MOLECULAR ECOLOGY

Molecular Ecology (2017) doi: 10.1111/mec.14051

INVITED REVIEWS AND SYNTHESES

Genetic architecture and balancing selection: the life and death of differentiated variants

VIOLAINE LLAURENS,* ANNABEL WHIBLEY† and MATHIEU JORON‡

*Institut de Systématique Evolution et Biodiversité (UMR 7205 CNRS, MNHN, UPMC, EPHE), Muséum National d'Histoire Naturelle - CP50, 45 rue Buffon, 75005 Paris, France, †Cell and Developmental Biology, John Innes Centre, Norwich Norfolk NR4 7UH, UK, ‡Centre d'Ecologie Fonctionnelle et Evolutive (UMR 5175 CNRS, Université de Montpellier, Université Paul Valéry Montpellier, EPHE), 1919 route de Mende, 34293 Montpellier, France

Abstract

Balancing selection describes any form of natural selection, which results in the persistence of multiple variants of a trait at intermediate frequencies within populations. By offering up a snapshot of multiple co-occurring functional variants and their interactions, systems under balancing selection can reveal the evolutionary mechanisms favouring the emergence and persistence of adaptive variation in natural populations. We here focus on the mechanisms by which several functional variants for a given trait can arise, a process typically requiring multiple epistatic mutations. We highlight how balancing selection can favour specific features in the genetic architecture and review the evolutionary and molecular mechanisms shaping this architecture. First, balancing selection affects the number of loci underlying differentiated traits and their respective effects. Control by one or few loci favours the persistence of differentiated functional variants by limiting intergenic recombination, or its impact, and may sometimes lead to the evolution of supergenes. Chromosomal rearrangements, particularly inversions, preventing adaptive combinations from being dissociated are increasingly being noted as features of such systems. Similarly, due to the frequency of heterozygotes maintained by balancing selection, dominance may be a key property of adaptive variants. High heterozygosity and limited recombination also influence associated genetic load, as linked recessive deleterious mutations may be sheltered. The capture of deleterious elements in a locus under balancing selection may reinforce polymorphism by further promoting heterozygotes. Finally, according to recent genomewide scans, balanced polymorphism might be more pervasive than generally thought. We stress the need for both functional and ecological studies to characterize the evolutionary mechanisms operating in these systems.

Keywords: adaptation, ecological genetics, evolutionary theory, frequency-dependent selection, polymorphism, population genetics – empirical

Received 12 July 2016; revision received 15 December 2016; accepted 19 December 2016

Introduction

Understanding how multiple functional variants may be maintained within populations is an important challenge for evolutionary biologists. The emergence and maintenance of such variation depend on the regimes of selection affecting the variants but also on the underlying genetic architecture, namely the number of genes involved, their effect sizes and their genetic linkage, and the coordination of the expression of their alleles (epistasis and dominance). Here, we review current knowledge on the genetic architecture of polymorphic traits. Investigating polymorphisms maintained within a population allows the simultaneous study of several adaptive variants within a common genetic background. This contrasts with studies focusing on traits under positive selection, where a single variant is usually

Correspondence: Mathieu Joron, Fax: +33 (0) 4 67 61 33 36;
E-mails: mathieu.joron@cefe.cnrs.fr, llaurens@mnhn.fr

observed within a population. We highlight how the persistence of polymorphisms maintained by balancing selection regimes generates specific population and genomic signatures.

Balancing selection usually describes any form of natural selection resulting in the persistence of multiple variants of a given trait at intermediate frequencies within populations. Traits for which the frequency spectrum contrasts with the frequencies expected by mutation–selection equilibrium are likely to be under a selection regime promoting polymorphism, as neutral alleles are unlikely to persist for long at intermediate frequencies within a population. Several selective mechanisms have been documented to promote balanced polymorphism, generally involving frequency-dependent processes, often negative frequency-dependent selection (referred to as NFDS hereafter). Balancing selection frequently results in the long-term persistence of alleles, which may be maintained through speciation events (e.g. see Takahata & Nei 1990 for MHC; Uyenoyama 1997 for S-locus in plants; van Diepen *et al.* 2013 for MAT-loci in fungi). NFDS also favours the introgression of new alleles via hybridization with closely related species. Similarly, NFDS tends to reduce population structure at the selected locus, because immigrant alleles are favoured by their low frequency in the recipient population (Leducq *et al.* 2011).

Overall, balancing selection preserves the coexistence of several adaptive variants for a given trait within population. This has important consequences for the genetic architecture of the traits involved and leads to the following theoretical predictions. First, when a trait is controlled by multiple loci, recombination tends to break up adaptive combinations of alleles: single-locus architecture or tight linkage between the loci involved (in extreme cases forming the so-called supergenes) is therefore expected in traits under balancing selection (e.g. Charlesworth & Charlesworth 1975). Structural rearrangements can play a role and inversions preventing recombination among loci underlying polymorphic traits are predicted to be favoured by balancing selection (Yeaman & Whitlock 2011). A second consequence of the persistence of polymorphism in sympatry is the high proportion of heterozygotes, which may promote the evolution of dominance at loci under balancing selection (Bourguet 1999), and may allow the accumulation of recessive deleterious mutations rarely expressed in heterozygous genotypes (van Oosterhout 2009a).

Here, we review current knowledge of balanced polymorphisms, from emergence to the decay of differentiation of the underlying alleles, and report how balancing selection regimes impact the genetic architecture governing trait variation. First, using evidence from a large number and broad taxonomic range of species, we document the selective mechanisms which promote polymorphism persistence within populations. Recent genome scans have detected a relatively high frequency of loci with signatures of balancing selection, implying that such selection regimes might be pervasive. We then follow the emergence and maintenance of these adaptive polymorphisms, demonstrating how specific genetic architectures may facilitate the persistence of differentiated variants. We finally show the current limitations in identifying causative mutations generating differentiation among adaptive variants and their relative impact on fitness.

Mechanisms underlying local polymorphism

Negative frequency-dependent selection (NFDS) is a straightforward driver of balanced polymorphism. It implies that the relative fitness of a variant increases when its frequency decreases, therefore protecting the polymorphism. NFDS is a widespread process often involved in traits where the fitness of variants depend on encounter rates, be that with alternative alleles, mates, or competitors, or with extrinsic factors such as resource type, pathogens or predators, which all inherently entail frequency-dependent dynamics.

Many traits involved in sexual recognition are influenced by NFDS, for instance when rare alleles and/or phenotypes benefit from greater reproductive success. The self-incompatibility locus (*S*-locus) in the *Brassicaceae* encodes both a pollen-coat protein and a corresponding pistil receptor; pollen can only germinate on pistil which does not express the receptor corresponding to its coat protein. Pollen expressing a rare haplotype at the *S*-locus benefits from increased mate availability and therefore a greater reproductive success. NFDS appears to promote the persistence of a large number of haplotypes at the *S*-locus and balances their frequency within populations (Llaurens *et al.* 2008; Stoeckel *et al.* 2008). Similarly, NFDS mechanisms generally promote polymorphism at loci underlying mating type in fungi, such as in MAT genes (van Diepen *et al.* 2013). However, the promotion of polymorphism by NFDS can be limited, because the advantage to rare alleles and the probability for a new allele to establish decrease as the number of alleles increases in the population (Gervais *et al.* 2011).

Similarly, different forms of disassortative mate preference also cause NFDS because rare phenotypes benefit from increased reproductive success. The disassortative mate preference of the white-throated sparrow (*Zonotrichia albicollis*), for instance, promotes plumage coloration and behavioural polymorphism (Thorneycroft 1975). A particular case of NFDS, referred to as rock-paper-scissor dynamics, has been described

in side-blotched lizards (*Uta stanburiana*) where three male morphs are maintained with oscillating frequencies, with each morph able to outcompete another one depending on their frequencies (Sinervo & Lively 1996).

Negative frequency-dependent selection in the narrow sense implies that the fitness associated with a given allele depends only on its frequency in the population, but such a mechanism is in fact often coupled with others. In the ruff (*Philomachus pugnax*), a wading bird whose males form competitive leks, three highly divergent male morphs persist: dark, aggressive and territorial 'independent' males defending small mating courts, white 'satellites' males codisplaying their feather collar but still competing with them, and finally rare 'faeder' males mimicking females. Faeders are likely to benefit from NFDS because their rarity allows them discreetly to access females, evading conflict. But the polymorphism of the three morphs seems to stem from a complex trade-off between sperm competition and male survival: faeders and satellites have poor survival compared with independents in crosses but are thought to benefit from increased fertilization success, as suggested by the large size of their testes (Kupper *et al.* 2016).

Negative frequency-dependent selection is also often observed in traits involved in competition for resources, such as the polymorphism in mandible orientation in crossbill finches feeding on pine cones: because pine cones are generally repeatedly visited by several individuals, birds with the rarer mandible orientation have better foraging efficiency (Benkman 1996). Polymorphism in behavioural traits can also be maintained by NFDS. For instance, the widespread *rover* and *sitter* allele polymorphism at the *foraging* locus in *Drosophila melanogaster* larvae has been demonstrated to be maintained by NFDS (Fitzpatrick *et al.* 2007). *Sitter* larvae tend to stay in the same food patch, whereas *rover* larvae move and usually explore more food patches; in conditions where competition for resources is high, each behaviour benefits from an increased fitness when rare. Similarly, polymorphism in foraging behaviour in *Caenorhabditis elegans* seems to be driven by density dependence: the major QTL controlling variations in foraging behaviour includes the gene *srx-43*, coding for a receptor protein sensitive to ascaroside pheromones, known to signal population density (Greene *et al.* 2016).

