REVIEW
Supergenes and their role in evolution

MJ Thompson[1,2] and CD Jiggins[1]

Adaptation is commonly a multidimensional problem, with changes in multiple traits required to match a complex environment. This is epitomized by balanced polymorphisms in which multiple phenotypes co-exist and are maintained in a population by a balance of selective forces. Consideration of such polymorphisms led to the concept of the supergene, where alternative phenotypes in a balanced polymorphism segregate as if controlled by a single genetic locus, resulting from tight genetic linkage between multiple functional loci. Recently, the molecular basis for several supergenes has been resolved. Thus, major chromosomal inversions have been shown to be associated with polymorphisms in butterflies, ants and birds, offering a mechanism for localised reduction in recombination. In several examples of plant self-incompatibility, the functional role of multiple elements within the supergene architecture has been demonstrated, conclusively showing that balanced polymorphism can be maintained at multiple coadapted and tightly linked elements. Despite recent criticism, we argue that the supergene concept remains relevant and is more testable than ever with modern molecular methods.
Heredity (2014) **113**, 1–8; doi:10.1038/hdy.2014.20; published online 19 March 2014

INTRODUCTION

'…the evolution of super-genes, whether consisting of a few closely-linked loci or of an inversion, should always be (but never, I think, is) treated as one of the fundamental properties of genetics.' (Ford, 1965)

One challenge for organisms as they adapt to their environment is that adaptive changes in multiple characters are often necessary to match complex differences between alternative environments. This poses a problem when individuals experiencing divergent adaptive challenges can still interbreed, such as in a balanced polymorphism or during local adaptation across a heterogeneous geographic range. Recombination between multiple loci involved in such adaptations imposes a selective cost on alleles due to recombination between co-adapted alleles. The fitness of individual alleles can be enhanced by a genetic architecture that maintains favourable combinations of alleles and reduces the production of recombinant genotypes. Genetic architectures that can maintain traits in favourable combinations against the disruptive forces of recombination are central to explaining many phenomena in evolutionary biology, from the evolution of genetic sex determination, to local adaptation in complex environments, to the origin of species.

The genomic architectures that can evolve range from the inversions found in many cases of local adaptation (Feder *et al.*, 2003; Cheng *et al.*, 2012) through to complex supergenes in which multiple characters are inherited as a single locus. In the case of local adaptation, theory predicts that selection against maladaptive intermediates between two locally adapted forms, unfit in either environment, can promote the spread of mechanisms that reduce recombination (Kirkpatrick and Barton, 2006). Several examples of

this phenomenon have now been documented, in which co-adapted phenotypic differences are maintained in favourable combinations by physical linkage and mechanisms to suppress recombination (for example, *Mimulus* inversions; Lowry and Willis, 2010).

Another example of the same phenomenon are 'supergenes', a term typically confined to cases of balanced polymorphism in which apparently complex adaptations have a simple genetic basis. Supergenes have been widely interpreted as consisting of multiple tightly linked loci that regulate a system of discrete phenotypes. Classic examples include alternative morphs in polymorphic Batesian mimics and heterostyly, in which polymorphism in floral morphology prevents inbreeding. These systems share a mechanism for maintenance of polymorphism, commonly negative frequency-dependent selection, and complex alternative phenotypes that influence multiple aspects of morphology. Supergenes therefore provide a mechanism whereby a complex balanced polymorphism can be maintained without the generation of maladaptive intermediates. The existence of highly discrete morphs without intermediates means that supergenes represent spectacular examples of morphological adaptation.

The term supergene was coined during the formulation of the Modern Synthesis. Although some have called into question the existence of supergenes, recent work has begun to dissect their molecular basis and shed new light on the concept. As we learn ever more about the molecular functioning of genomes, an updated and more nuanced definition of the term supergene is required. Here we examine the history and use of the term before reviewing recent criticisms and synthesising a new definition in the light of recent examples.

HISTORY
Over 100 years ago, in one of the first stirrings of the Modern Synthesis, Punnett (1905) argued that the works of Mendel allowed a

[1]Department of Zoology, University of Cambridge, Cambridge, UK and [2]Department of Life Sciences, Natural History Museum, London, UK
Correspondence: MJ Thompson, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.
E-mail: mjt68@cam.ac.uk

new perspective on Darwinism, one that overcame earlier problems. Punnett believed in the finite and discrete nature of the variation that provides the raw material for evolutionary change. In contrast to some earlier views, this variation was seen as independent from the direction of evolutionary change, with natural selection acting as a filter upon variation. Punnett categorised variation as either non-heritable, environmental variation, which he termed 'fluctuations', or heritable changes, known as mutations. By referring to systems where complex mimetic wing patterns could be ascribed to changes in single genes, Punnett argued that major morphological changes evolved from single mutations and that the evolution of mimicry often involved parallel mutations in both model and mimic (Punnett, 1915).

The idea of multiple linked changes regulating a phenotype arose from the gradualist argument against this mutationist perspective. Fisher (and later Ford) considered parallel mutation as highly improbable and instead argued that mimicry evolves via a gradualist adaptive walk of many steps (Fisher, 1930). They interpreted the observation that polymorphisms in butterflies are often determined by one or a few Mendelian factors as the outcome of a long process of evolution, rather than evidence for a single mutation of large effect. Fisher, (1930a) gives a recognisable definition of a supergene (although Fisher himself did not use the term) in the context of the genetics of mimetic polymorphism in *Papilio polytes*:

'…polymorphism in this case, and probably in similar cases, is dependent on one or more Mendelian factors the function of which is to switch on one or other of the possible alternatives.'

Fisher (1930a) drew parallels between the linked genes presumed to underlie butterfly mimicry and the evolution of the sex chromosomes: in both cases complex differences between phenotype classes are determined by inheritance of a single locus, although likely representing many genes and whose identity is maintained by the absence of recombination.

Dobzhansky's pioneering work on *Drosophila* chromosomes provided a potential mechanism for the evolution of tight linkage. He demonstrated the widespread occurrence of inversions (Dobzhansky and Sturtevant, 1938) which were quickly recognised as suppressors of recombination such that alleles within an inversion are 'locked' together in physical linkage. The observation of clines in inversion frequency in species of *Drosophila*, as well as seasonal fluctuations within individual populations, remain canonical examples of local adaptation (Dobzhansky, 1947).

