particular, the genomic revolution has highlighted a surprising amount of variation in the genomic architecture of supergenes. Genomic architecture plays a critical role in the continued evolution of supergenes as it dictates two of their key features: recombination rate and selective pressures on the supergene. First, large structural variants (SVs; e.g. chromosomal inversions, chromosomal fusions and indels), which are common supergene architectures, alter the effective recombination rate. However, different SVs vary in the extent and pattern of recombination suppression [20]. Second, selection can act on supergenes at three different scales, depending on genetic architecture: (i) there can be direct selection on the supergene itself, if potential breakpoints or deletions disrupt genes, create new genes [21] or separate loci from their regulatory elements; (ii) the allelic content of the haplotypes is also under selection, generating indirect selection at the level of the supergene through linkage disequilibrium; and (iii) there can be indirect selection on the supergene owing to the reduction in effective recombination between supergene haplotypes (i.e. selection because of a recombination load; [8,10,22,23]). Determining the relationship between genomic architecture and the evolution of supergenes is imperative for explaining and predicting supergene evolution and its consequences in natural populations [18].

The present collection of papers integrates (i) theoretical (including simulation) studies that generate predictions about supergene evolution, (ii) empirical studies that dissect genomic variation associated with supergenes, and (iii) synthetic opinion and review papers that tie together these diverse aspects of supergene evolution. Papers in this theme issue showcase the stunning architectural and taxonomic diversity of supergenes and cover a broad range of subjects relevant to our understanding of supergenes including local adaptation versus gene flow; parallel evolution; mutation accumulation and balanced lethality; sexually antagonistic selection; sex chromosome evolution; speciation; chromatin architecture and gene expression; and selfish genetic elements. In this introduction, we review the advances made by these papers, summarize the genomic architecture they uncover (§2) and detail how it connects to three major questions in supergene research: the role of supergenes in adaptation (§3), maintenance of supergene polymorphism (§4), and mutation accumulation in supergenes (§5). We highlight methodologies that can move the field forward and focus on outstanding questions in supergene evolution.

2. Genomic architecture of supergenes

Characterizing the genomic architecture of supergenes is the first step towards understanding their evolution. A major focal point is to comprehend the underlying genetic mechanisms of recombination suppression between supergene haplotypes. Certain genetic mechanisms can generate unbalanced gametes in supergene heterozygotes (e.g. chromosomal inversions or translocations), which has the two-pronged effect of generating underdominance as well as reducing

effective recombination. For example, improper segregation of chromosomes in translocation heterozygotes can lead to the creation of aneuploid gametes (lacking critical genes) at a rate of 18–80% [24,25]. The extent of the reduction in effective recombination can also evolve over the lifetime of a supergene, through expansion or accumulation of additional SVs. Reconstructing the structural changes that generate recombination suppression is now possible using a combination of long-read sequencing and comparative genomics. Several papers in this issue dissect the genomic architecture of different supergenes, ranging from supergenes controlling colour polymorphism [21,26] to those contributing to female meiotic drive [27], in order to uncover the sources of recombination suppression.

Kim *et al.* [26] use trio-binning and long-read sequencing to construct chromosomal level assemblies of the BC supergene, associated with colour polymorphism in the African monarch butterfly, *Danaus chrysippus*. They find three alleles that differ dramatically in sequence, each containing multiple duplicated regions and inversions. Comparing BC supergene structure across the *Danaus* phylogeny, the authors reconstruct this supergene's evolutionary history and infer that it initially arose when a large genomic region was repeatedly duplicated approximately 7.5 Myr, with several inversions arising subsequently within this region of duplicated segments. Recombination suppression across the supergene probably spread through the fixation of these inversions, in addition to the genetic divergence of duplicated segments followed by differential loss of some of these duplicated regions. This second process of divergence and loss of duplicates may be a common contributor to supergene evolution.

Komata *et al.* [21] also use phylogenomic methods to understand the genomic architecture of supergenes. They focus on a supergene found in multiple species of *Papilio* butterflies that controls female-limited mimicry [28,29]. Two differentiated haplotypes of an autosomal genomic region (termed H and h) are associated with the mimetic and non-mimetic morphs, respectively. While the genomic architecture of H is known to include an inversion in *Papilio polytes*, no inversion is present in H in *Papilio memnon* [28,30,31]. Thus, this system is ideal for investigating the role of different genomic architectures in shaping supergene evolution, even when they lead to the same phenotype. Komata *et al.* [21] put together data from whole-genome resequencing studies, functional analyses and expression studies, to better understand the genomic architecture of the supergene in *P. memnon* and *P. polytes*. Using knockdown and expression studies, the authors are able to directly link genes within the supergene to different aspects of the female wing colour patterns. The authors show that both systems exhibit strong linkage disequilibrium in the supergene region as well as accumulation of repetitive sequences, a hallmark of low recombination regions [32–36]. Surprisingly, this indicates that the presence of an inversion may not have drastically changed the recombination rate in the supergene. Instead, the inversion in *P. polytes* seems to have generated a new gene, *U3X*, that regulates the expression of other loci within the supergene. Overall, this suggests that selection for the inversion may be acting on the breakpoints (i.e. direct selection on the inversion itself), as opposed to other characteristics such as reduced recombination.