Many traits where learning processes play a role, for example in predators, prey, pollinators or even sexual partners, have been shown to result in NFDS. NFDS drives polymorphism in mouth orientation in scale-eating fish because of host behaviour: scale-eating fish attack their fish hosts preferentially on one side depending of mouth orientation, causing attacked fish to be more cautious to attacks from one particular side. Individuals with the rarer mouth orientation enjoy a better attack efficiency (Hori 1993). Polymorphism in prey cryptic patterns can also be maintained by NFDS, because visual predators tend to form search images for common patterns, so rare ones are less likely to be found by predators (Bond & Kamil 2002). In prey species displaying vivid colour patterns mimicking chemically defended species (Batesian mimicry), polymorphism is also common in the mimic and stems from NFDS. Indeed, the protection provided by predators learning to avoid the toxic species is weakened when palatable mimetic forms become common (Turner 1987). NFDS can also drive colour polymorphism in deceptive flowers, because deceived pollinators learn to avoid common nonrewarding flower morphs, and tend to visit rarer morphs more frequently (Smithson & Macnair 1997). NFDS also drives polymorphism in female damselflies where (male-mimicking) andromorph females may evolve in response to high harassment by males on gynomorphs (Takahashi *et al.* 2010). A similar female-limited polymorphism in colour pattern is observed within populations of *Drosophila erecta*, with a light morph and a male-like dark morph, probably maintained by NFDS due to male harassment; signals of balancing selection have been detected in the enhancer of the melanization gene *tan* controlling this phenotypic variation (Yassin *et al.* 2016a).

Heterozygote advantage (also referred to as overdominance) enhances polymorphism, because heterozygotes benefit from a higher fitness compared with homozygotes. Note that overdominance refers to fitter heterozygotes within a locus whereas hybrid vigour (or heterosis) describes improved fitness in hybrids among lineages and may involve several or many loci. Heterozygote advantage generates NFDS, because common alleles tend to occur as homozygotes more frequently than rarer alleles and therefore suffer from a reduced fitness. For instance in most social Hymenoptera, sex is determined by the diploidy at the complementary sex determiner (*CDS*) locus, haploids (single allele) developing as males and diploids (two alleles) developing as females. However, diploid homozygotes with two identical *CDS* alleles develop as inviable or infertile males, causing reduced colony fitness. NFDS therefore favours rare *CDS* alleles and promotes high polymorphism at the *CDS* locus, as recently demonstrated in the invasive honey bee species *Apis cerana* (Gloag *et al.* 2016). Many traits involved in resistance to parasites display heterozygote advantage because heterozygotes may recognize a larger range of pathogens than homozygotes. Aside from the classical example of the *S*-gene involved in sickle-cell disease and malaria resistance, overdominance has been suggested to be implicated in the evolution of immune-related genes such as *MEFV*, the gene responsible for familial

Mediterranean fever in humans (Fumagalli *et al.* 2009) and MHC genes (Kekalainen *et al.* 2009), especially in cases of multipathogen infections (Oliver *et al.* 2009). However, heterozygote advantage is also reported for other traits. One striking example is the balanced polymorphism in horn size in males of the Soay sheep, mainly controlled by a single gene *RXFP2* (Johnston *et al.* 2013). Homozygotes Ho^{+}/Ho^{+} at this locus have larger horns and greater reproductive success but reduced survival, whereas homozygotes Ho^{P}/Ho^{P} have smaller horns and greater survival but reduced reproductive success. In this case, the resulting trade-off between reproductive success and survival provides an overall greater fitness for heterozygotes Ho^{+}/Ho^{P}. Other trade-offs are reported to contribute to polymorphism, for example in the marine isopod *Paracerceis sculpta* where a trade-off between reproductive tenure and maturation time appears to maintain three different male morphs (Shuster & Wade 1991). Colour polymorphism in the meadow spittlebug *Philaenus spumarius* may also reflect a complex trade-off between survival and reproductive success between the different morphs (Silva *et al.* 2015). Finally, a case of double overdominance was described for insecticide resistance in the mosquito *Culex pipiens*, where resistance alleles with sublethal effects when homozygous have undergone two successive duplications, resulting in the persistence of four alleles (Labbe *et al.* 2007b).

Selection varying through time is also considered an important driver of balanced polymorphism. When the fitness of different alleles fluctuates in time, it may promote the persistence of polymorphism, for instance through red queen dynamics (Van Valen 1973), as demonstrated in host–parasite co-evolution where pathogen composition varies through time (Decaestecker *et al.* 2007). Other host–parasite interactions can also lead to NFDS. For instance, polymorphism in egg colour pattern is observed in bird species whose nests are regularly parasitized by cuckoos, because new egg phenotypes in the host species may escape parasitic mimicry (Spottiswoode & Stevens 2012).

Selection varying through space. Spatially heterogeneous selection is a ubiquitous mechanism allowing the persistence of polymorphism through migration/selection balance (Levene 1953), but whether high levels of polymorphism may be maintained within populations will depend on the grain of variation and the selective mechanisms at play. In the social bacterium *Myxococcus xanthus*, fruiting body formation excludes competitors, therefore creating a patchy environment and maintaining high polymorphism at a fine spatial scale (Rendueles *et al.* 2015). At larger scale, populations with different local conditions and connected by migration may maintain local polymorphisms for traits such as dispersal, as described for instance in plants (Lopez *et al.* 2008) or butterflies (Hanski *et al.* 2011).

Sexual antagonism is also a classic example of a mechanism promoting a diversity of alleles with antagonistic effects in individuals of different sexes. In gynodioecious plants such as *Plantago lanceolata*, both hermaphrodites and females are maintained within populations. This polymorphism is maintained by an evolutionary arms race between cytoplasmic male sterility (*CMS*) genes located in the mitochondrial genome and male fertility restorers located in the nuclear genome (Gouyon *et al.* 1991). However, the conditions required for those conflicts to resolve in the stable maintenance of polymorphisms may not be trivial. In bank voles, for instance, alleles favouring male mating success are detrimental to female fecundity, but the maintenance of polymorphism is greatly facilitated when male-benefiting alleles are themselves under NFDS (Mokkonen *et al.* 2011).

Intralocus sexual conflicts may also generate balancing selection. In *Drosophila* populations for instance, polymorphism at the gene *Cyp6g1*, identified for its role in DDT resistance, was maintained before DDT treatment introduction, through sexually antagonistic effects between female fecundity and male mating success (Rostant *et al.* 2015). Similarly, the gene *VGLL3* exhibits sex-dependent dominance in salmon, promoting alleles for earlier and later maturation in males and females, respectively, resulting in balanced polymorphism in age at maturity (Barson *et al.* 2015).

Various mechanisms allow the persistence of stable polymorphisms within populations. However in many cases where balanced polymorphism is observed, the underlying mechanisms remain unknown. For instance, the trans-specific persistence of the ABO blood group polymorphism in apes shows a clear signature of balancing selection, but it is still unclear why (Segurel *et al.* 2013). Theoretical predictions on the spatial and temporal distribution of allele frequencies under different balancing selection regimes are crucially needed to decipher and distinguish the selective processes involved in the maintenance of polymorphisms.

Detecting balancing selection in genomes

Despite the documentation of a few well-researched case studies where the adaptive traits have been extensively studied, the general importance of balancing selection as an evolutionary process is still unclear. However, in recent years mounting evidence has started to suggest that balancing selection might not be restricted to rare case studies (Key *et al.* 2014). Although genome scans have received much attention and may help the molecular identification of adaptive variants

(for instance, see Haasl & Payseur 2016 for a review), how to recover loci under balancing selection efficiently from genomic data sets is still poorly studied. Part of the difficulty arises because signatures of balancing selection can vary depending on the balancing mechanism and the genetic architecture of the trait. Most methods rely on departures from neutral expectations of the structure and distribution of genetic variation, but in directions which may be different from the signatures left by selective sweeps (a significant reduction in polymorphism in linked SNPs, for instance (Nielsen *et al.* 2005)). Generally, heterozygote excess, an excess of nonsynonymous mutations, an excess of common polymorphism, a lack of spatial genetic structure or trans-specific polymorphism, may indicate the persistence of balanced polymorphisms at a given locus (see Fijarczyk & Babik 2015 for a critical review). Crucially, the use of different detection criteria might impact findings as different mechanisms of selection can produce different molecular signatures. For instance, a lack of spatial structure, indicated by a locally depressed *Fst* compared with neutral markers, can be used to detect balancing selection associated with negative frequency-dependent selection (Weedall & Conway 2010), but an opposite trend is expected when the balanced polymorphism is promoted by spatially variable selection, with an above-average *Fst* predicted at the balanced locus. Similarly, some selective mechanisms can maintain polymorphism for long evolutionary periods, but such trans-specific polymorphisms require that the balancing mechanism is conserved in daughter species, which is not necessarily the case for all balancing selection regimes. Signatures of balancing selection have been detected in the human genome in 60 of 13 400 genes studied using statistics based on an excess of nonsynonymous mutations, Tajima's D or HKA tests (Andres *et al.* 2009). These tests detect significant departures of nucleotide differences and their frequency from neutral expectations: excesses of nonsynonymous mutations and intermediate frequencies are used as signatures of balancing selection. Integration of these statistics and the development of more sophisticated methods taking account intra- and interspecific divergence have led to the identification of a similarly high number of loci putatively under balancing selection: Leffler *et al.* (2013) detected 125 genomic regions showing signatures of trans-specific polymorphism in a genomewide scan of human and chimpanzees, highlighting an enrichment in immune-related regions. Using a composite likelihood method, DeGiorgio *et al.* (2014) successfully detected similar regions, including *HLA* genes, but also discovered new candidate genes such as *FANK1*, putatively involved in segregation distortion. A recent review confirmed trans-specific polymorphisms between apes and humans for loci putatively involved in immune reactions (Azevedo *et al.* 2015; e.g. polymorphism of ABO group, Segurel *et al.* 2012). Although most reported scans have been performed on the human genome, the recent development of robust statistical methods to detect genomic regions under balancing selection (Gao *et al.* 2015) promises to broaden our view on the importance of balancing selection in nonmodel organisms. Despite the increasing number of loci showing signatures of balancing selection, the underlying balancing mechanisms which protect polymorphism at each locus remain largely unknown. Investigating the fitness effects of the different variants is now needed to characterize the mechanisms maintaining variation. Empirical data on the temporal and spatial grain of variation in allelic composition for those loci may also enable deciphering mechanisms, such as temporal cycles suggesting NFDS for instance, while stably balanced frequencies would be more likely in loci exhibiting heterozygote advantage. Distinguishing between balancing mechanisms also requires a clearer understanding of the genomic signatures to be expected in the different cases.