In the light of these observations, Darlington and Mather (1949) coined the term supergene as '*a group of genes acting as a mechanical unit in particular allelic combinations*' and emphasised the role of inversions as a means of 'pegging' together combinations of genes and suppressing recombination. Mather (1950, 1955) characterised supergenes as a series of '*ganged switches*' that direct development down different developmental trajectories. His hypothesis, informed by work on heterostyly in *Primula*, was that selection would act on initially unlinked loci, such as the *SI* genes in *P. sinensis*, and would favour increasing linkage between components of the switch. This would lead to evolution of a composite unit, as in *P. viscosa*, and finally one single switch that cannot be unravelled by recombination, as in *P. vulgaris* (Mather, 1950).

In parallel to the work on *Primula*, polymorphic mimetic butterflies also provided much of the evidence for the origins of supergenes. Clarke and Sheppard (reviewed in Clarke *et al.*, 1995) envisaged the gradual evolution of supergenes through accumulation

of both linked mutations but also unlinked 'modifiers'. Evidence for linked mutations came from the observation of rare phenotypes in both *P. dardanus* and *P. memnon* that were interpreted as recombinants. In some cases the order of the various components within the supergene could be inferred (Clarke *et al*, 1968), and recombination within a supergene architecture can therefore have a role in generating diversity. The evidence for unlinked modifier alleles came from crosses between geographic populations in which the phenotypes controlled by the supergene broke down (Clarke and Sheppard, 1971). Hence, unlinked modifier alleles are inferred to act epistatically with alternate alleles at the supergene locus in order to perfect the mimetic phenotypes in local populations.

By the mid 1960s the 'supergene' model of complex adaptation was considered as a core element of the modern synthesis. When Hamilton sought to explain the evolution of altruism, supergenes were key to the 'green beard' model (Hamilton, 1964; Dawkins, 1976). This hypothesis states that if individuals can recognise the allelic state of other individuals at a locus (the 'green beard'), they could behave in such a way as to maximise the inclusive fitness of the allele that they possess. Hamilton recognised that such a mechanism would require a complex genomic architecture:

'That genes could cause the perception of the presence of like genes in other individuals may sound improbable; at simplest we need to postulate something like a supergene …' (Hamilton, 1964).

In summary, by the early 1970s there was a clear definition of a supergene, perhaps best encapsulated in Dobzhansky's definition:

'coadapted combinations of several or many genes locked in inverted sections of chromosomes and therefore inherited as single units.' (Dobzhansky, 1970).

MODELLING THE ORIGINS OF A SUPERGENE

The supergene concept was firmly established by the mid/late 20th century; however, there was little theoretical treatment of the mechanisms by which such an architecture could arise. There was, however, a considerable body of work on modifiers of recombination rates in the context of local adaptation and clines (Kimura, 1956; Nei *et al.*, 1967; Charlesworth and Charlesworth, 1979), showing that selection could act to modify the recombination rate between two already linked loci and that the rate of change in recombination was greatest when the modifier was already in strong linkage disequilibrium with the loci under selection (Charlesworth and Charlesworth, 1979).

It was initially considered that selection for genetic linkage of co-adapted traits might result in translocation of genes involved in specifying a co-adapted trait, leading to a concentration of tightly linked genes on one chromosome. However, in line with emerging theoretical work on inversion clines, it was realised that a multilocus multiallele polymorphism is unlikely to be stable without some initial degree of linkage, which can be further tightened by selection

'The difficulty of building two or more genes into a co-adapted system must be much increased when they start on different chromosomes' (Ford, 1965).

The 'translocation' model of supergene evolution was considered less plausible and it was thought more likely that new mutations arose in linkage with the supergene. This eventually became known as the 'sieve' (or 'Turner's sieve' model, after Turner, 1967)

The 'sieve' argument was formalised by Charlesworth and Charlesworth (1975a, b, c) who modelled the evolution of polymorphic Batesian mimicry involving modifier and mimicry genes with differing dominance relationships, epistasis and fitness effects. The Charlesworths demonstrated that to maintain two loci together in a polymorphism, a recombination fraction of between 0.01 and 0.001 was required (the exact number depending on other parameters such as model abundance).

More recent theoretical work has mainly focused on local adaptation. In particular, Kirkpatrick and Barton (2006) explored the conditions under which an inversion is favoured in locally adapted populations that exchange migrants. They find that the rate of spread of a new inversion locking locally adapted loci together is independent of the relative fitnesses of the two competing sets of alleles and is determined by the migration rate and the number of loci captured. Alleles within inversions spread as they escape the fitness costs of recombination into disadvantageous genetic backgrounds. This theory has been further advanced by Yeaman (2013), who has suggested that local adaptation can lead to the evolution of a single-locus genetic architecture by gene translocation. In these local adaptation models, low migration rates between populations allow establishment of a stable multilocus polymorphism, providing a starting point from which gene translocation can occur. In contrast, under negative frequency-dependent selection, selection has to be extremely strong to maintain multilocus polymorphism and translocation is considered unlikely (Charlesworth and Charlesworth, 1975c).

Using an approach known as adaptive dynamics, which considers the likelihood that novel alleles can invade an existing population, several recent papers have considered frequency-dependent disruptive selection (Kopp and Hermisson, 2006; Van Doorn and Dieckmann, 2006). These have also shown that the genetic architecture underlying polymorphism can evolve from an initially multilocus architecture towards genetic control by one or a few major effect loci. Kopp and Hermisson (2006) additionally study the effects of linkage on the evolution of genetic architecture, finding that where loci are initially tightly linked ($r < 0.1$), several loci could be maintained as polymorphic. Increasing linkage between such loci can evolve if the selection coefficient is greater than the recombination rate between loci (Barton, 2000). In summary, theory implies that rearrangements such as inversions and perhaps even long-range translocations may have a role in supergene evolution, and importantly these ideas are now directly testable with genomic data.

A NEW PERSPECTIVE FROM DEVELOPMENTAL GENETICS

The supergene hypothesis has also been criticised by developmental biologists. Nijhout (1991, 1994) argued that in most butterfly polymorphisms the designation of a 'supergene' architecture rested on a '*supposition that the differences in wing pattern between the various female forms are simply too great to be accounted for by the effects of a single gene*' (1994). Nijhout instead advocates a model influenced by insight from molecular genetics, citing examples of homeotic *Drosophila* mutants in which many distinct phenotypes map to a single coding region (Nijhout, 1994), and suggests that switching of multiple phenotypes can be explained by a single transcription factor with complex downstream targets, or by alternative splicing of a single gene (Nijhout, 1991). This provides a molecular mechanism to explain how one locus could take over control of a complex phenotype. More recently, Nijhout (2003) has directly contrasted his model of '*a polymorphic regulatory gene*' with a supergene (defined following Charlesworth and Charlesworth, 1975b, 1975c as '*modifying mutations…tightly linked to the gene which*

conferred the initial advantage') with *P. memnon* given as an example of the latter.