Not all supergenes show suppressed recombination between haplotypes. Dufresnes *et al.* [37] investigate sex chromosomes that lack recombination suppression in the heterogametic sex. In most species, archetypal sex chromosomes come in two

differentiated versions that do not recombine over a large stretch in the heterogametic sex (the male in XY systems such as mammals or the female in WZ systems such as birds). Because of this strong reduction in recombination, sex chromosomes are predicted to accumulate mutations at a faster rate than autosomes. This may cause sex chromosomes to behave like 'supergenes of speciation', with the non-recombining section accumulating genes responsible for reproductive isolation [38]. In fact, two rules of speciation—Haldane's rule (heterogametic hybrids show relatively higher sterility/inviability) and the large X-/Z-effect (X/Z chromosomes excessively accumulate incompatibilities relative to autosomes)—are based on the idea that sex chromosomes accumulate mutations. However, sex determination does not always follow the simple rules outlined above [39]. Amphibians lack the genomic architecture that suppresses recombination between the sex chromosomes. Dufresnes *et al.* [37] review this amphibian defiance of the classical laws of sex chromosome evolution and examine what it can teach us about the role of sex chromosomes as supergenes of speciation. The generally homomorphic sex chromosomes of amphibians do not appear to play a disproportionally large role in reproductive isolation as compared to the autosomes. These results are in contrast with support of both Haldane's rule and the large X-effect in insects, mammals, and fishes (who also have chromosomes without large-scale recombination suppression) [40]. Dufresnes *et al.* [37] highlight amphibians as a promising model to contrast with systems where sex chromosomes act as supergenes of speciation.

In addition to affecting linkage between loci, supergenes may also go beyond DNA changes to alter the epigenome, i.e. chromatin and methylation patterns. The ensuing consequences of supergenes on epigenomic modifications have been little explored up to now. In their paper, Wright & Schaeffer [41] explore the question of whether and how chromosomal rearrangements such as inversions might alter chromatin organization and modify gene expression; effects that might cause selection against these SVs. They map the breakpoints of seven pairs of inversion breakpoints in the fly *Drosophila pseudoobscura* to a map of topologically associated domains (TADs). TADs, which represent a form of higher order chromatin interactions, are self-interacting regions of the genome hypothesized to regulate gene expression [42–45]. Wright & Schaeffer [41] examine whether persisting inversion breakpoints are more likely to occur at boundaries between TADs and find that, indeed, breakpoints occur at TAD boundaries more often than expected by chance. Their study thus suggests that some inversion breakpoints might affect gene expression within TADs; this lends support to the hypothesis that position effects can contribute to the establishment of inversions.

Supergenes may also be selfish genetic elements, containing alleles that are specialized in over-representing themselves in the next generation through segregation distortion [46,47]. There are many well-known selfish supergenes, including *t*-haplotypes in mice [48], the *Sp* haplotype in the Alpine silver ant, *Formica selysi* [49] and *Segregation distorter* in *Drosophila melanogaster* [50]. Finseth *et al.* [27] provide evidence that the previously identified selfish centromere (D) in *Mimulus guttatus* is a supergene, segregating in multiple *Mimulus* populations in the Pacific northwest. They show that in several populations, D is a large (10–12 Mb) non-recombining region that leads to female meiotic drive and contributes to pollen inviability when homozygous (female-driving). D shows evidence of prior selective sweeps (possibly separate sweeps in two populations), but appears to be maintained within populations as a balanced polymorphism. Transcriptomes of individuals with and without the D allele indicate that many genes located within the supergene show reduction of expression in developing and reproductive tissues, suggesting many of the effects of D are *cis*-acting, similar to findings for other non-driving supergenes [51–53]. Finseth *et al.* [27] identify a number of potential linked candidate genes differing in expression that may lead to meiotic drive, including a centromere chaperone (NASP^{SIM3}), further contributing to our knowledge of selfish supergenes.

3. Supergenes facilitate adaptation

Supergenes provide a mechanism that allows multiple favourable trait combinations to be maintained in the face of recombination. While the role of supergenes in adaptation has been well established [3–5,7,8,18,19,54], there remain fundamental questions about the ways in which supergenes facilitate adaptation. One of the most prominent of these questions is: do supergenes capture adaptive variants at the time of origin or gain these later? While some empirical studies find evidence that the accumulation of mutations inside a supergene over time caused it to establish and persist ('gain' or 'accumulation'; [55]), others find that the supergene captured a beneficial combination of locally adapted alleles when it originated ('capture'; [56,57]). Schaal *et al.* [58] use simulated data to explore the dynamics of capture and gain, while Stenløkk *et al.* [59] examine the evidence for capture versus gain using data from the Atlantic salmon.