Birth of new variants: how are highly differentiated variants formed?

One challenging question in evolutionary genetics is how the multiple changes required to generate a functionally new variant can accumulate despite of drift and/or negative selection exerted on intermediate stages. This question is particularly challenging for traits under balancing selection as new variants arise and co-occur with alternative variants, with ample opportunity for recombination between them. Three main mechanisms have been recurrently proposed to promote the emergence of differentiated phenotypes: first, the recruitment of major-effect mutations, second, gene duplication, and finally introgression of 'ready-made' alleles from hybridizing species.

Recruitment of large-effect mutations

Classical models propose that new adaptive alleles could be formed through a first mutation of large beneficial effect, followed by successive mutations of smaller effect which further hone the phenotype ('two step theory', (Poulton 1912)). For example, in species exhibiting polymorphic mimicry, an initial mutation might result in imperfect mimicry of a new model, with sufficient resemblance to the new model to confer some degree of protection from predators, and this mutation could be followed by successive mutations of smaller effect size, refining the mimetic phenotype (Nicholson 1927). However, for traits showing balanced polymorphisms,

recombination might prevent successive mutations from accumulating, which might restrict the emergence of new functional variants to cases where one or few mutational events in 'key' genes are sufficient to produce differentiated phenotypes.

Genes with many pleiotropic effects or with effects in multiple tissues during development ('developmental genes') are sometimes proposed as good candidates where point mutations could induce large phenotypic changes. Key genes identified as underlying polymorphic mimicry in butterflies indeed tend to have a large phenotypic effect and are involved in developmental pathways known to influence spatial identity in wing cells. For instance, the gene *engrailed*, a major transcription factor controlling morphogenesis in different tissues and stages, is associated with female-limited polymorphic mimicry in the *Papilio dardanus* (Timmermans *et al.* 2014). Similarly, the gene *doublesex,* well-known for its role in the sex-determination cascade as well as in sexually dimorphic morphologies in *Drosophila* and other insects, controls female-limited polymorphic mimicry in *Papilio polytes* (Kunte *et al.* 2014; Nishikawa *et al.* 2015) as well as in *Papilio memnon* (Komata *et al.* 2016). The recruitment of *doublesex* for sex-specific variations in wing colour pattern might be explained by its pre-existing sex-specific role. The repeated use of the same genes involved in different wing phenotypes in different species, as found in the Lepidoptera, may indicate that tweaking the regulatory regions of genes expressed in the target tissue could facilitate the rapid emergence of new adaptive variants (Martin *et al.* 2014; van't Hof *et al.* 2016; Nadeau *et al.* 2016; Wallbank *et al.* 2016). Similarly, the transcription factor gene *pdm3* has been repeatedly recruited in several species of the *Drosophila montium* subgroup in the control of female-limited pigmentation polymorphism, probably involving several independently evolved variants of its *cis*-regulatory region (Yassin *et al.* 2016b). By contrast, this evolution of similar genetic architectures contrasts with that found in *Drosophila erecta* where female-limited pigmentation polymorphism is controlled by an enhancer of the well-known melanization gene *tan* (Yassin *et al.* 2016a). In summary, it is unclear whether pleiotropic genes are recruited in the control of differentiated forms because their pleiotropy allows them to control multiple differences simultaneously. An alternative is that pleiotropic genes are especially likely to be expressed at a key developmental time and/or in relevant developing tissues where phenotypic differences may be controlled, making them repeated targets of balancing selection.

More generally, it is still unclear to what extent genes expressed in multiple different tissues or positioned at certain stages in development are more likely to evolve into polymorphic loci switching between differentiated phenotypic forms. One of the reasons may be that we still ignore much of the genetic underpinnings of modularity itself, that is how easily the recruitment of a gene positioned upstream in a developmental cascade may result in the co-option of its entire downstream effects when expressed in a new tissue. The expectations regarding the distribution of mutational effects at a given locus are also still poorly defined and we may find that the multiple elements participating to a composite locus or a supergene may be expected to repeatedly hit the same gene, or target genes with large mutational possibilities, for example those with long and complex *cis*-regulatory regions (e.g. genes *optix* and *doublesex* in butterflies, Martin *et al.* 2014).

Gene duplication

The emergence of divergent alleles characterized by the accumulation of several mutations affecting traits under balancing selection might be facilitated by gene duplications. Gene duplications frequently arise as a result of errors in DNA replication and are thought to be an important source of genetic variation. In the case of heterostyly in *Primula vulgaris*, the switch between pin (long style/short anthers) and thrum (short style/long anthers) morphs is now known to be derived from the duplication of a floral homeotic gene, followed by a neo-functionalization of one copy allowing the emergence of the thrum phenotype (Li *et al.* 2016). Several other documented traits under balancing selection, in particular those involved in pathogen recognition, display frequent gene duplications (*R*-genes in plants (McDowell & Simon 2006), *MHC* in Vertebrates (Piertney & Oliver 2006), resulting in large copy number variations: for instance, the number of MHC genes is highly variable across species (Mehta *et al.* 2009), but also within species (Bonhomme *et al.* 2008) and populations (Eimes *et al.* 2011). Copy number variations can result from birth and death processes well described in multigene families (Nei & Rooney 2005), where duplicated copies can be promoted by natural selection and subsequently lost because of drift or changes in adaptive value. In the evolutionary arms race between hosts and pathogens, new MHC gene copies are indeed regularly promoted when they recognize frequently encountered pathogens, but some copies are also frequently lost because they are no longer adapted to current pathogen communities. The presence of several copies creates a range of variation which can be neutral most of the time but may sometimes allow a rapid response to the invasion of new pathogens. However, gene

conversion, by allowing genetic exchange among the different copies, might favour rapid adaptations to new pathogens but may also homogenize variations among copies, and therefore limit divergence. Furthermore, in some cases, duplication can erase heterozygote advantage by creating a superior haplotype with both alleles in tandem, as observed in the insecticide resistance gene *ace-1* in the mosquito *Culex pipiens* (Labbe *et al.* 2007a). Duplications can therefore facilitate the emergence of differentiated variants, but may also limit within-locus polymorphism.

Introgression

Another source of highly divergent adaptive variants highlighted in balanced polymorphisms is the introgression of new haplotypes favoured by selection. This is illustrated by the recent adaptive introgression of a MHC allele from domestic goat to *Alpine Ibex* (Grossen *et al.* 2014), or the multiple introgressions of alleles from the self-incompatibility locus in the closely related species *Arabidopsis halleri* and *A. lyrata* (Castric *et al.* 2008). In the white-throated sparrow, the *white* haplotype, an inversion-based supergene allele of >100 Mb which confers differences in pigmentation and components of social behaviour compared with the alternate *white* haplotype, was also recently suggested to be introgressed from a closely related species (Tuttle *et al.* 2016). The strength of balancing selection acting on alternative variants probably favours the capture of new haplotypes already shaped by selection in isolation in a related lineage. Introgression thus appears as a potentially important driver of adaptation, allowing the recruitment of variants carrying multiple co-adapted mutations. Adaptive introgression is also particularly likely when NFDS operates on new alleles which benefit from their rarity upon entering a new population.

Life of adaptive alleles: genetic architecture controlling adaptive polymorphism

By maintaining several variants for a given trait, balancing selection shapes the underlying genetic architecture, in terms of number of loci involved, recombination rates, dominance between alleles and perhaps even the chromosomal location (see Box 1).

In many cases, traits under balancing selection require the coordinated action of several genes, and the adaptive combinations of alleles at those genes may be continuously destroyed by recombination occurring every generation. Therefore, genetic architectures limiting recombination, for instance through close physical linkage of these genes or through chromosomal inversions, are predicted to be favoured under balancing selection. Loci located close to each other on the same chromosome possess a limited recombination risk; similarly, chromosomal inversions, by preventing chiasmata formation between inverted segments in heterozygotes, limit crossing-over and therefore recombination among genes captured within the inversion.