West-Eberhard (2003) has similarly advocated 'regulatory linkage' as an alternative to chromosomal linkage in determining co-expression of genes. West-Eberhard considers the cases of Batesian mimicry and other supergenes to be special cases of developmental switches, which are widespread and reasonably well understood. In effect, this is the same point made by Nijhout—a single regulatory locus with complex downstream effects could act in the same way as the linked set of loci described in the classical conception of a supergene. Such a regulatory locus is similar to the 'input–output' genes that regulate the development of organs such as *Drosophila* sensory bristles and become the target for evolutionary diversification (Stern and Orgogozo, 2008). The 'regulatory gene' hypothesis is proposed as an alternative hypothesis to explain the empirical data; however, in reality this is a false dichotomy. These perspectives are useful and clearly a more nuanced consideration of the supergene is needed, taking into account a modern view of developmental genetics. However, it is also clear that a single regulatory gene can possess multiple co-adapted cis-regulatory elements that control its expression in different developmental contexts, which would be genetically equivalent to the classical conception of a supergene. The emerging evidence from empirical examples indicates that in several cases putative supergenes do indeed show high levels of linkage disequilibrium (for example, Huynh *et al.*, 2010a, b; Wang *et al.*, 2013) and chromosomal rearrangements (for example, Joron *et al.*, 2011), providing support for the classical conception of a supergene involving multiple linked and co-adapted elements. In most cases, what remains to be determined is the mechanism by which such loci control multiple downstream targets.

A NEW DEFINITION

Drawing upon these examples in the light of recent criticisms we can synthesise a new definition of a supergene:

> A genetic architecture involving multiple linked functional genetic elements that allows switching between discrete, complex phenotypes maintained in a stable local polymorphism.

We have made this definition as inclusive as possible and avoid any mention of the number of functional open-reading frames included in the locus. Variation in multiple *cis*-regulatory elements could underlie switching between developmental trajectories through regulatory control of a single expressed protein. In common with historical definitions (Darlington and Mather, 1949), we would include sex chromosomes as a special case of supergenes, although due to their distinct nature these are more usefully considered separately.

There is also conceptual overlap with some cases of local adaptation; however, we restrict the term 'supergene' to cases where a stable polymorphism within a population is controlled by a single locus that segregates in a Mendelian manner and is maintained at intermediate frequency. Local adaptation, such as that in *Mimulus*, typically does not involve coexistence of both morphs at intermediate frequencies within populations (Kirkpatrick and Barton, 2006). Nonetheless, putative supergenes of *Heliconius numata, Cepaea nemoralis* and some brood parasites may be maintained by local adaptation to such a fine-grained habitat that the result is a stable local polymorphism rather than between population adaptive differentiation. We have therefore avoided any mention of the specific selection regime that maintains polymorphism, as this can be hard to determine with certainty in particular cases. Furthermore, we make no explicit

mention of mechanisms to reduce recombination rates. A supergene could arise from initially very tightly linked elements, such that selection for any further reduction in recombination rate would be weak.

Supergenes may also partially overlap with examples of ongoing speciation with gene flow. Two paraphyletic sympatric 'sister' species of *Petunia* illustrate this continuum. The species *P. axillaris* and *P. exserta* are pollinated by hawkmoths and hummingbirds, respectively. Flowers of the two species are adapted to their divergent pollinators in terms of colour (both visible and UV), scent and floral morphology. The genes underlying the co-adaptations have been mapped to a single locus, components of which can be separated by rare recombinants (Hermann *et al.*, 2013). Floral traits are undoubtedly under divergent selection to adapt them to their pollinators (Venail *et al.*, 2010), and the traits map to distinct but very closely linked loci, meaning that this system shares many characteristics with other supergenes. The key difference differentiating this case from the following examples is that the co-adapted traits are maintained in separate populations with the traits themselves under selection as prezygotic isolating mechanisms, rather than segregating as a polymorphism in a single population.

In summary, to qualify as a supergene, a biological system needs to demonstrate clear evidence of a complex phenotype of multiple co-adapted elements, with a pattern of inheritance essentially identical to alternative alleles at a single locus, and maintained in a polymorphism in a single population. Our definition does not explicitly require suppression of recombination or non-additive fitness effects among components of the supergene, although these may be important in some systems. Here we review putative examples of supergenes in the light of this definition.

SELF-INCOMPATIBILITY AND HETEROSTYLY

The best characterised supergenes are to be found in the self-incompatibility (SI) and heterostyly systems that have evolved in several groups of flowering plants. SI describes a multitude of different mechanisms that have evolved convergently and that ensure outcrossing through the prevention of fertilisation by self-pollen—this can be molecular homomorphic SI, for example in *Papaver*, or it can also include physical separation of male and female parts within a flower known as heteromorphic self-incompatibility or heterostyly, such as occurs in *Primula*. Typically, tight linkage between genes determining male- and female-specific responses are crucial to the maintenance of SI, thus favouring a supergene architecture (Takayama and Isogai, 2005).

In homomorphic SI, incompatibility is accomplished through intercellular interactions, generally leading to inhibited growth or death of self-pollen. Homomorphic SI has evolved independently multiple times; the particular molecular mechanisms involved are unrelated, although all share linked multi-allelic multi-gene architectures (Figure 1, reviewed in Takayama and Isogai, 2005). The best characterised SI mechanism comes from the Brassicaceae. Here, the *S* locus contains at least three genes: *SLG*, *SLK* and *SP11*. SLK is necessary and sufficient to induce the female reaction, whereas SP11 encodes the male determinant. Small RNAs also encoded at *S* act in a *trans*-regulatory manner to determine the dominance relationship among different *S* haplotypes—SI in the Brassicaceae is determined by the genes of the haploid sporophyte. The inclusion of a dominance-modifying factor within the supergene is in agreement with arguments from the earliest days of supergene theory. Recognition of self is mediated through interaction of a secreted peptide (SP11) with a transmembrane receptor kinase (SLK). That *SLK* and
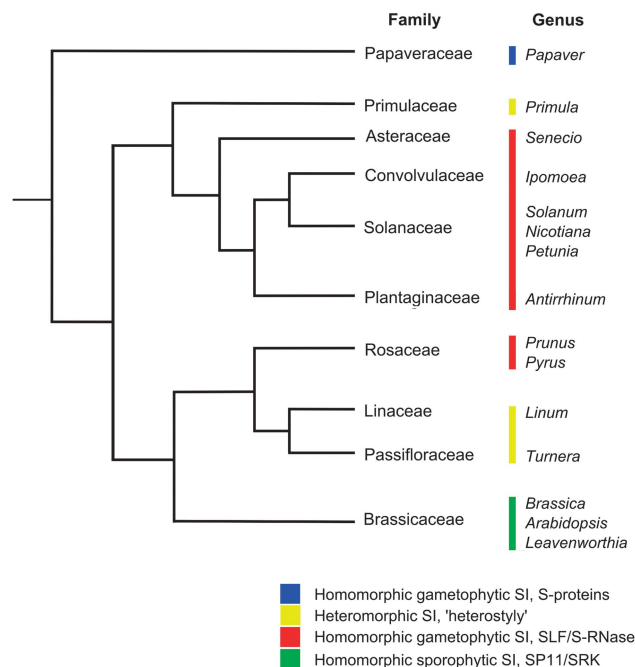
Figure 1 Convergent evolution of SI mechanisms across flowering plants. The S-RNases are shared between Asterid and Rosid lineages, which diverged c. 100 mya. Genera where SI has been documented are annotated within each family. Phylogenetic relationships sourced from http://tolweb.org, divergence estimate from http://timetree.org.