Schaal *et al.* [58] address gaps in our understanding of capture and gain with whole-genome simulations of local adaptation with polygenic architectures and supergenes. Focusing on supergenes where recombination suppression is owing to inversions (i.e. inversion-based supergenes), their study relaxes assumptions made by the previous theory of fixed inversion size and location by allowing inversions of any size to mutate at any location in the genome. Importantly, this allows them to compare the characteristics of inversions involved in local adaptation (i.e. supergenes) to those that are not. The inversion-based supergenes that evolved in their simulations have many characteristics found in empirical studies (e.g. multiple old and large inversions that are outliers in genome scans, sometimes overlapping with other inversions), which highlights that empirical observations are consistent with a highly polygenic architecture and that supergenes do not necessarily need to contain any interesting large-effect genes to play an important role in local adaptation. Their simulations also show how, under polygenic architectures, supergenes can capture a small set of favourable alleles early in adaptation, which sets them on a trajectory to accumulate additional favourable alleles until an equilibrium level of local adaptation is achieved. Specifically, the simulations predict that adaptive inversion-based supergenes will harbour a significant proportion of the additive genetic variance in the trait under selection, but that a significant proportion will also be harboured in the collinear genome (i.e. the non-rearranged regions). Thus, these simulations produce specific hypotheses that can be tested by comparing the characteristics of supergenes to other inversions.

Stenløkk *et al.* [59] use empirical data to test for capture versus gain of adaptive alleles in Atlantic salmon. The authors

take advantage of advances in sequencing and analysis to conduct a comparative analysis of all inversions segregating in the metapopulation of Atlantic salmon across the species range. For each inversion, they conduct an analysis of genes interrupted by the breakpoints, identified putatively deleterious mutations and identified single-nucleotide polymorphisms (SNPs) putatively under spatially heterogeneous selection by the environment. For this latter set of locally adapted SNPs, they were also able to determine if an allele was probably captured at the time of origin of the supergene or gained later, by comparing patterns to the ancestral state inferred from a closely related species. For the one inversion that shows clear patterns of local adaptation to the environment (e.g. a supergene), they find it is of a large size and both the capture and accumulation of adaptive variation was important in forming the supergene, findings that are consistent with the predictions of Schaal et al. [58].

A second critical question regarding the role of supergenes in adaptation is whether supergenes repeatedly arise. Observations of the same supergene are common in isolated populations of the same species or in closely related species and could result from several different evolutionary scenarios: multiple independent origins of the supergene, the presence of the supergene in a common ancestor or adaptive introgression of supergenes. Kay et al. [60] and Westram et al. [61] use both theoretical and empirical methods to differentiate between these possibilities.

Kay et al. [60] examine an elaborate 'social polymorphism' in the colony structure of ants found in many species. Ant colonies are either headed by a single queen (monogyne colonies) or by multiple queens (up into the hundreds; polygyne colonies). This social polymorphism, associated with a battery of phenotypic and life-history traits, reflects highly distinct evolutionary strategies, for which success depends on spatial environmental variation. Monogyne colonies fare better in habitat patches that have newly become available (and queens that found such colonies are, among many adaptations, better colonizers). Polygyne colonies have the upper hand in habitat patches that are already occupied (such colonies contain more workers and occur at higher densities). Remarkably, in the five ant lineages studied to date, it turns out that supergenes are involved in the social polymorphism, each with an independent origin. However, within some of these lineages, the social supergene has been co-opted by additional species within these lineages (in which the social polymorphism never evolved) via adaptive introgression [62]. Thus, both independent origins in addition to spread via adaptive introgression have been important in establishing the ant social supergenes.

The mechanisms behind the role of supergenes in parallel adaptation with gene flow is reviewed and modelled by Westram et al. [61]. The authors review multiple examples in which the same inversions are repeatedly implicated in local adaptation in different locations, indicating that supergenes may play a disproportionate role in parallel evolution. Using theory and simulation, they show how a supergene haplotype can function as an efficient 'transport vehicle' that can bring whole sets of adaptive alleles to a new location. Using models of local adaptation that span from a few patches with different environments to an environmental cline, they show how supergenes in one area of the species range can traverse space to facilitate parallel adaptation in another area of the species range. Their models predict that the presence of

an inversion can create a supergene which speeds up the process of parallel adaptation and sometimes facilitates parallel adaptation when it would not have been possible otherwise. A limitation of the model is the assumption of only two adaptive loci. The authors highlight that an important direction for future research is to explore the dynamics with more polygenic architectures.

A third challenge in studying how supergenes facilitate adaptation is identifying adaptive variants in large regions of linkage disequilibrium generated by recombination suppression. Jay et al. [63] use a multivariate genome-wide association study (GWAS) to dissect the genetic basis of a supergene-based wing pattern polymorphism in the butterfly *Heliconius numata*. Within the supergene, which is itself composed of three chromosomal inversions, the authors identify several independent, putatively adaptive loci that are associated with different aspects of wing patterning. The results of Jay et al. [63] are consistent with a model whereby the inversions making up this supergene might have captured adaptive haplotypes, which predate the origin of the inversions. For some other loci, however, the results suggest that they might have been gained after the formation of the inversions (as modelled in Schaal et al. [58]).