In many cases of local adaptation with gene flow, selection in favour of a higher linkage between locally adapted alleles is expected to increase with the intensity of migration between populations, because migration tends to break successful combinations of locally adapted alleles. For instance, the well-documented inversion *3RP* in *Drosophila melanogaster* is associated with several independent latitudinal gradients (Rane *et al.* 2015), suggesting that the linkage imposed by inversions might allow a more efficient adaptive response to climatic gradients. Conversely, recombination also permits the formation of new allelic combinations and therefore may promote the emergence of adaptive variants. Theoretical approaches investigating the evolution of recombination by tracking down the success of recombination modifiers suggest that recombination enhancers could be favoured by spatial heterogeneity in selection, because they allow the recurrent formation of new adaptive combinations (Lenormand & Otto 2000).

Balanced multilocus architecture

Few cases of balancing selection involving a multilocus architecture have been reported. Our current knowledge might be limited by the theoretical and empirical challenges encountered in the identification of balancing selection acting on multiple loci. Indeed, the selective conditions required for the persistence of polymorphism at multiple additive loci might be restrictive, but are theoretically possible in certain cases of spatially or temporally variable selection (Turelli & Barton 2004). For instance, if the environmental conditions where individuals develop are largely unpredictable, then variation may be maintained at several unlinked loci by habitat heterogeneity. This seems to be the case in the American eel, which behaves as a single panmictic population maintaining considerable polymorphism at many loci (Cote *et al.* 2013). Adaptive variation is preserved by strong selection on multiple traits involved in ecological adaptation in the juvenile forms, in the varied coastal and riverine habitats where they develop. In this marine species, juveniles disperse from the breeding waters into all freshwater environments found along the coast from subtropical to temperate conditions, which filter different allelic combinations at multiple loci (Pavey *et al.* 2015). Adults of all ecotypes return to a single shared breeding area where the different adaptive variants mix. In this species, both spatially heterogeneous selection

Box 1. Traits displaying highly differentiated variants

Many traits under balancing selection display highly differentiated variants, with several epistatic loci controlling classically assumed to control a different features of the trait. For instance, in the land snail species *Cepaea nemoralis*, shell appearance is polymorphic and involves both colour and the number of bands, controlled by the loci *C* and *D*, respectively. Genomic mapping of these loci based on RAD markers revealed that these two loci cosegregate, therefore suggesting a supergene architecture (Richards *et al.* 2013).

Heterostyly in *Primula* can also be described as a highly differentiated trait, with a minimum of three tightly linked loci involved forming a supergene: *G* controlling the length of the style (short *vs.* long), *A* controlling the position of the anthers and *P* the size of pollen grains (Lewis & Jones 1992). The dominance of alleles appears to be coordinated, the dominant allele *G* encoding for short style being associated with the dominant allele *A* encoding for high anthers.

Similarly in the Batesian mimic *Papilio memnon*, several loci are involved in the mimicry of females towards chemically defended species (Fig. 1). For instance, locus *T* (controlling tail presence/absence), locus *W* (controlling location of white patterning in the hindwing) and locus *B* controlling body colour (black *vs.* rayed yellow) exhibit a linkage disequilibrium and might therefore belong to a supergene (Clarke *et al.* 1968; Clarke & Sheppard 1971).

Fig. 1 Example of *Papilio memnon* female morph diversity (courtesy of the National Natural History Museum in Paris). Top row: *P. memnon agenor hiera*, *P. memnon agenor vimus*, *P. agenor butlerianus*. Bottom row: *P. memnon achates*, *P. memnon achates*, *P. memnon achates alcanor*. Note correlations among the phenotypes shown by the arrows: loci T (controlling presence vs. absence of tail), B (black vs. yellow body colour) and W (brown vs. white hindwing colour pattern) (see Clarke *et al.* 1968).

Pollination, domestication or behavioural syndromes are also relevant examples of traits combining multiple developmentally distinct features. For instance, domestication syndromes in mammals include a suite of behavioural, physiological and morphological traits not observed in their wild counterparts, such as for instance increased docility and tameness, coat colour changes and nonseasonal oestrus cycles (Wilkins *et al.* 2014). Pollination syndromes can involve variation in petal shape and colour, as well as shape of reproductive organs as well as odours, and selection acts on trait combinations acting on pollinators' attractions, as investigated in hummingbirds pollinating *Silene* flowers (Fenster *et al.* 2015). Behavioural syndromes describe suites of correlated behaviours, like in the temperate comb-footed spider, *Anelosimus studiosus* where two distinct types of behaviours coexist: (1) individuals defending their asocial nests against intrusion by conspecifics or (2) individuals cooperating in multifemale nests. These two types are associated with a series of different behaviours, (response to predators and prey, degree of superfluous killing, exploratory behaviour and general level of activity) which may stem from either pleiotropic effects or linkage between several traits (Pruitt *et al.* 2008), and exemplifying the high differentiation maintained within natural populations.

It should be noted that even in traits controlled by a single gene, coordinated variations of several independent genetic elements are required. In the famous example of the gene *shavenbaby*, controlling for trichome pattern differences among *Drosophila* species, McGregor *et al.* (2007) demonstrated variations in three different *cis*-located enhancers are collectively required to make the *D. sechellia* trichome pattern different from *Drosophila melanogaster*. This highlights that epistatic interactions are also an important feature of the genetic architecture of trait variations.

and complete panmixia favour the persistence of a high level of polymorphism at multiple adaptive loci. Selection in different directions is exerted on multiple traits involved in adaptation to temperature, pollution or salinity which are not necessarily correlated in all environments. This multilocus architecture of local adaptation allows the decoupling of the different traits and might be a specificity of the migration and reproduction behaviour observed in this species.

Is single-locus architecture promoted by balancing selection?

Adaptive polymorphisms are frequently controlled by simple, single-locus architectures locking together successful mutations with epistatic benefits. Distinct adaptive variants are indeed frequently formed by several mutations with epistatic interactions. Linkage among functional variants may be advantageous in loci under positive selection (Kirkpatrick & Barton 2006), because it allows the joint recruitment of several epistatic loci. However under directional selection, local adaptation may also be permitted by the successive fixation of adaptive alleles at different loci because of limited gene flow. Contrary to traits under positive selection where new adaptive combinations may rapidly replace ancestral ones, in traits under balancing selection, several alternative combinations may be maintained at relatively high frequencies, providing ample opportunity for recombination to break down successful combinations of mutations. Selection favouring the persistence of adaptive polymorphism may therefore promote tighter linkage among the targeted functional sites. For instance, at the *S*-locus controlling self-incompatibility in *Arabidopsis halleri* and *A. lyrata*, recombination is indeed suppressed between epistatic genes controlling pollen protein and pistil receptors, and quickly returns to the genomewide level in flanking regions (Roux *et al.* 2013), suggesting a strong role for selection on recombination limitation.

Few empirical cases allow deciphering the evolution of genetic architecture by contrasting different selection regimes. In butterflies from the genus *Heliconius*, most species experience strong directional selection towards mimicry of the wing pattern providing the best protection locally (Brown 1981), resulting in a geographical mosaic of wing pattern races following the distribution of multispecies mimicry communities (Merrill *et al.* 2015). The genetic architecture underlying the switch between geographical races is controlled by several well-identified loci (Reed *et al.* 2011; Martin *et al.* 2012; Huber *et al.* 2015; Nadeau *et al.* 2016). In contrast, locally polymorphic butterflies maintaining several mimetic forms within populations, such as *Heliconius numata* or several *Papilio* species, display a single-locus architecture controlling wing pattern variation (Clarke & Sheppard 1959, 1972; Clarke *et al.* 1968; Joron *et al.* 2006). In Batesian mimicry, single-locus architecture may be beneficial, otherwise recombination between different mimetic alleles would destroy adaptive combinations of mutations and generate intermediate phenotypes suffering from decreased fitness because of poor mimicry. In the unpalatable species *H. numata*, migration-selection equilibrium in a fine spatial mosaic of mimicry communities may be sufficient to generate balancing selection on wing coloration (Joron *et al.* 1999; Chouteau *et al.* 2016). In contrast to the supergene architecture observed in *H. numata*, mimicry phenotypes in the sister species *H. hecale* and *H. ismenius* are under local directional selection and are associated with oligogenic control (Huber *et al.* 2015). However, although balanced polymorphism may be associated with single-locus architecture here, which one evolved first and promoted the other is not clear.

Chromosomal inversions associated with balancing selection

Supergenes can be defined as genetic architectures involving multiple linked functional genetic elements that allow switching between discrete, complex phenotypes maintained in a stable local polymorphism (Thompson & Jiggins 2014). This architecture has long been associated with the evolution of adaptive polymorphisms (e.g. population genetic model predicted a supergene architecture in polymorphic mimicry Charlesworth & Charlesworth 1975), and more recent theory suggests supergenes should evolve in the general context of balancing selection (Yeaman 2013). The nature and content of several supergenes controlling adaptive polymorphism have been recently uncovered in a variety of taxa (Thompson & Jiggins 2014) and are in most cases associated with chromosomal rearrangements (see Schwander *et al.* 2014 for a review). Examples are not confined to the original pin-and-thrum flower

Table 1 Comparative summary of documented cases of adaptive polymorphisms involving differentiated alleles.