SP11 are indeed the genes responsible seems certain: it is possible to induce an SI reaction in the normally self-fertile *Arabidopsis thaliana* (with a degenerate *S* locus) by transferring SLK and SP11 from the self-incompatible *A. lyrata* (Kusaba *et al.*, 2001; Nasrallah *et al.*, 2002). Intriguingly, it appears that different SI genes within the Brassicaceae are not homologous. Species in the genus *Leavenworthia* lack the 'classical' Brassicaceae SI genes SLK/SP11—in these species SI is determined by paralogues of SLK/SP11, *LaLal2* and *LaSCRL* (Chantha *et al.*, 2013). *LaLal2* and *LaSCRL* have syntenic homologues in *A. lyrata*, yet are unlinked to *S* genes *SLK* and *SP11*, suggesting that *Leavenworthia* has convergently (re)evolved a SI supergene from an existing linked architecture. This is strong evidence that selection for co-segregation of loci determining complex traits can give rise to supergenes, and that whereas the requirement of initial linkage (Turner's sieve) may seem strict, there is abundant evidence that supergenes can indeed arise by this mechanism.

In contrast to the sporophytically determined SI in Brassicaceae, the two other systems for which SI has been well characterised demonstrate gametophytic determination of pollen type. In the Solanaceae, Asteraceae, Convolvulaceae, Plantaginaceae and Rosaceae, SI is mediated by the reaction of a secreted pistil-expressed RNase (S-RNase) and a pollen-expressed F-box protein, SLF (Murfett *et al.*, 1994; Sijacic *et al.*, 2004). Both female (for example, Ushijima *et al.*, 1998) and male (for example, Entani *et al.*, 2003) components exhibit high sequence diversity, with novel haplotypes possibly derived by recombination (Wang *et al.*, 2001) or gene duplication: in the case of SLF, more than five different subgroups have been detected and several SLF-like genes are found in close linkage at the *S* locus and are co-expressed in pollen (Kubo *et al.*, 2010). In the tomato, the SI system is an important mediator of interspecific unilateral incompatibility (Li and Chetelat, 2010), once again highlighting the importance of supergenes in ecological speciation.

The third SI system with a well-characterised mechanism is in Papaveraceae. In this system, the female *S* locus gene encodes a secreted 15-kDa protein, which is thought to interact with a *S*-encoded, pollen tube-expressed transmembrane receptor (Foote and Ride, 1994; Wheeler *et al.*, 2009). Interaction of non-compatible *S* proteins leads to disruption of pollen-tube growth through calcium-dependent depolymerisation of the pollen-tube cytoskeleton and apoptosis (Wilkins *et al.*, 2011).

SI mechanisms in flowering plants therefore represent the best-characterised examples of a supergene architecture in which the distinct molecular components have been identified. All cases show switching of a complex trait involving both male and female effects, underlain by inheritance of a multi-allele, multi-locus composite locus. In many cases, the molecular functioning of the switch has been identified.

Less well characterised are the heteromorphic systems, such as that in primroses (genus *Primula,* Primulaceae), which consist of the production of two flower types, pin and thrum. Pin flowers have an elongated style with the stamens set deep within the flower, whereas thrum flowers have a short style and the stamens are much higher within the flower. The divergent floral morphs reduce the chance of a flower selfing and are controlled by a single locus, *S*. Some species show a breakdown in heterostyly, producing self-fertile homostyle individuals (reviewed for *P. vulgaris* by Bodmer, 1960). When crossed experimentally, these homostyles behave as though they are the result of recombination within the *S* locus, suggestive of a multilocus supergene (Wedderburn and Richards, 1992). A linkage map of *P. sieboldii* has been constructed and used to map quantitative traint loci (QTLs) for heterostyly traits, many of which map to the *S* locus, although there are several unlinked QTLs, analogous to modifier loci from other supergene systems (Yoshida *et al.*, 2011). Work has also been conducted to produce a physical map of the *S* locus: Li *et al.* (2011) produced several *S*-linked BAC tile paths, together spanning 2.2 Mb of *S*-linked sequence. In summary, there is an increasingly detailed physical and genetic map of the *Primula S*, although functional sites are yet to be identified.

Heterostyly has arisen convergently in the genus *Turnera,* with some species exhibiting distyly in addition to homomorphic SI (Labonne *et al.*, 2010). A series of experiments utilising hybrid crosses has uncovered a large *S*-linked region of over 1.6 megabases in three contigs containing several genes (Labonne and Shore, 2011), although it is not known whether these genes are functional in determining the SI phenotypes. X-ray-derived mutants show differing degrees of deletion of *S*-linked genes and a range of phenotypes affecting SI and flower morphology, indicating a multilocus architecture (Labonne *et al.*, 2010). Although flower morphs are thought to be inherited as a single unit, there is no statistically significant decrease in recombination in the *S* locus compared with other loci, arguing against a genomic architecture such as an inversion, although it has been suggested that the *S* locus may reside in a collinear region with reduced recombination, such as near a centromere (Labonne and Shore, 2011).

The striking fact that hetero- and homomorphic SI, and the various independently evolved mechanisms underlying these systems, all show the same type of genetic architecture of very tightly linked male- and female-determinant genes argues strongly for the reality of selection favouring a 'supergene' genetic architecture. However, a number of unanswered questions remain, especially with regard to the origins of these supergenes. Some authors (for example, Ford, 1964, p. 121; McKinnon and Pierotti, 2010) have argued that gene duplication may be a route to supergene formation; however, the fact that male and female genes are often unrelated (the glycoprotein/receptor kinase and S-protein in Brassicaceae, the S-RNase and F-box gene system of Solanaceae and so on) demonstrates that this is certainly not universally the case. The SI mechanisms considered here evolved early in the diversification of the various families; the supergenes are shared by distantly related genera and coalescence is deep in the ancestry of these plants. To gain insight into processes by which supergenes initially arise and the evolutionary pressures shaping their evolution, examples where the supergenes are younger and specific to individual species may be more informative. The classical examples of polymorphic mimetic butterflies fit this description, with mimetic species showing independent origins of supergene systems.