Campoy et al. [64] review and analyse data from both GWAS and functional analyses to understand the phenotypic effects of two large inversions in humans that might likely represent supergenes. One of these inversions, 17q21.31, is associated with multiple complex phenotypes, including brain-related traits, red and white blood cells, lung function, male and female-specific traits, and disease risk. By combining data on nucleotide variation and gene expression, the authors pinpoint the role of three candidate genes (*CRHR1*, *KANLS1* and *MAPT*) that might underlie these phenotypes. The second inversion studied, 8p23.1, is also associated with several related phenotypes and gene expression differences; however, the complex breakpoint structure, and the apparent lack of genetic divergence within the inverted region away from the breakpoints, renders understanding the effects of this inversion challenging. Studying the properties of supergenes in humans is fundamentally important, not only for gaining evolutionary insights, but also for an improved understanding of how supergenes such as inversions impact human health and disease.

4. Maintaining supergene polymorphism

Many known supergenes are over 1 Myr old [18], begging the question of how these polymorphisms are maintained over large time scales in the face of drift and selection. Supergenes that persist over long time scales within a population are especially puzzling, given that supergene haplotypes often degrade by accumulating deleterious mutations [65–67]. Complex selective environments with multiple forms of balancing selection (e.g. overdominance, spatially variable selection, disassortative mating, antagonistic selection and negative frequency-dependent selection) are often necessary to protect supergene polymorphism. The non-trivial dynamics of polymorphism maintenance are examined theoretically by Tafreshi et al. [49] and Berdan et al. [68]. Dagilis et al. [69] take a mixed empirical and simulation approach, searching for signatures of sexually antagonistic selection, a particularly

powerful form of balancing selection, in the three-spine stickleback.

Multiple forms of balancing selection are probably needed to protect polymorphism over long time scales, but we understand little about which combinations of selective pressures can achieve this. Tafreshi *et al.* [49] examine the evolutionary fate of a supergene given different combinations of balancing selection. They present a model based on an ancient supergene (approx. 20–40 Myr) in one particular ant species: the Alpine silver ant, *F. selysi*. This supergene has two haplotypes, *Sm* and *Sp*, that control colony structure in this haplodiploid species. In monogyne colonies, all queens and workers are *Sm/Sm* and males are *Sm*. In polygyne colonies queens and workers are *Sm/Sp* or *Sp/Sp*. However, the males produced in these colonies are only *Sp*, indicating the transmission ratio distortion caused by the *Sp* haplotype via a maternal killing effect [70]. Tafreshi *et al.* [49] find that this maternal killing effect creates a substantial challenge for polymorphism maintenance. In their model a stable polymorphism is only reached when there is both assortative mating, as well as large fitness differences between supergene genotypes. The narrow parameter space that allows polymorphism to be maintained highlights the complex dynamics necessary for supergenes to persist in nature.

One of the major challenges of supergene polymorphism maintenance is that the allelic content of the supergene haplotypes can shift over time. In particular, supergenes may accumulate deleterious mutations that lessen the selective edge of certain haplotype combinations [67]. Using simulations, Berdan *et al.* [68] examine the consequences of mutation accumulation on the maintenance of supergene polymorphism. Recessive deleterious mutations that are private to each supergene haplotype are masked in the heterozygote, generating associative overdominance (AOD; [71]). AOD describes a heterozygote advantage experienced by a neutral variant owing to selection on linked sites [71,72]. Berdan *et al.* [68] find that AOD aids invasion of a new supergene haplotype but that further mutation accumulation can destabilize the system, leading to the loss of the polymorphism. For this reason, the authors conclude that AOD alone is probably insufficient to maintain the polymorphism, and in most systems, multiple forms of balancing selection are probably necessary to maintain polymorphism over long time scales.

Another form of balancing selection that may maintain a supergene polymorphism within a population is sexually antagonistic selection, in which an allele that increases the fitness of males decreases the fitness of females, or vice versa. Dagilis *et al.* [69] investigate the possibility that sexually antagonistic selection is acting on stickleback sex chromosomes. Sexually antagonistic selection may lead to the spread of recombination suppression on sex chromosomes [73–75], but is often difficult to measure on older sex chromosomes where recombination suppression is already complete. Dagilis *et al.* [69] take advantage of the neo-Y chromosome that has been characterized in the Japan Sea stickleback (*Gasterosteus nipponicus*) [76], which formed through the fusion of the typical three-spine stickleback Y with an autosome. The authors compare sequences of male and female sticklebacks across the fused sex chromosome to determine if sexually antagonistic selection is driving genetic divergence in this region. Their results suggest that recombination suppression on the neo-Y chromosome has arisen relatively

recently, perhaps only in the last few hundred thousand years. In the still recombining 'pseudo-autosomal' region of the neo-Y chromosome, the authors identify four genes involved in nervous system processes that show sequence divergence between males and females. However, using simulations, the authors found that demographic history alone could generate patterns of genetic divergence similar to these candidate regions. Thus, sexually antagonistic selection remains difficult to detect, even on recently evolved and recombining sex chromosomes.