Complex traits	Species	Balancing selection mechanisms	Underlying loci	Causative genes	Mechanism preventing recombination	Dominance relationships	Associated genetic load	References
Mimetic wing colour pattern	*Heliconius numata*	Spatially heterogeneous selection	*Supergene P – 31 genes*	Unknown number and identity	400 kb inversion	Inverted > Ancestral; Coordination of dominance among pattern elements for sympatric alleles.	Unknown (deficit of dominant homozygotes in the wild)	Chouteau, M. (unpublished data); Joron *et al.* (2011), Le Poul *et al.* (2014)
Mimetic wing colour pattern	*Papilio polytes*	Negative frequency-dependent selection	*Locus H – 5 genes*	*Doublesex*	300 kb inversion	Inverted > Ancestral	Unknown	Kunte *et al.* (2014)
Mimetic wing colour pattern	*Papilio dardanus*	Negative frequency-dependent selection	*Locus H*	*Engrailed*	Duplication of unknown size associated with *lamborni* allele	Derived > Ancestral Duplication is intermediate dominant.	Unknown (nonsynonymous changes in key gene)	Timmermans *et al.* (2014), Thompson *et al.* (2014)
Colonies' social organization	*Solenopsis invicta*	Overdominance	Supergene *SB* – 616 genes	Unknown number and identity	9 Mb inversion	Inverted > Ancestral	Recessive lethal	Wang *et al.* (2013)
Colonies' social organization	*Formica selysi*	Unclear	Supergene *Sm*	Unknown number and identity	Recombination suppressed in a 183 cM region by an unknown mechanism	*Sp* > *Sm*	No lack of heterozygous, suggesting no associated load	Purcell *et al.* (2014)
Mating type	*Arabidopsis halleri and lyrata*	Negative frequency-dependent selection	S-locus - 3 genes	Pollen-coat protein *SCR* and pistil receptor *SRK*	Variations in gene order	Overall consistent dominance hierarchy between *SCR* and *SRK* alleles.	Partially recessive genetic load	Llaurens *et al.* (2009b), Stift *et al.* (2013)
Pathogen and self-recognition	*Homo sapiens and other vertebrate species*	Fluctuating selection	MHC loci (or HLA in humans)	MHC class I and II	Recombination not prevented	Codominance	Recessive genetic load?	Shiina *et al.* (2006), van Oosterhout (2009a)
Sexual chromosomes	*Primates, birds, plants...*	Negative frequency-dependent selection	Sex chromosomes XY and ZW		Structural rearrangement + high divergence	Derived chromosome dominant over ancestral	Degeneration of Y chromosome	Charlesworth & Charlesworth (2000)
Shell colour and pattern	*Cepaea nemoralis*	Negative frequency-dependent selection	Supergene. Five linked loci incl. Colour (*C*) and Band (*B*)		No recombination between Colour and Band loci			Richards *et al.* (2013)
Male behaviour and plumage	*Philomachus pugnax*	Negative frequency-dependent selection	Supergene ~125 genes		4.4 Mb inversion	Inverted > Ancestral	Homozygous lethal + deleterious effect on heterozygotes	Kupper *et al.* (2016)
Male behaviour and plumage	*Zonotrichia albicollis*	Negative frequency-dependent selection	*ZAL* Supergene - *ca.* 1000 genes	*Estrogen receptor 1 (ESR1)*	Pair of included pericentric inversions (at least 98 Mb)	Inverted > Ancestral	Inverted allele deleterious although not lethal	Horton *et al.* (2013, 2014), Thomas *et al.* (2008)
Male segregation distorter	*Mus musculus*	Overdominance	*t*-complex ~700 genes	Putatively one responder gene *Tcr*, and several distorter genes *Tcd*, including *Tagap1*	4 nonoverlapping inversions (33 Mb)	Inverted > Ancestral	Sperm of homozygotes is sterile Associated recessive lethal mutation s detected	Bauer *et al.* (2005), Lyon (2003), Sugimoto (2014)
Heterostyly	*Primula veris*	Negative frequency-dependent selection	*S*-locus – 113 genes		278 kb polymorphic insertion. Hemizygosity	Short style > Long style Coordination of dominance of the three main loci.	Homozygotes for the dominant allele (short style) are lethal.	Nowak *et al.* (2015), Li *et al.* (2016)

polymorphism seen in *Primula* or the buckwheat, and in fact cover drastically different traits and diverse taxa, such as social organization of *Formica selysi* ants (Purcell *et al.* 2014), mimetic wing colour pattern in *Papilio polytes* butterflies (Kunte *et al.* 2014; Nishikawa *et al.* 2015), or mating behaviour in birds, notably the ruff *Philomachus pugnax*. The spectacular polymorphism of male morphs in this wading bird was recently shown to be encoded by an ancestral and two rearranged alleles of a single supergene (Kupper *et al.* 2016; Lamichhaney *et al.* 2016). By strongly limiting recombination between loci and locking together co-adapted genes, chromosomal inversions can enable the persistence of functional polymorphism for a given trait. Table 1 details documented cases of inversion polymorphisms underlying traits under different selection regimes. The size of the regions and the number of genes captured vary considerably, from a single, long and complex gene such as *doublesex* controlling wing patterning variation the mimetic butterfly *P. polytes*, to entire chromosome arms such as those controlling changes in social organization of colonies in the fire ant *Solenopsis invicta* (Wang *et al.* 2013). This variation in inversion size may indicate differences in the balancing mechanisms as well, perhaps, as in the number of traits and sites affected by selection within the rearranged regions, but theoretical predictions are needed to disentangle the effects of those different factors.

The inversion underlying mimetic variation in the butterfly *P. polytes* seems to be restricted to the causative gene *doublesex*, whereas the fire ant inversion responsible for changes in social organization of colonies contains a large number of genes, many of which presumably have little role in the trait(s) targeted by balancing selection, although they may influence the maintenance of polymorphism through linked negative effects. For instance, degeneration of genes in the *ZAL2m* inversion were shown to have negative effects in *ZAL2m* homozygotes in the white-throated sparrows *Zonotrichia albicollis* (Tuttle *et al.* 2016); similarly, inverted haplotypes in the ruff *Philomachus pugnax*, associated with higher mating success, also have negative effects on survival and may play a role in the maintenance of inversion polymorphism (Kupper *et al.* 2016).

Influence of dominance on allele fate

Allelic dominance is another important feature of the genetic architecture of adaptive traits which strongly influences the fate of alleles, such as their equilibrium number and frequency but also their spatial distribution. Dominance is known to influence the probability of new alleles to become established. However, because balancing selection in diploid outbreeding species entails the formation of a high proportion of heterozygotes in natural populations, dominance in traits under balancing selection is also expected to affect allele dynamics and equilibria through time and space.

New alleles are expected to invade populations differently whether they are dominant or recessive. A nice example of this is provided by the dominance of the convergent mutations under positive selection allowing two species of lizards to switch from dark to white phenotypes in white sand areas (Rosenblum *et al.* (2010). One species *Aspidoscelis inornata* has a recessive derived mutation which could invade dark habitats, while the other species *Sceloporus undulatus* has a dominant derived mutation unable to invade dark habitats. For traits under NFDS however, the initial rarity of new alleles increases their fitness if they are expressed, thereby enhancing effective migration as compared to unlinked neutral loci (Schierup *et al.* 2000). Therefore, the probability for dominant alleles under NFDS to invade new populations is higher than for recessive alleles of equivalent effect which drift may easily eliminate (Schierup *et al.* 1997; Vekemans *et al.* 1998). This effect is a sophistication of the famous 'Haldane's sieve' effect (Haldane 1927) and predicts a general increase in dominance over the course of allelic diversification. In the supergene of the mimetic butterfly *H. numata*, alleles carried by the ancestral haplotype are strictly recessive to the derived alleles carried by inversions (Le Poul *et al.* 2014), showing the same Haldane's sieve effect. Similarly, in *P. polytes*, the mimetic allele *polytes* is derived (associated with the inversion) and dominant to the ancestral, non-mimetic allele *cyrus*. Bird supergenes affecting plumage and behavioural traits in the ruff *P. pugnax* and the white-throated sparrow *Z. albicollis* also display derived character combinations with dominant effects, born by rearrangements found only in heterokaryotypes. Dominance of new variants to pre-existing ones appears as a key parameter influencing the evolution of balanced polymorphism (see Table 1).

Dominance also influences allele frequency and therefore their persistence within populations. For instance, at the sporophytic *S*-locus of *A. halleri*, where balanced polymorphism is maintained by a NFDS, recessive alleles maintain higher allelic frequencies than dominant ones, as frequency-dependent selection is based upon expressed alleles and is blind to hidden recessive alleles (Billiard *et al.* 2007; Llaurens *et al.* 2008, 2013). In polymorphic mimicry where balanced polymorphism is maintained by selection/migration equilibrium, recessive alleles are able to enter populations even in the absence of local comimics, because they are rarely expressed. Their overall frequency at a larger scale is then predicted to reach higher levels than for dominant alleles (Llaurens *et al.* 2013), which suggests they are likely to persist in

populations for longer evolutionary periods and escape being lost by drift. In the stick insect *Timema cristinae*, three colour patterns co-exist, with a green pattern easily camouflaged on the host-plant *Ceanothus spinosus* leaves, a green striped pattern camouflaged on *Adenostomas fasciculatum* and a melanic morph camouflaged on the wood of both plant species; the melanic form is recessive to both green specialist morphs, and displays high between host-plants' migration rate, favouring the persistence of polymorphism within species, despite divergent selection exerted by host-plant specialization (Comeault *et al.* 2015). Clearly, dominance not only influences the probability for new alleles to be picked by selection, especially in ecological situations where rare alleles are favoured, but may also favour the persistence of polymorphism within populations.

Discordant or concordant dominance?