MIMICRY

Many of the classical examples of supergenes involve mimicry, which commonly leads to negative frequency-dependent selection, leading to a balanced polymorphism and the need for correlated expression of co-adapted characters. Classically, some of the best examples of supergenes were considered to be polymorphic mimetic butterflies (Ford, 1964). Although early work focused on the genus *Papilio*, the best studied example is the South American butterfly *Heliconius numata* (Figure 2a), in which a polymorphism in mimetic wing pattern is maintained by spatially heterogeneous selection for locally adapted Müllerian mimicry (Joron *et al.*, 2001). A large genomic region of around 200 kb shows extremely high linkage disequilibrium in wild populations implying reduced effective recombination (Joron *et al.*, 2011). Mapping of the wing pattern locus and genomic comparisons with related species have shown that this results from multiple chromosomal rearrangements that are associated with different wing patterns within *H. numata*. Furthermore, the region has arisen from several linked ancestral wing-pattern genes known in other *Heliconius* species, suggesting an ancestral architecture that predisposed this locus to evolution of a supergene (Ferguson *et al.*, 2010). Although it remains unclear what the functional elements are within the locus, *H. numata* provides a clear example of several essential elements for a 'classical' supergene—a selective regime that maintains a local polymorphism combined with evidence for reduced recombination between ancestrally linked functional elements.

The African Mocker Swallowtail, *Papilio dardanus* is a female-limited Batesian mimic in which a single locus, termed H, controls all aspects of the wing-pattern polymorphism. Intermediates between morphs are rare to non-existent, and clear dominance hierarchies exist within populations (Ford, 1964; Nijhout, 2003). The *H* locus has been physically mapped through linkage mapping and SNP-association analysis to a narrow region that includes the candidate genes *engrailed* and *invected* (Timmermans *et al.*, personal communication, Clark *et al.*, 2008). In contrast to the situation in *H. numata*, however, there is no clear evidence for a region of reduced recombination and only limited evidence for enhanced linkage disequilibrium. It seems possible therefore that this case is closer to the Nijhout hypothesis in which a single regulatory locus, with multiple downstream regulatory targets, controls patterning. *Engrailed* and *invected* are excellent candidates in this regard, although direct evidence for their involvement remains to be demonstrated.

Older studies of *P. dardanus* also highlight other aspects of the evolution of the supergene architecture. Thus, it has been shown that the dominance relationships between alleles break down in inter-race crosses, suggesting that unlinked modifier loci act epistatically to fine-tune the supergene locus (Clarke and Sheppard, 1962). In addition,
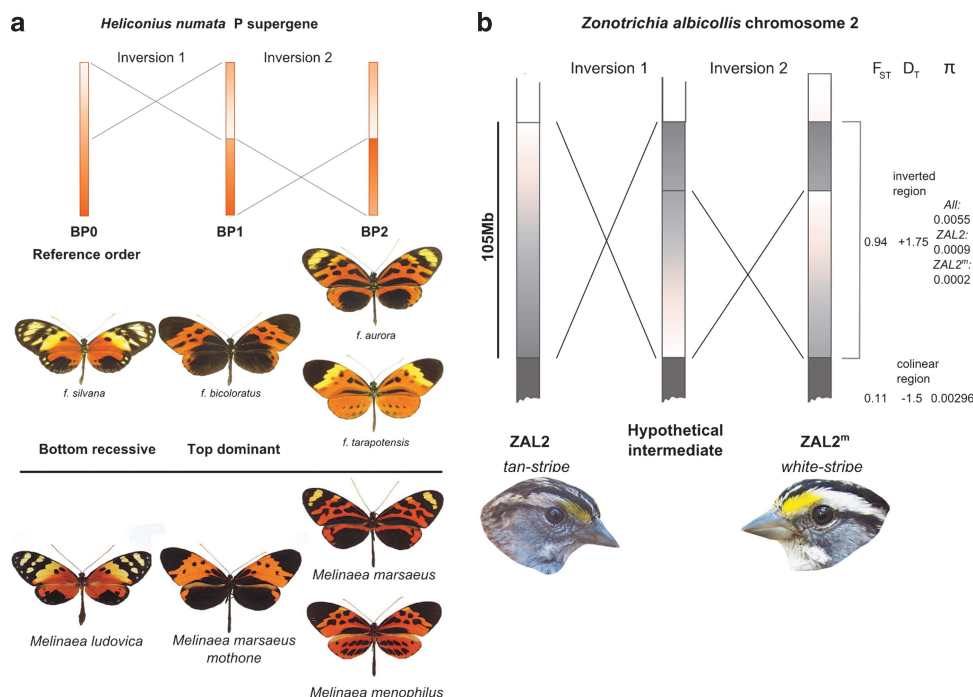
Figure 2 (**a**) Different forms of the butterfly *H. numata* are associated with different genome arrangements in the region of the P supergene. Forms of *H. numata* are shown above the horizontal line, with their corresponding models beneath. Photos from Mathieu Joron. (**b**) Representation of the rearrangements on chromosome 2 of *Zonotrichia albicollis* associated with reproductive dimorphism. The inversions cover c. 105 Mb. Population genetics figures from Thomas *et al.* (2008). Figures for collinear region are for the marker SUPT3H. Pi values for the inverted region are for both loci considered together, and each separately. Photographs modified from Figure 1 of Horton *et al.* (2012).

two populations have tailed females and the gene determining the presence/absence of tails is unlinked to the mimicry locus, *H*.

The remaining butterfly examples have not been studied using molecular methods. *P. memnon* and *P. polytes* are long-standing examples of polymorphic Batesian mimics, in which multiple morphs are inherited as a single trait. In both cases, a second locus determines the presence/absence of tails (Clarke *et al.*, 1968, 1971; Clarke and Sheppard, 1971, 1973), such that not all mimicry traits are inherited at the supergene. Some rare phenotypes (or geographically restricted forms) have been interpreted as recombinants, allowing the ordering of components within the supergene (Clarke *et al.*, 1968). These species were both studied extensively in the 20th century and are now excellent candidates for the application of modern molecular techniques.