5. Mutation accumulation in supergenes

The reduction in effective recombination between supergene haplotypes can be viewed as a double-edged sword: it can facilitate adaptive processes when beneficial alleles are brought together, but might also speed up the accumulation of deleterious mutations [18]. This is because suppressed recombination also reduces the efficacy of purifying selection. The interplay between recombination, drift and linked selection has long-term consequences for the fate of supergenes, which are only now being fully recognized. Advances in the computational power of forward simulations (e.g. SLiM, [77]) are enabling us to theoretically model and predict the evolutionary consequences of supergenes in unprecedented detail. Berdan *et al.* [68] use simulations to look at mutation accumulation in supergenes and the ensuing consequences, while Stenløkk *et al.* [59] look for signatures of this mutation accumulation in Atlantic salmon.

Berdan *et al.* use simulations in SLiM [77,78] to examine the costs and consequences of mutation accumulation in supergenes. While AOD owing to fixed recessive deleterious alleles could facilitate the invasion of a new supergene haplotype and the origin of the supergene polymorphism, continued mutation accumulation strongly threatened the polymorphism. Asymmetry in mutational load between the two haplotypes often leads to the least fit haplotype degenerating more rapidly, which makes the polymorphism sensitive to loss through drift. This is because a high mutational load drives down the frequency of the haplotype, which translates into purifying selection being far less effective and an increase in mutation accumulation. This further decreases the frequency of the more mutationally loaded haplotype, which further decreases the efficacy of purifying selection, in a feedback loop [65]. The authors found that even small differences in load between the two haplotypes could trigger this feedback loop, leading to the loss of the polymorphism. The feedback loop is driven by strong differences in the efficacy of purifying selection in the two haplotypes, which is tied to differences in effective population size. The only way for both haplotypes to degrade at a similar rate is a strong reduction in population size so that drift becomes the main evolutionary force. In small populations, purifying selection will be less effective in general and in the fitter haplotype in particular. This decreases the differential in the efficacy of purifying selection between haplotypes. When both haplotypes degrade at the same rate, they can reach an evolutionary stable state known as a balanced lethal system. As both supergene homozygotes are inviable, and only heterozygotes contribute to subsequent generations, the polymorphism can never be lost without the extinction of the population. This maladaptive situation has been described in plants, insects and vertebrates [79,80].

Stenløkk *et al.* [59] directly test for mutation accumulation in inversions segregating in the metapopulation of Atlantic salmon. For each inversion they assess accumulation of deleterious mutations (or lack thereof), and analyse change through time, by comparing the focal inversion to its ancestral state of inversions in a sister species. They find a lack of evidence for the accumulation of deleterious mutations as predicted by theory, but conclude that this is probably owing to the young age of the inversions.

6. Concluding remarks

By bringing together diverse empirical and theoretical studies of supergenes, this theme issue provides new insights into supergene evolution. While supergenes are a classical subject in evolutionary biology, investigating the evolution of supergenes has only really taken off within the twenty-first century [3,18,19,54]. This renewed interest in supergenes, driven by new methodological and theoretical developments, has revealed that supergenes abound in nature and underlie many fantastically complex phenotypes. It is clear that the evolution of supergenes is incredibly complicated, as both the external environment as well as the allelic content of different supergene haplotypes shift over time. Multiple different processes and selective pressures must then work together to maintain these polymorphisms over long time scales. However, the emergence and subsequent interaction of these different selective pressures over the evolutionary history of a supergene remain unknown. As the papers in this theme issue show, integration of both theoretical and empirical techniques can help bring these puzzles into focus. By bringing together work on diverse approaches, species and supergene structures, we hope this theme issue stimulates future discussion and research on the topic of supergene evolution.

Guest editor profiles

Emma Berdan is a researcher at Gothenburg University. Her research focuses broadly on the role of recombination in evolution and the importance of structural variants. To understand the evolution of large structural variants over long time scales, her research incorporates forward simulations, population genomic techniques, and field and laboratory experiments. Most of her empirical work focuses on insects, in particular the seaweed fly *Coelopa frigida*.

Thomas Flatt is a professor of Evolutionary Biology and Head of the Department of Biology at the University of Fribourg, Switzerland. His research focuses on the experimental population genetics and genomics of adaptation, mainly by using *D. melanogaster* as a model. His studies examine genetic variation for fitness components; life-history clines; inversion polymorphisms maintained by spatially varying selection and other forms of balancing selection; and the evolution of ageing and trade-offs associated with longevity and other fitness-related traits.

Genevieve Kozak is an assistant professor at the University of Massachusetts-Dartmouth. She studies how genomic divergence and phenotypic plasticity contribute to divergence in reproductive traits. She is interested in determining how the genetic architecture of reproductive isolation influences the process of speciation with gene flow. Her research focuses primarily on Lepidoptera.

Katie E. Lotterhos is an associate professor in the Department of Marine and Environmental Sciences at Northeastern University. Her research focuses on the evolution, population genomics, and statistical inference of adaptation and plasticity. To better understand how to conserve species in the face of climate change, her research integrates empirical studies with simulations of virtual species. Much of her work is focused on marine fish and invertebrates, and other non-model species.