Balancing selection generates a high frequency of heterozygotes, whose phenotypes will depend on the dominance of the different genes involved in the trait. Because differentiated traits may often involve several genes, one might predict a coordinated dominance of these genes (see Table 1). In the self-incompatibility locus of *A. halleri*, a general consistency in dominance levels is observed between the two genes *SCR* and *SRK* although the encoded proteins and receptors are totally different and have drastically different molecular properties (Hiscock & McInnis 2003). Similarly, the supergene controlling polymorphic mimicry in *H. numata* appears to show highly coordinated dominance shaped by selection: heterozygotes obtained from crosses between alleles from distinct geographic regions showed uncoordinated dominance, while heterozygotes naturally found in large numbers within polymorphic populations displayed consistent dominance across all elements of wing pattern (Le Poul *et al.* 2014). Natural selection favouring mimicry of local models in natural heterozygotes would explain this coordination. These examples show how dominance relationships may be finely tuned by the ecology of the balancing mechanisms acting on these loci.

Evolution of dominance

Contrary to loci under positive selection where heterozygosity is generally transient and selection on dominance might occur only during the invasion phase (*i.e.* Haldane's sieve effect), alleles in traits under balancing selection are mostly found in heterozygotes so that their level of expression in heterozygous condition is likely to be under selection throughout their existence in natural populations. Using a theoretical approach, Otto & Bourguet (1999) indeed showed that natural selection is strong enough for dominance to evolve in loci under balancing selection because of the high frequencies of heterozygotes at such loci. Such evolution of dominance in balanced polymorphism has been demonstrated in specific theoretical approaches assuming the existence of dominance modifiers in sporophytic self-incompatibility (Llaurens *et al.* 2009a; Schoen & Busch 2009) and empirical evidence is accumulating in various systems (see Billiard & Castric 2011 for a review). In the mimicry gene *doublesex* in *Papilio dardanus*, dominant alleles prevent the expression of recessive alleles, suggesting an active switch-off mechanism (Nishikawa *et al.* 2015). In the mimetic butterfly *H. numata*, dominance mechanisms appear to have evolved during allele diversification (Le Poul *et al.* 2014). Finally, the existence of small RNAs acting as dominance modifiers in the *S*-locus has been demonstrated in *A. halleri* (Durand *et al.* 2014). The small RNAs are associated with dominant alleles and are specifically silencing the recessive ones, suggesting the evolution of specific dominance mechanisms at this locus, in response to strong natural selection. This is the first demonstration of evolution of dominance independent from the encoding protein and stresses that balancing selection is also shaping patterns of expression in the genes involved in adaptive variations.

Decay and loss of adaptive variation

Sheltered genetic load associated with loci under balancing selection

Regions of suppressed recombination are expected to accumulate deleterious recessive mutations. First, when structural variation (e.g. inversion polymorphism) occurs and is positively selected, deleterious mutations may be captured within the chromatid undergoing the initial structural change(s) (Kirkpatrick 2010). Such deleterious recessives may subsequently be protected by selection on the beneficial variants located within the inversion. Second, high heterozygosity can further shelter recessive deleterious mutations and slow their purge. This should result in the build-up of a sheltered genetic load, as was suggested for MHC regions (van Oosterhout 2009a) or the *S*-locus (Llaurens *et al.* 2009b). The high frequency of genetic diseases associated with the *HLA* gene region in humans suggests a possible accumulation of deleterious mutations at this balanced locus (de Bakker *et al.* 2006; Shiina *et al.* 2006). The enrichment of transposable elements in this region also supports the inefficiency of shedding parasitic or deleterious genetic elements in these regions (van Oosterhout

2009b). Significant deficit of homozygotes at the *S*-locus in the progeny of forced incompatible crosses in *A. halleri* (Llaurens *et al.* 2009b) and *A. lyrata* (Stift *et al.* 2013) confirmed the existence of a sheltered genetic load at this locus. Lethal haplotypes in the fire ant social supergene have been hypothesized to account for the lack of one the two possible homozygotes (Wang *et al.* 2013). Similarly, the colour and behavioural polymorphism in the white-throated sparrow are associated with near-perfect disassortative mating between ancestral and rearranged alleles at the supergene, and a lack of homozygotes for the rearranged allele (Thomas *et al.* 2008). Recent nucleotide polymorphisms and expression studies from the two alleles have hinted at a functional degradation of the rearranged allele, which might be expected if this allele is invariably found in heterozygotes and accumulate deleterious mutations (Tuttle *et al.* 2016). A similar situation occurs in the ruff (*Philomachus pugnax*) where the dominant and most derived allele ('satellite') is known to be lethal when homozygous, which is putatively attributed to the disruption of an essential gene, *CENP-N* gene (encoding centromere protein N) by an inversion breakpoint (Kupper *et al.* 2016). The most striking example of such decayed evolution at loci maintaining balanced polymorphism is sex chromosomes where *Y* or *W* chromosomes are degenerating because of the lack of recombination (Charlesworth & Charlesworth 2000). However, the proportion of a genetic load due to mutation captured initially vs. that accumulated secondarily is still largely unknown in most cases and may depend on the level of purging opportunities. Generally, derived haplotypes stemming from inversions and being dominant to ancestral ones tend to accumulate a higher mutation load, as illustrated for the *S*-locus, where dominant haplotypes display an elevated number of transposable elements as compared to recessive ones (Goubet *et al.* 2012).

Genetic load may strengthen balancing selection

Deleterious mutations associated with some alleles of a locus under balancing selection can drive overdominance, whereby heterozygotes enjoy better fitness than homozygotes expressing the associated genetic load. This further reinforces the persistence of balanced polymorphism in natural populations and can explain the long branches observed, for instance, in the phylogenies of alleles from the self-incompatibility locus (Uyenoyama 2003).

However, the association of adaptive alleles with increasing albeit sheltered genetic load might also be expected to contribute to their elimination from the populations, because other alleles without such load could benefit from higher fitness. Genetic load associated with adaptive alleles at polymorphic loci has indeed been shown to be partially recessive, for example in *A. halleri S*-locus (Llaurens *et al.* 2009b) or in the ruff supergene where heterozygous carriers of the lethal inversion suffer from a slightly reduced survival (Kupper *et al.* 2016). When the genetic load is also frequently expressed in heterozygotes, heterozygote advantage may decrease correspondingly, which may eventually lead to the extinction of the associated allele. Altogether, however, the capture and/or accumulation of deleterious mutations can strengthen balancing selection over long evolutionary timescales.

Mechanisms of balancing selection posing a threat to natural populations

The persistence of polymorphism often rests on balancing mechanisms, some of which might restrict invasion capacities or even jeopardize population persistence. In cases of strict disassortative mating for instance, range expansions are limited by the simultaneous migration of several distinct morphs, as in self-incompatible plants (Petanidou *et al.* 2012). Similarly, small-sized populations with self-incompatibility systems are also believed to suffer from an increase in extinction risk (*i.e.* Allee effect), because of the limited mate availability (Levin *et al.* 2009).

In the polymorphic bird Gouldian finch (*Erythrura gouldiae*), red, black and yellow morphs coexist and are associated with variation in aggressiveness and parental investment differences. Imbalances between morph frequencies due to habitat destruction, for instance enhancing the frequency of the competitive red morphs which exhibit poor parental care, are predicted to increase population extinction risk (Kokko *et al.* 2014).

When differentiation at adaptive traits leads to speciation

Co-adapted haplotypes maintained within populations can also result in disruptive selection, which may ultimately lead to sympatric speciation (Rueffler *et al.* 2006). Loci under strong disruptive selection where recombination is limited might constitute speciation islands (Feder & Nosil 2010): for instance, the genetic clustering of traits linked to pollination syndromes observed in *Petunia* species is thought to favour ecological speciation because changing the community of attracted pollinators leads to reproductive isolation between plants living in sympatry (Hermann *et al.* 2013). Coordinated polymorphism at this genetic locus may be associated with transient polymorphism and quickly lead to a permanent split into separate taxa. Inversion polymorphisms, by immediately generating a

strong hitchhiking effect, might further promote speciation, as theoretically predicted by Feder *et al.* (2014). The chromosomal inversion polymorphism observed within the species *Mimulus guttatus*, which might have been promoted through the frequent contact among ecotypes in the transition zone, results not only in local adaptation but also reproductive isolation (Lowry & Willis 2010). In the case of balanced polymorphism generated by migration/selection balance, the evolution of clusters of adaptive genes protected from recombination can eventually result in divergent selection and speciation, that is the cessation of recombination at the genome scale. In that case, adaptive variation may become partitioned between speciating taxa, essentially transforming within-population polymorphism into variation at higher taxonomic levels, and departing from the definition of adaptive, within-population polymorphism.