Avian brood parasites offer another case of mimicry in which negative frequency-dependent selection exerted by hosts can maintain polymorphism, and where effective mimicry requires a correlated response in several phenotypic characters. In the case of Common Cuckoo, *Cuculus canorus*, host races (gentes) are adapted to particular hosts—for example, with egg and hatchling mimicry (Davies, 2011 and references therein). As males mate at random with respect to female host race (Gibbs *et al.*, 2000), some mechanisms are required to keep gens-specific adaptations together, and egg polymorphism may be W-linked. If this can be shown to be the case then this would then undoubtedly fulfil our criteria of a supergene.

Another well-studied brood parasite is the Greater Honeyguide, *Indicator indicator*. Honeyguides have two deeply diverged mitochondrial lineages, specialising on burrow- and hole-nesting hosts (Spottiswoode *et al.*, 2011). The available data suggest that honeyguide host races have been host-specific (ground versus tree-nesting lineages) for a long time, but that gene flow through males has prevented host races from speciating. The location of the host

specificity trait(s) is/are unknown; however, the results are consistent with either W-linkage or a sex-limited autosomal locus.

SOCIAL POLYMORPHISM
Hamilton (1964) proposed that altruism could evolve if altruistic behaviour were directed towards other carriers of the altruistic allele, later termed the 'green beard' hypothesis (Dawkins, 1976). Hamilton recognised that a supergene architecture was likely to be required to retain the tight association of the signal and behaviour. This idea remained theoretical until the recent discovery of examples such as *Solenopsis invicta*.

The Imported Red Fire Ant *S. invicta* has a social polymorphism affecting worker behaviour and queen reproductive strategies that maps to the odorant-binding protein gene *Gp-9*, and the linked gene *pgm-3* (Ross and Keller, 1998). *Gp-9* has been called a 'green beard' (Keller and Ross, 1998) due to its association with tolerance of multiple queens within a nest; all queens in polygynous nests are *Bb* heterozygotes at *Gp-9*—in these nests *BB* queens are killed, predominantly by *Bb* workers (*bb* is lethal). Selection favouring *b* due to its greater success through polygynous colony form is balanced by the lower fitness of *b* males and the lethality of *bb* homozygote females. The *B* and *b* alleles are characterised by a c. 13-Mb non-recombining region, including *Gp-9* and more than 615 other coding regions, many of which are known to be differentially expressed between ants of the two social forms (19 out of 27 total differentially expressed genes). The non-recombining region contains a 48-kb inversion that alters the expression of a candidate effector gene for cuticular odour differences between queens of the two forms (Wang *et al.*, 2013). The existence of a large non-recombining region determining alternative morphs is very similar to the role of the Y (or W) chromosome in sex

determination, and the situation seen in the reproductive dimorphism in *Zonotrichia albicollis*.

The white-throated sparrow *Z. albicollis* (Figure 2b) has a balanced reproductive dimorphism maintained by negative assortative mating. There are two morphs, '*tan-stripe*' and '*white-stripe*,' that show divergent plumage, behaviour and mate preference. The morphs show strong negative assortative mating with 96% of all pairings being between individuals of opposite morphs, and are associated with two alternative forms of chromosome 2, characterised by an extremely large (c. 104 Mb out of c. 110 Mb of chromosome 2) region of linkage disequilibrium and high levels of genetic divergence compared with collinear regions (Huynh *et al.*, 2010a), associated with one or more inversions (Thomas *et al.*, 2008; Huynh *et al.*, 2010b; Davis *et al.*, 2011). The existence of a partially non-recombining chromosome containing on the order of 1000 genes (Thomas *et al.*, 2008) and maintained in a balanced state by assortative mating offers a striking parallel to a sex chromosome and effectively represents a second and separate set of sex chromosomes. The genetic determination of reproductive morphs in *Z. albicollis* fulfils many of the criteria necessary for designation of a supergene: there is a mechanism to reduce recombination and a balanced polymorphism of a complex, polygenic and co-adapted phenotype.

OTHER ADAPTIVE COLOURATION

The guppy *Poecilia reticulata* displays a spectacular male-limited colour and size polymorphism. Colour polymorphism is manifested as multiple male morphs, probably determined by Y-linked loci (Lindholm *et al.*, 2004): Tripathi *et al.* (2009) detail karyotypically XX males that lack usual Y-linked colour and pattern traits. At a sub-chromosomal scale, the colour pattern loci map to a pseudo-autosomal region on the Y (Tripathi *et al.*, 2009), suggesting a possible role for recombination with the X as a source of novel male patterns: female colour pattern loci map to the X and there is recombination between colour patterns on X and Y (Khoo *et al.*, 1999). The guppy illustrates a difficulty in separating sex-linked traits and true supergenes; this case could be considered a supergene due to the complex, multi-trait nature of the male polymorphism when compared with simpler sex-linked traits.

The snail *Cepaea nemoralis* displays a shell-pattern polymorphism in terms of colour and the presence and number of contrasting stripes. The determination of shell pattern is genetic (Cain *et al.*, 1968), and polymorphism is known to be maintained in populations over long periods of time (Silvertown *et al.*, 2011). Evidence for the selective forces maintaining the polymorphism is mixed, with several selective factors having a role (Studies on Cepaea, 1968), including differential predation, climatic selection, heterosis (Cook, 2007) or random processes (Bellido *et al.*, 2002). Controlled laboratory crosses indicate strong linkage between colour and pattern loci, and recent molecular analysis has used RAD-seq to examine segregation of shell colouration (Richards *et al.*, 2013). There was no recombination between the colour and banding loci in 323 individuals from five segregating broods. The RAD-seq analysis found 44 markers potentially linked to the pattern-determining region of the genome and led to the construction of a 35.8-cM genetic map of the putative supergene region. Overall, despite decades of intensive study, both the origins and maintenance of the putative supergene in *Cepaea* remain unclear.