Ben Wielstra is an assistant professor at the Institute of Biology Leiden, Leiden University and an honorary researcher at Naturalis Biodiversity Center. His research interests focus on how genomes change composition in response to the associated positive and negative fitness consequences of hybridization, combining the genetic material of two distinct species, and structural rearrangements, splitting up a single stretch of DNA into two independently evolving units. Most of his work focuses on salamanders.

References

1. Charlesworth D. 2016 The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol. Appl.* **9**, 74–90. (doi:10.1111/eva.12291)

2. Kimura M. 1956 A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**, 278–287. (doi:10.1111/j.1558-5646.1956.tb02852.x)

3. Thompson MJ, Jiggins CD. 2014 Supergenes and their role in evolution. *Heredity* **113**, 1–8. (doi:10.1038/hdy.2014.20)

4. Turner JR. 1967 On supergenes. I. The evolution of supergenes. *Am. Nat.* **101**, 195–221. (doi:10.1086/282485)

5. Darlington CD, Mather K. 1949 *The elements of genetics*. London: George Allen & Unwin.

6. Charlesworth B, Barton N. 1996 Recombination load associated with selection for increased recombination. *Genet. Res.* **67**, 27–41. (doi:10.1017/S0016672300033450)

7. Charlesworth B, Charlesworth D. 1975 An experiment on recombination load in *Drosophila melanogaster*. *Genet. Res.* **25**, 267–273. (doi:10.1017/S001667230001569X)

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

9. Lenormand T. 2002 Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183–189. (doi:10.1016/S0169-5347(02)02497-7)

10. Charlesworth B. 1974 Inversion polymorphism in a two-locus genetic system. *Genet. Res.* **23**, 259–280. (doi:10.1017/S0016672300014919)

11. Charlesworth B, Flatt T. 2021 On the fixation or nonfixation of inversions under epistatic selection. *Mol. Ecol.* **30**, 3896–3897. (doi:10.1111/mec.16026)

12. Dobzhansky T. 1949 Observations and experiments on natural selection in *Drosophila*. *Hereditas* **35**, 210–224. (doi:10.1111/j.1601-5223.1949.tb03334.x)

13. Dobzhansky T. 1950 Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. *Genetics* **35**, 288–302. (doi:10.1093/genetics/35.3.288)

14. Lewontin R, Kojima KI. 1960 The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472. (doi:10.2307/2405995)

15. Ford EB. 1965 *Genetic polymorphism*. Cambridge, MA: The MIT Press.

16. Mayr E. 1963 *Animal species and evolution*. Cambridge, MA: Harvard University Press.

17. Fisher RA. 1930 *The genetical theory of natural selection*. Oxford, UK: Clarendon.

18. Gutiérrez-Valencia J, Hughes W, Berdan EL, Slotte T. 2021 The genomic architecture and evolutionary fates of supergenes. *Genome Biol. Evol.* **13**, evab057. (doi:10.1093/gbe/evab057)

19. Schwander T, Libbrecht R, Keller L. 2014 Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294. (doi:10.1016/j.cub.2014.01.056)

20. Berdan EL, Blanckaert A, Slotte T, Suh A, Westram AM, Fragata I. 2021 Unboxing mutations: connecting mutation types with evolutionary

consequences. *Mol. Ecol.* **30**, 2710–2723. (doi:10. 1111/mec.15936)

21. Komata S, Kajitani R, Itoh T, Fujiwara H. 2022 Genomic architecture and functional unit of mimicry supergene in female limited Batesian mimic *Papilio* butterflies. *Phil. Trans. R. Soc. B* **377**, 20210198. (doi:10.1098/rstb.2021.0198)

22. Charlesworth B, Barton NH. 2018 The spread of an inversion with migration and selection. *Genetics* **208**, 377–382. (doi:10.1534/genetics.117.300426)

23. Charlesworth B, Charlesworth D. 1973 Selection of new inversions in multi-locus genetic systems. *Genet. Res.* **21**, 167–183. (doi:10.1017/S0016672300013343)

24. Morel F, Douet-Guilbert N, Le Bris MJ, Herry A, Amice V, Amice J, De Braekeleer M. 2004 Meiotic segregation of translocations during male gametogenesis. *Int. J. Androl.* **27**, 200–212. (doi:10. 1111/j.1365-2605.2004.00490.x)

25. Talukdar D. 2010 Reciprocal translocations in grass pea (*Lathyrus sativus L.*): pattern of transmission, detection of multiple interchanges and their independence. *J. Hered.* **101**, 169–176. (doi:10. 1093/jhered/esp106)

26. Kim K-W, De-Kayne R, Gordon IJ, Omufwoko KS, Martins DJ, ffrench-Constant R, Martin SH. 2022 Stepwise evolution of a butterfly supergene via duplication and inversion. *Phil. Trans. R. Soc. B* **377**, 20210207. (doi:10.1098/rstb.2021.0207)