Conclusions and perspectives

Traits maintaining multiple variants within populations provide a rare opportunity to study allele diversification and its association with functional phenotypic variation in a shared genetic and environmental background. Balancing selection gives us insights into how selection shapes the underlying genetic architecture from the birth to the death of differentiated variants: coordinated dominance, lack of recombination and accumulation of deleterious mutations have been repeatedly found to be associated with balanced polymorphism (see Fig. 2 for a summary). However, we argue that there is often a continuum between directional and balancing selection: for example, local adaptations in connected populations can display signatures of directional selection but, as migration increases, may enter into a balancing selection regime where local

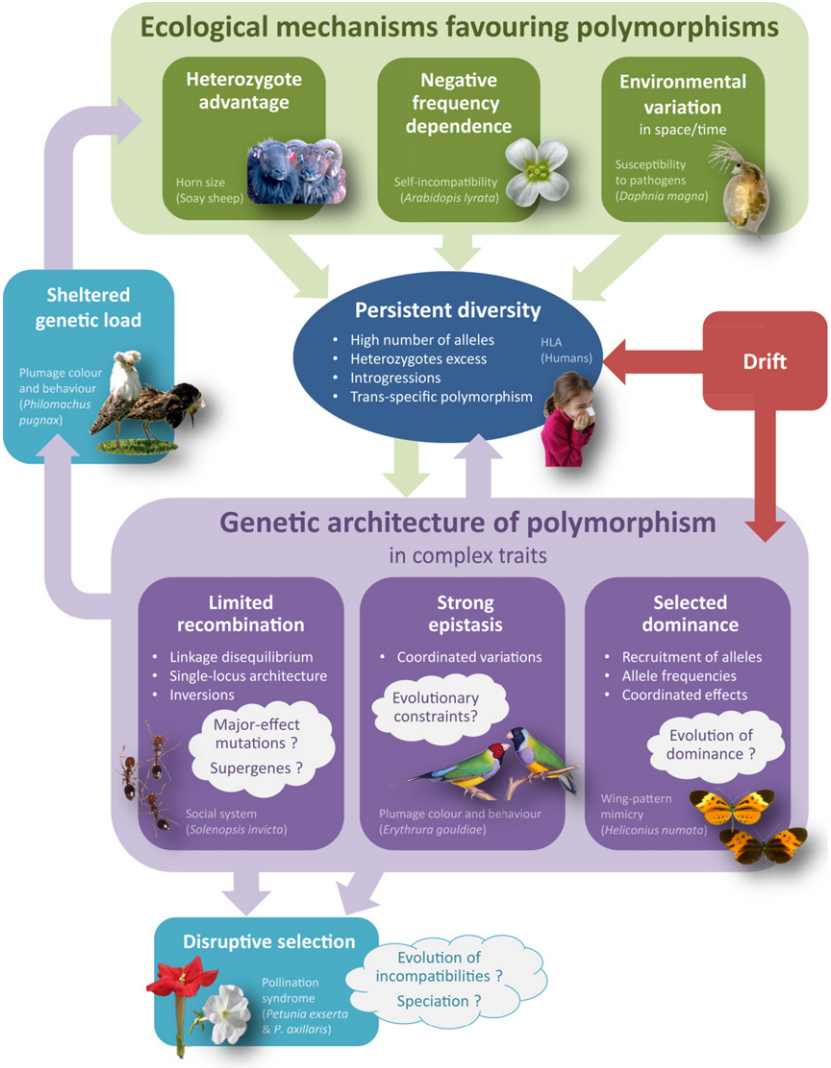

Fig. 2 Balancing selection on complex traits: interactions between evolutionary forces (green), drift (red) and genetic architecture (purple) shaping adaptive polymorphism, illustrated with emblematic examples described in the text. Clouds indicate suggested directions for future research. Credits to Jill Pilkington (Soay sheep), Vincent Castric (*Arabidopsis lyrata*), Dieter Ebert (*D. magna*), Arjan Haverkamp (*Philomachus pugnax*), Violaine Llaurens (human), Stephen Ausmus (ants), Mathieu Joron (*Heliconius numata*), Cris Kuhlemeier (*Petunia*)

polymorphism can drive specific genetic architecture. Theoretical approaches explicitly modelling different selective mechanisms promoting polymorphism and predicting their effect on the underlying genetic architecture are therefore crucially needed to pinpoint the evolutionary origin of polymorphisms observed in natural populations.

Although single-locus architecture has been repeatedly reported for traits under balancing selection, our knowledge might be biased towards large effect polymorphism with simple Mendelian inheritance. Balancing selection acting on polygenic, quantitative traits is undeniably more challenging to study and characterize. In fact, the large number of genomic regions displaying signatures of balancing selection suggests an important role for this selection regime and also leaves us with a rich array of possible balancing mechanisms. Experimental approaches characterizing the functional differences between different variants and their potential interactions are needed, as are reliable estimations of their fitness effects. Furthermore, disentangling the precise targets of selection for well-differentiated traits is challenging but nevertheless needed to reconstruct their evolutionary history: for instance, experiments would be required to infer which specific features of male morphs (e.g. behaviour, plumage colour, size or shape) have the strongest effect on fitness in the ruff.

Precise identification of causative variation *via* association mapping is generally challenging and is even more complicated when linkage among causative loci is strong or even when recombination is suppressed through inversions, as often observed in polymorphic traits. However, characterizing the number of causative genes and/or mutations within inverted regions and their epistatic interactions by other approaches, such as the study of differential expression among haplotypes, is a promising avenue for a better understanding of the evolution of differentiated variations in natural populations. The fine dissection of functional variants will allow estimating the relative contributions of point mutations, recombination events and introgression in the emergence of new functional variants. Many polymorphic traits are well differentiated, but some are caused by a single deletion or mutational events, as for instance a unique amino acid change in the acetylcholinesterase gene *ace-1* leading to insecticide resistance in the mosquito *Culex pipens* (Weill *et al.* 2003) or the insertion of a single transposable element in the *cortex* gene generating the melanic form in the peppered moth *Biston betularia* from Britain (van't Hof *et al.* 2016). It is possible that in many cases of pathogen or insecticide resistance genes, limited genetic changes are sufficient to generate a new function. However, for polymorphisms maintaining well-differentiated syndromes, sustained recombination among causative mutations within population should limit the build-up of phenotypic differentiation. Recently, a growing number of studies have highlighted the previously neglected role of introgression in adaptation, where initial differentiation may build up in a separate lineage. We suspect that introgression might be even more favoured in cases of balancing selection because of the limitations imposed by recombination. Similarly, inversions have been also recently found in a large proportion of loci under balancing selection (see Table 1), putatively because they prevent recombination between mutations involved in the switch between phenotypes. Inversions act as strongly selected allelic blocks, but we also suggest that the capture of deleterious mutations within inversions may enhance or even cause polymorphism through heterozygote advantage (Kupper *et al.* 2016).

Another important challenge to understanding the underpinnings of the maintenance of polymorphisms resides in the precise characterization of the interactions between alleles (dominance) and loci (epistasis) and their potential evolution. Investigating allele interactions might bring new light not only on their relative contribution to trait variation but also on the chronology of recruitment of the different loci.

Acknowledgements

We thank Louis Bernatchez for inviting us to write this review, Thomas Lenormand, Vincent Castric and two anonymous referees for valuable comments on earlier versions of this manuscript. We would like to thank Mathilde Cordelier, Vincent Castric, Jon Slate, Jill Pilkington, Sarah Pryke, Laurent Keller, Yannick Wurm, Romain Libbrecht, Elaina Tuttle, Brent Horton, Marie Manceau, Christophe Thébault, Cris Kuhlemeier and Clemens Küpper for sharing pictures with us and for their enthusiasm regarding this review. Our research is supported by grants from the Agence Nationale de la Recherche (ANR-13-JSV7-0003-01 DOMEVOL to VL and ANR-12-JSV7-0005 HYBEVOL to MJ), the European Research Council (ERC StG-243179 MimEvol to MJ) and an 'Emergence' grant from the Paris Regional Council to VL.

References

Andres AM, Hubisz MJ, Indap A *et al.* (2009) Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, **26**, 2755–2764.

Azevedo L, Serrano C, Amorim A, Cooper DN (2015) Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Human Genomics*, **9**, 21. doi:10.1186/s40246-015-0043-1.

de Bakker PIW, McVean G, Sabeti PC *et al.* (2006) A high-resolution HLA and SNP haplotype map for disease association

studies in the extended human MHC. *Nature Genetics*, **38**, 1166–1172.

Barson NJ, Aykanat T, Hindar K *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**, 405–408.

Bauer H, Willert JR, Koschorz B, Herrmann BG (2005) The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature Genetics*, **37**, 969–973.

Benkman CW (1996) Are the ratios of bill crossing morphs in crossbills the result of frequency-dependent selection? *Evolutionary Ecology*, **10**, 119–126.

Billiard S, Castric V (2011) Evidence for Fisher's dominance theory: how many 'special cases'? *Trends in Genetics*, **27**, 441–445.

Billiard S, Castric V, Vekemans X (2007) A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. *Genetics*, **175**, 1351–1369.

Bond AB, Kamil AC (2002) Visual predators select for crypticity and polymorphism in virtual prey. *Nature*, **415**, 609–613.

Bonhomme M, Doxiadis GGM, Heijmans CMC *et al.* (2008) Genomic plasticity of the immune-related Mhc class I B region in macaque species. *BMC Genomics*, **9**, 514. doi:10.1186/1471-2164-9-514.

Bourguet D (1999) The evolution of dominance. *Heredity*, **83**, 1–4.

Brown KS (1981) The biology of *Heliconius* and related genera. *Annual Review of Entomology*, **26**, 427–456.

Castric V, Bechsgaard J, Schierup MH, Vekemans X (2008) Repeated adaptive introgression at a gene under multiallelic balancing selection. *PloS Genetics*, **4**, e1000168.

Charlesworth D, Charlesworth B (1975) Theoretical genetics of batesian mimicry.2. Evolution of supergenes. *Journal of Theoretical Biology*, **55**, 305–324.

Charlesworth B, Charlesworth D (2000a) The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **355**, 1563–1572.

Chouteau M, Arias M, Joron M (2016) Warning signals are under positive frequency-dependent selection in nature. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 2164–2169.

Clarke CA, Sheppard PM (1959) The genetics of *Papilio dardanus*, Brown. 1. Race cenea from South Africa. *Genetics*, **44**, 1347–1358.

Clarke CA, Sheppard PM (1971) Further studies on genetics of mimetic butterfly *Papilio memnon* L. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **263**, 35–70.

Clarke CA, Sheppard PM (1972) Genetics of mimetic butterfly *Papilio polytes* L. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **263**, 431–458.