CONCLUDING REMARKS

From the very first attempts to integrate the laws of genetics with the theory of evolution, the adaptive importance of linkage between loci has been realised. Physical linkage reduces the probability that particular trait combinations are rearranged by recombination to give offspring with novel combinations. Where different loci define co-adapted characters, increasing linkage is advantageous as it maintains traits in favourable combinations. There is now clear evidence that balanced polymorphisms can result in tightly linked functional elements and in some cases for the evolution of genetic architectures that reduce recombination between functional elements (*H. numata* butterflies, *Solenopsis* fire ants and sparrows). Consistent with early theoretical predictions, none of the case studies that have been analysed at a molecular level have involved long-distance translocations of functional elements around the genome. It seems therefore that the supergene architectures studied to date have involved a combination of 'Turner's sieve' and local genomic rearrangement in order to further reduce recombination between already linked elements. In some cases, genomic architectures may evolve, which result in a cessation of recombination, allowing alternative haplotypes to further differentiate within a particular region, but allowing gene flow across the rest of the genome (*Zonotrichia albicollis*). Despite criticism of the usefulness of the supergene concept in the face of rapidly changing views of the term gene, we believe that supergene control of complex traits is widespread and study of supergene evolution has the potential to offer insight into convergent evolution of particular architectures and the role that the genome structure can have in adaptation.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Barton NH (2000). Genetic hitchhiking. *Phil Trans Roy Soc Lond B* **355**: 1553–1562.
Bellido A, Madec L, Arnaud J, Guiller A (2002). Spatial structure of shell polychromatism in populations of *Cepaea nemoralis*: new techniques for an old debate. *Heredity* **88**: 75–82.
Bodmer WF (1960). The genetics of homostyly in populations of *Primula vulgaris*. *Phil Trans Roy Soc B* **242**: 517–549.
Cain AJ, Sheppard PM, King JMB (1968). The genetics of some morphs and varieties of *Cepaea nemoralis* L. *Phil Trans Roy Soc B* **253**: 383–396.
Chantha S-C, Herman AC, Platts AE, Vekemans X, Schoen DJ (2013). In: Nasrallah JB (ed.) Secondary evolution of a self-incompatibility locus in the brassicaceae genus *Leavenworthia*. *PLoS Biol* **11**: e1001560.
Charlesworth D, Charlesworth B (1975a). Theoretical genetics of Batesian mimicry I. Single locus models. *J Theor Biol* **55**: 293–303.
Charlesworth D, Charlesworth B (1975b). Theoretical genetics of Batesian mimicry III. Evolution of dominance. *J Theor Biol* **55**: 325–337.
Charlesworth D, Charlesworth B (1975c). Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *J Theor Biol* **55**: 305–324.
Charlesworth D, Charlesworth B (1979). Selection on recombination in clines. *Genetics* **91**: 581–589.
Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW *et al.* (2012). Ecological genomics of anopheles gambiae along a latitudinal cline in cameroon: a population resequencing approach. *Genetics* **190**: 1417–1432.
Clark R, Brown SM, Collins SC, Jiggins CD, Heckel DG, Vogler AP (2008). Colour pattern specification in the Mocker swallowtail *Papilio dardanus*: the transcription factor *invected* is a candidate for the mimicry locus H. *Proc Roy Soc Lond B* **275**: 1181–1188.
Clarke CA, Clarke FMM, Gordon IJ (1995). Mimicry and other controversial topics in east African Lepidoptera. *J East African Nat Hist* **84**: 3–18.
Clarke CA, Sheppard PM (1962). Disruptive selection and its effect on a metrical character in the butterfly *Papilio dardanus*. *Evolution* **16**: 214–226.

8

Clarke CA, Sheppard PM (1971). Further studies on the genetics of the mimetic butterfly Papilio memnon L. *Phil Trans Roy Soc Lond B* **263**: 35–70.

Clarke CA, Sheppard PM (1973). The genetics of four new forms of the mimetic butterfly Papilio memnon L. *Proc Roy Soc Lond B* **184**: 1–14.

Clarke CA, Sheppard PM, Thornton IW (1968). The genetics of the mimetic butterfly Papilio memnon L. *Phil Trans Roy Soc Lond B* **263**: 35–70.

Clarke CA, Sheppard PM, Thornton IW (1971). The genetics of the mimetic butterfly Papilio polytes L. *Phil Trans Roy Soc Lond B* **263**: 35–70.

Cook LM (2007). Heterosis in Cepaea. *Biol J Linn Soc* **90**: 49–53.

Darlington CD, Mather K (1949). *Elements of Genetics*. George Allen & Unwin Ltd: London.

Davies NB (2011). Cuckoo adaptations: trickery and tuning. *J Zool* **284**: 1–14.

Davis JK, Mittel LB, Lowman JJ, Thomas PJ, Maney DL, Martin CL et al. (2011). Haplotype-based genomic sequencing of a chromosomal polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *J Hered* **102**: 380–390.

Dawkins R (1976). *The Selfish Gene*. Oxford University Press: Oxford.

Dobzhansky T (1947). Adaptive changes induced by natural selection in wild populations of *Drosophila*. *Evolution* **1**: 1–16.

Dobzhansky T (1970). *Genetics of the Evolutionary Process*. Columbia University Press: New York, NY, USA.

Dobzhansky T, Sturtevant AH (1938). Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* **23**: 28–64.

Van Doorn GS, Dieckmann U (2006). The long-term evolution of multilocus traits under frequency-dependent disruptive selection. *Evolution* **60**: 2226–2238.

Entani T, Iwano M, Shiba H, Che F-S, Isogai A, Takayama S (2003). Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes Cells* **8**: 203–213.

Feder JL, Roethele JB, Filchak KE, Niedbalski J, Romero-Severson J (2003). Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163**: 939–953.

Ferguson LC, Lee SF, Chamberlain NL, Nadeau NJ, Joron M, Baxter SW et al. (2010). Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol Ecol* **19**(Suppl 1): 240–254.

Fisher RA (1930). *The Genetical Theory of Natural Selection*. Oxford University Press: Oxford.

Foote H, Ride J (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc Natl Acad Sci USA* **91**: 2265–2269.

Ford EB (1964). *Ecological Genetics*. Methuen: London.

Ford EB (1965). *Genetic Polymorphism*. Faber & Faber: London.

Gibbs HL, Sorenson MD, Marchetti K, Brooke MD, Davies NB, Nakamura H (2000). Genetic evidence for female host-specific races of the common cuckoo. *Nature* **407**: 183–186.

Hamilton WD (1964). The genetical evolution of social behaviour. II. *J Theor Biol* **7**: 17–52.

Hermann K, Klahre U, Moser M, Sheehan H, Mandel T, Kuhlemeier C (2013). Tight genetic linkage of prezygotic barrier loci creates a multifunctional speciation island in *Petunia*. *Curr Biol* **23**: 873–877.

Horton BM, Hauber ME, Maney DL (2012). Morph matters: aggression bias in a polymorphic sparrow. *PLoS One* **7**: e48705.

Huynh LY, Maney DL, Thomas JW (2010a). Contrasting population genetic patterns within the white-throated sparrow genome (*Zonotrichia albicollis*). *BMC Genet* **11**: 96.

Huynh LY, Maney DL, Thomas JW (2010b). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *Heredity* **106**: 537–546.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR et al. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**: 203–206.