27. Finseth F, Brown K, Demaree A, Fishman L. 2022 Supergene potential of a selfish centromere. *Phil. Trans. R. Soc. B* **377**, 20210208. (doi:10.1098/rstb. 2021.0208)

28. Nishikawa H et al. 2015 A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47**, 405–409. (doi:10.1038/ng.3241)

29. Palmer DH, Kronforst MR. 2020 A shared genetic basis of mimicry across swallowtail butterflies points to ancestral co-option of doublesex. *Nat. Commun.* **11**, 1–10. (doi:10.1038/s41467-019-13859-y)

30. Iijima T, Kajitani R, Komata S, Lin CP, Sota T, Itoh T, Fujiwara H. 2018 Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Sci. Adv.* **4**, eaao5416. (doi:10.1126/sciadv.aao5416)

31. Kunte K, Zhang W, Tenger-Trolander A, Palmer D, Martin A, Reed R, Mullen SP, Kronforst M. 2014 Doublesex is a mimicry supergene. *Nature* **507**, 229–232. (doi:10.1038/nature13112)

32. Charlesworth B, Charlesworth D. 1997 Rapid fixation of deleterious alleles can be caused by Muller's ratchet. *Genet. Res.* **70**, 63–73. (doi:10.1017/S0016672397002899)

33. Dolgin ES, Charlesworth B. 2008 The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**, 2169–2177. (doi:10.1534/genetics.107.082743)

34. Gilbert KJ, Pouyet F, Excoffier L, Peischl S. 2019 Transition from background selection to associative overdominance promotes diversity in regions of low recombination. *Curr. Biol.* **30**, 101–107. (doi:10. 1016/j.cub.2019.11.063)

35. Kliman RM, Hey J. 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**, 1239–1258.

36. Muller HJ. 1964 The relation of recombination to mutational advance. *Mutation Res./Fundamental Mol. Mech. Mutagenesis* **1**, 2–9. (doi:10.1016/0027-5107(64)90047-8)

37. Dufresnes C, Crochet P-A. 2022 Sex chromosomes as supergenes of speciation: why amphibians defy the rules? *Phil. Trans. R. Soc. B* **377**, 20210202. (doi:10. 1098/rstb.2021.0202)

38. Qvarnström A, Bailey RI. 2009 Speciation through evolution of sex-linked genes. *Heredity* **102**, 4–15. (doi:10.1038/hdy.2008.93)

39. Bachtrog D et al. 2014 Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899. (doi:10.1371/journal.pbio.1001899)

40. Presgraves DC. 2018 Evaluating genomic signatures of 'the large X-effect' during complex speciation. *Mol. Ecol.* **27**, 3822–3830. (doi:10.1111/mec.14777)

41. Wright D, Schaeffer SW. 2022 The relevance of chromatin architecture to genome rearrangements in *Drosophila*. *Phil. Trans. R. Soc. B* **377**, 20210206. (doi:10.1098/rstb.2021.0206)

42. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380. (doi:10.1038/nature11082)

43. Gonzalez-Sandoval A, Gasser SM. 2016 On TADs and LADs: spatial control over gene expression. *Trends Genet.* **32**, 485–495. (doi:10.1016/j.tig.2016.05.004)

44. Pombo A, Dillon N. 2015 Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257. (doi:10.1038/nrm3965)

45. Yu M, Ren B. 2017 The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289. (doi:10.1146/annurev-cellbio-100616-060531)

46. Bravo-Núñez MA, Nuckolls NL, Zanders SE. 2018 Genetic villains: killer meiotic drivers. *Trends Genet.* **34**, 424–433. (doi:10.1016/j.tig.2018.02.003)

47. Lindholm AK et al. 2016 The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31**, 315–326. (doi:10.1016/j.tree.2016.02.001)

48. Schimenti J. 2000 Segregation distortion of mouse t haplotypes: the molecular basis emerges. *Trends Genet.* **16**, 240–243. (doi:10.1016/S0168-9525(00)02020-5)

49. Tafreshi AG, Otto SP, Chapuisat M. 2022 Unbalanced selection: the challenge of maintaining a social polymorphism when a supergene is selfish. *Phil. Trans. R. Soc. B* **377**, 20210197. (doi:10.1098/rstb. 2021.0197)

50. Wong HW, Holman L. 2020 Fitness consequences of the selfish supergene segregation distorter. *J. Evol. Biol.* **33**, 89–100. (doi:10.1111/jeb.13549)

51. Berdan EL, Mérot C, Pavia H, Johannesson K, Wellenreuther M, Butlin RK. 2021 A large chromosomal inversion shapes gene expression in seaweed flies (*Coelopa frigida*). *Evol. Lett.* **5**, 607–624. (doi:10.1002/evl3.260)

52. Lavington E, Kern AD. 2017 The effect of common inversion polymorphisms In(2 L)t and In(3R)Mo on patterns of transcriptional variation in *Drosophila melanogaster*. *G3* **7**, 3659–3668. (doi:10.1534/g3. 117.1133)