Joron M, IANR Wynne, Lamas G, Mallet J (2001). Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evol Ecol* **13**: 721–754.

Keller L, Ross KG (1998). Selfish genes: a green beard in the red fire ant. *Nature* **251**: 573–576.

Khoo G, Lim T, Chan W, Phang V (1999). Linkage analysis and mapping of three sex-linked color pattern genes in the guppy, *Poecilia reticulata*. *Zool Sci* **16**: 893–903.

Kimura M (1956). A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**: 278–287.

Kirkpatrick M, Barton NH (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Kopp M, Hermisson J (2006). The evolution of genetic architecture under frequency-dependent disruptive selection. *Evolution* **60**: 1537–1550.

Kubo K, Entani T, Takara A, Wang N, Fields AM, Hua Z et al. (2010). Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science* **330**: 796–799.

Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME (2001). Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* **13**: 627–643.

Labonne JJD, Shore JS (2011). Positional cloning of the s haplotype determining the floral and incompatibility phenotype of the long-styled morph of distylous *Turnera subulata*. *Mol Genet Genomics* **285**: 101–111.

Labonne JJD, Tamari F, Shore JS (2010). Characterization of X-ray-generated floral mutants carrying deletions at the S-locus of distylous *Turnera subulata*. *Heredity* **105**: 235–243.

Li W, Chetelat RT (2010). A pollen factor linking inter- and intraspecific pollen rejection in tomato. *Science* **330**: 1827–1830.

Li J, Webster Ma, Smith MC, Gilmartin PM (2011). Floral heteromorphy in *Primula vulgaris*: progress towards isolation and characterization of the S locus. *Ann Bot* **108**: 715–726.

Lindholm AK, Brooks R, Breden F (2004). Extreme polymorphism in a Y-linked sexually selected trait. *Heredity* **92**: 156–162.

Lowry DB, Willis JH (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* **8**: pii: e1000500.

Mather K (1950). The genetical architecture of heterostyly in *Primula sinensis*. *Evolution* **4**: 340–352.

Mather K (1955). Polymorphism as an outcome of disruptive selection. *Evolution* **9**: 52–61.

McKinnon JS, Pierotti MER (2010). Colour polymorphism and correlated characters: genetic mechanisms and evolution. *Mol Ecol* **5**101–5125.

Murfett J, Atherton T, Mou B, Gassert C, McClure B (1994). S-RNase expressed in transgenic Nicotiana causes S-allele-specific pollen rejection. *Nature* **367**: 563–566.

Nasrallah ME, Liu P, Nasrallah JB (2002). Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* **297**: 247–249.

Nei M, Kojima K, Schaffer HE (1967). Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* **57**: 741–750.

Nijhout HF (1991). *The Development and Evolution of Butterfly Wing Patterns*. Smithsonian Institution Scholarly Press: Washington.

Nijhout HF (1994). Developmental perspectives on evolution of butterfly mimicry. *Bioscience* **44**: 148–157.

Nijhout HF (2003). Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. *Evol Dev* **5**: 579–592.

Punnett RC (1905). *Mendelism*, 1st edn. Macmillian and Company: London.

Punnett RC (1915). *Mimicry in Butterflies*. Cambridge University Press: Cambridge.

Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A (2013). RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Mol Ecol* **22**: 3077–3089.

Ross KG, Keller L (1998). Genetic control of social organization in an ant. *Proc Natl Acad Sci USA* **95**: 14232–14237.

Sijacic P, Wang X, Skirpan A, Wang Y (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429**: 2–5.

Silvertown J, Cook LM, Cameron R, Dodd M, McConway K, Worthington J et al. (2011). Citizen science reveals unexpected continental-scale evolutionary change in a model organism. *PLoS One* **6**: e18927.

Spottiswoode CN, Stryjewski KF, Quader S, Colebrook-Robjent JFR, Sorenson MD (2011). Ancient host specificity within a single species of brood parasitic bird. *Proc Natl Acad Sci USA* **108**: 17738–17742.

Stern DL, Orgogozo V (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* **62**: 2155–2177.

Studies on Cepaea (1968). *Phil Trans Roy Soc Lond B* **253**: 383–595.

Takayama S, Isogai A (2005). Self-incompatibility in plants. *Annu Rev Plant Biol* **56**: 467–489.

Thomas JW, Cáceres M, Lowman JJ, Morehouse CB, Short ME, Baldwin EL et al. (2008). The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics* **179**: 1455–1468.

Tripathi N, Hoffmann M, Weigel D, Dreyer C (2009). Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (*Poecilia reticulata*). *Genetics* **182**: 365–374.

Tripathi N, Hoffmann M, Willing E-M, Lanz C, Weigel D, Dreyer C (2009). Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. *Proc Biol Sci* **276**: 2195–2208.

Turner JRG (1967). On supergenes. I. The evolution of supergenes. *Evolution* **101**: 195–221.

Ushijima K, Sassa H, Tao R, Yamane H, Dandekar AM, Gradziel TM et al. (1998). Cloning and characterization of cDNAs encoding S-RNases from almond (*Prunus dulcis*): primary structural features and sequence diversity of the S-RNases in Rosaceae. *Mol Gen Genet* **260**: 261–268.

Venail J, Dell'olivo A, Kuhlemeier C (2010). Speciation genes in the genus Petunia. *Phil Trans Roy Soc Lond B* **365**: 461–468.

Wang X, Hughes AL, Tsukamoto T, Ando T, Kao T (2001). Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (S) locus in *Petunia inflata*. *Plant Physiol* **125**: 1012–1022.

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D et al. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**: 664–668.

Wedderburn FM, Richards AJ (1992). Secondary homostyly in Primula L.; evidence for the model of the 'S'supergene. *New Phytol* **121**: 649–655.

West-Eberhard MJ (2003). *Developmental Plasticity and Evolution*. Oxford University Press: New York, NY, USA.

Wheeler MJ, de Graaf BHJ, Hadjiosif N, Perry RM, Poulter NS, Osman K et al. (2009). Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature* **459**: 992–995.

Wilkins KA, Bancroft J, Bosch M, Ings J, Smirnoff N, Franklin-Tong VE (2011). Reactive oxygen species and nitric oxide mediate actin reorganization and programmed cell death in the self-incompatibility response of papaver. *Plant Physiol* **156**: 404–416.

Yeaman S (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc Natl Acad Sci USA* **110**: E1743–E1751.

Yoshida Y, Ueno S, Honjo M, Kitamoto N, Nagai M, Washitani I et al. (2011). QTL analysis of heterostyly in *Primula sieboldii* and its application for morph identification in wild populations. *Ann Bot* **108**: 133–142.