53. Tuttle EM et al. 2016 Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350. (doi:10.1016/j.cub.2015.11.069)

54. Black D, Shuker DM. 2019 Supergenes. *Curr. Biol.* **29**, R615–R6R7. (doi:10.1016/j.cub.2019.05.024)

55. Lamichhaney S et al. 2016 Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88. (doi:10.1038/ng.3430)

56. Coughlan JM, Willis JH. 2019 Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Mol. Ecol.* **28**, 1343–1357. (doi:10.1111/mec.14804)

57. Kirubakaran TG et al. 2016 Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol. Ecol.* **25**, 2130–2143. (doi:10.1111/mec.13592)

58. Schaal SM, Haller BC, Lotterhos KE. 2022 Inversion invasions: when the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Phil. Trans. R. Soc. B* **377**, 20210200. (doi:10.1098/rstb.2021.0200)

59. Stenløkk K, Saitou M, Rud-Johansen L, Nome T, Moser M, Árnyasi M, Kent M, Barson NJ, Lien S. 2022 The emergence of supergenes from inversions in Atlantic salmon. *Phil. Trans. R. Soc. B* **377**, 20210195. (doi:10.1098/rstb.2021.0195)

60. Kay T, Helleu Q, Keller L. 2022 Iterative evolution of supergene-based social polymorphism in ants. *Phil. Trans. R. Soc. B* **377**, 20210196. (doi:10.1098/rstb. 2021.0196)

61. Westram AM, Faria R, Johannesson K, Butlin R, Barton N. 2022 Inversions and parallel evolution. *Phil. Trans. R. Soc. B* **377**, 20210203. (doi:10.1098/rstb.2021.0203)

62. Kirkpatrick M, Barrett B. 2015 Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Mol. Ecol.* **24**, 2046–2055. (doi:10. 1111/mec.13074)

63. Jay P, Leroy M, Le Poul Y, Whibley A, Arias M, Chouteau M, Joron M. 2022 Association mapping of colour variation in a butterfly provides evidence that a supergene locks together a cluster of adaptive loci. *Phil. Trans. R. Soc. B* **377**, 20210193. (doi:10. 1098/rstb.2021.0193)

64. Campoy E, Puig M, Yakymenko I, Lerga-Jaso J, Cáceres M. 2022 Genomic architecture and functional effects of potential human inversion supergenes. *Phil. Trans. R. Soc. B* **377**, 20210209. (doi:10.1098/rstb.2021.0209)

65. Berdan EL, Blanckaert A, Butlin RK, Bank C. 2021 Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genet.* **17**, e1009411. (doi:10.1371/journal.pgen.1009411)

66. Faria R, Johannesson K, Butlin RK, Westram AM. 2019 Evolving inversions. *Trends Ecol. Evol.* **34**, 239–248. (doi:10.1016/j.tree.2018.12.005)

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

68. Berdan EL, Blanckaert A, Butlin RK, Flatt T, Slotte T, Wielstra B. 2022 Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals. *Phil. Trans. R. Soc. B* **377**, 20210199. (doi:10.1098/rstb.2021.0199)

69. Dagilis AJ, Sardell JM, Josephson MP, Su Y, Kirkpatrick M, Peichel CL. 2022 Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. *Phil. Trans. R. Soc. B* **377**, 20210205. (doi:10.1098/rstb.2021.0205)

70. Avril A, Purcell J, Béniguel S, Chapuisat M. 2020 Maternal effect killing by a supergene controlling ant social organization. *Proc. Natl Acad. Sci. USA* **117**, 17 130–17 134. (doi:10.1073/pnas.2003282117)

71. Ohta T. 1971 Associative overdominance caused by linked detrimental mutations. *Genet. Res.* **18**, 277–286. (doi:10.1017/S0016672300012684)

72. Jones DF. 1917 Dominance of linked factors as a means of accounting for heterosis. *Genetics* **2**, 466. (doi:10.1093/genetics/2.5.466)

73. Charlesworth D. 2021 When and how do sex-linked regions become sex chromosomes? *Evolution* **75**, 569–581. (doi:10.1111/evo.14196)

74. Charlesworth D, Charlesworth B. 1980 Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214. (doi:10.1017/S0016672300014051)

75. Rice WR. 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**, 911–914. (doi:10.1111/j.1558-5646.1987.tb05864.x)

76. Kitano J *et al.* 2009 A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079–1083. (doi:10.1038/nature08441)

77. Haller BC, Messer PW. 2019 SLiM 3: forward genetic simulations beyond the Wright-Fisher model. *Mol. Biol. Evol.* **36**, 632–637. (doi:10.1093/molbev/msy228)

78. Haller BC, Messer PW. 2016 SLiM 2: flexible, interactive forward genetic simulations. *Mol. Biol. Evol.* **34**, 230–240. (doi:10.1093/molbev/msw211)

79. Grossen C, Neuenschwander S, Perrin N. 2012 The balanced lethal system of crested newts: a ghost of sex chromosomes past? *Am. Nat.* **180**, E174–EE83. (doi:10.1086/668076)

80. Wielstra B. 2020 Balanced lethal systems. *Curr. Biol.* **30**, R742–R7R3. (doi:10.1016/j.cub.2020.05.011)