Clarke CA, Sheppard PM, Thornton IW (1968) Genetics of mimetic butterfly *Papilio memnon* L. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **254**, 37–89.

Comeault AA, Flaxman SM, Riesch R *et al.* (2015) Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. *Current Biology*, **25**, 1975–1981.

Cote CL, Gagnaire PA, Bourret V *et al.* (2013) Population genetics of the American eel (*Anguilla rostrata*): FST=0 and North Atlantic Oscillation effects on demographic fluctuations of a panmictic species. *Molecular Ecology*, **22**, 1763–1776.

Decaestecker E, Gaba S, Raeymaekers JAM *et al.* (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, **450**, 870–U816.

DeGiorgio M, Lohmueller KE, Nielsen R (2014) A model-based approach for identifying signatures of ancient balancing selection in genetic data. *Plos Genetics*, **10**, e1004561.

van Diepen LTA, Olson A, Ihrmark K, Stenlid J, James TY (2013) Extensive trans-specific polymorphism at the mating type locus of the root decay fungus Heterobasidion. *Molecular Biology and Evolution*, **30**, 2286–2301.

Durand E, Meheust R, Soucaze M *et al.* (2014) Dominance hierarchy arising from a complex small RNA regulatory network. *Science*, **346**, 1200–1205.

Eimes JA, Bollmer JL, Whittingham LA *et al.* (2011) Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and copy number variation. *Journal of Evolutionary Biology*, **24**, 1847–1856.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Feder JL, Nosil P, Flaxman SM (2014) Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. *Frontiers in Genetics*, **5**, 295.

Fenster CB, Reynolds RJ, Williams CW, Makowsky R, Dudash MR (2015) Quantifying hummingbird preference for floral trait combinations: The role of selection on trait interactions in the evolution of pollination syndromes. *Evolution*, **69**, 1113–1127.

Fijarczyk A, Babik W (2015) Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, **24**, 3529–3545.

Fitzpatrick MJ, Feder E, Rowe L, Sokolowski MB (2007) Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature*, **447**, 210–U215.

Fumagalli M, Cagliani R, Pozzoli U *et al.* (2009) A population genetics study of the Familial Mediterranean Fever gene: evidence of balancing selection under an overdominance regime. *Genes and Immunity*, **10**, 678–686.

Gao ZY, Przeworski M, Sella G (2015) Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. *Evolution*, **69**, 431–446.

Gervais CE, Castric V, Ressayre A, Billiard S (2011) Origin and diversification dynamics of self-incompatibility haplotypes. *Genetics*, **188**, 625–U205.

Gloag R, Ding G, Christie JR *et al.* (2016) An invasive social insect overcomes genetic load at the sex locus. *Nature Ecology & Evolution*, **1**, 0011.

Goubet PM, Berges H, Bellec A *et al.* (2012) Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in Arabidopsis. *Plos Genetics*, **8**, e1002495.

Gouyon P-H, Vichot F, Van Damme J (1991) Nuclear-cytoplasmic male sterility: single-point equilibria versus limit cycles. *American Naturalist*, **137**, 498–514.

Greene JS, Brown M, Dobosiewicz M *et al.* (2016) Balancing selection shapes density-dependent foraging behaviour. *Nature*, **539**, 254–258.

Grossen C, Keller L, Biebach I, Croll D, Int Goat Genome C (2014) Introgression from domestic goat generated variation at the major histocompatibility complex of alpine ibex. *Plos Genetics*, **10**, e1004438.

Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, **25**, 5–23.

Haldane JBS (1927) A mathematical theory of natural and artificial selection, Part V: selection and mutation. *Proceedings of the Cambridge Philosophical Society*, **23**, 838–844.

Hanski I, Mononen T, Ovaskainen O (2011) Eco-evolutionary metapopulation dynamics and the spatial scale of adaptation. *American Naturalist*, **177**, 29–43.

Hermann K, Klahre U, Moser M *et al.* (2013) Tight genetic linkage of prezygotic barrier loci creates a multifunctional speciation Island in Petunia. *Current Biology*, **23**, 873–877.

Hiscock SJ, McInnis SM (2003) Pollen recognition and rejection during the sporophytic self-incompatibility response: Brassica and beyond. *Trends in Plant Science*, **8**, 606–613.

van't Hof AE, Campagne P, Yung C *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, **534**, 102–105.

Hori M (1993) Frequency-dependent natural selection in the handedness of scale-eating Cichlid fish. *Science*, **260**, 216–219.

Horton BM, Hu Y, Martin CL *et al.* (2013) Behavioral characterization of a white-throated sparrow homozygous for the ZAL2(m) chromosomal rearrangement. *Behavior Genetics*, **43**, 60–70.

Horton BM, Hudson WH, Ortlund EA *et al.* (2014) Estrogen receptor alpha polymorphism in a species with alternative behavioral phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 1443–1448.

Huber B, Le Poul Y, Whibley A *et al.* (2015) Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. *Heredity*, **114**, 515–524.

Johnston SE, Gratten J, Berenos C *et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, **502**, 93 + .

Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evolutionary Ecology*, **13**, 721–754.

Joron M, Papa R, Beltran M *et al.* (2006) A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. *Plos Biology*, **4**, 1831–1840.

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

Kekalainen J, Vallunen JA, Primmer CR, Rattya J, Taskinen J (2009) Signals of major histocompatibility complex overdominance in a wild salmonid population. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 3133–3140.

Key FM, Teixeira JC, de Filippo C, Andres AM (2014) Advantageous diversity maintained by balancing selection in humans. *Current Opinion in Genetics & Development*, **29**, 45–51.

Kirkpatrick M (2010) How and why chromosome inversions evolve. *Plos Biology*, **8**, e1000501.

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

Kokko H, Griffith SC, Pryke SR (2014) The hawk-dove game in a sexually reproducing species explains a colourful polymorphism of an endangered bird. *Proceedings of the Royal Society B-Biological Sciences*, **281**, 20141794.

Komata S, Lin C-P, Iijima T, Fujiwara H, Sota T (2016) Identification of *doublesex* alleles associated with the female-limited Batesian mimicry polymorphism in *Papilio memnon*. *Scientific Reports*, **6**, 34782.

Kunte K, Zhang W, Tenger-Trolander A *et al.* (2014) doublesex is a mimicry supergene. *Nature*, **507**, 229–232.

Kupper C, Stocks M, Risse JE *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, **48**, 79–83.

Labbe P, Berthomieu A, Berticat C *et al.* (2007a) Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. *Molecular Biology and Evolution*, **24**, 1056–1067.

Labbe P, Berticat C, Berthomieu A *et al.* (2007b) Forty years of erratic insecticide resistance evolution in the Mosquito *Culex pipiens*. *Plos Genetics*, **3**, 2190–2199.

Lamichhaney S, Fan G, Widemo F *et al.* (2016) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, **48**, 84 + .

Le Poul Y, Whibley A, Chouteau M, Prunier F, Llaurens V, Joron M. (2014) Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nature Communications*, **5**, doi:10.1038/ncomms6644.

Leducq JB, Llaurens V, Castric V *et al.* (2011) Effect of balancing selection on spatial genetic structure within populations: theoretical investigations on the self-incompatibility locus and empirical studies in *Arabidopsis halleri*. *Heredity*, **106**, 319–329.

Leffler EM, Gao ZY, Pfeifer S *et al.* (2013) multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, **339**, 1578–1582.

Lenormand T, Otto SP (2000) The evolution of recombination in a heterogeneous environment. *Genetics*, **156**, 423–438.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, **87**, 331–333.

Levin DA, Kelley CD, Sarkar S (2009) Enhancement of Allee effects in plants due to self-incompatibility alleles. *Journal of Ecology*, **97**, 518–527.

Lewis D, Jones DA (1992) The genetics of heterostyly. In: *Evolution and Function of Heterostyly* (ed. Barrett SCH), pp. 129–150. Springer-Verlag, New York, New York.

Li J, Cocker JM, Wright J *et al.* (2016) Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nature Plants*, **2**, 16188.

Llaurens V, Billiard S, Leducq J-B, Castric V, Vekemans X (2008) Does frequency-dependent selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in *Arabidopsis halleri? Evolution*, **62**, 2545–2557.

Llaurens V, Billiard S, Castric V, Vekemans X (2009a) Evolution of dominance in sporophytic self-incompatibility systems: I. Genetic load and co-evolution of levels of dominance in pollen and pistil. *Evolution*, **63**, 2427–2437.

Llaurens V, Gonthier L, Billiard S (2009b) The sheltered genetic load linked to the S-locus: new insights from theoretical and empirical approaches in sporophytic self-incompatibility. *Genetics*, **183**, 1105–1118.

Llaurens V, Billiard S, Joron M (2013) The effect of dominance on polymorphism in Mullerian mimicry. *Journal of Theoretical Biology*, **337**, 101–110.

Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O (2008) Migration load in plants: role of pollen and seed dispersal in heterogeneous landscapes. *Journal of Evolutionary Biology*, **21**, 294–309.

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

Lyon MF (2003) Transmission ratio distortion in mice. *Annual Review of Genetics*, **37**, 393–408.

Martin A, Papa R, Nadeau NJ *et al.* (2012) Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 12632–12637.

Martin A, McCulloch KJ, Patel NH, Briscoe AD, Gilbert LE, Reed RD. (2014) Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *Evodevo*, **5**, doi:10.1186/2041-9139-5-7.

McDowell JM, Simon SA (2006) Recent insights into R gene evolution. *Molecular Plant Pathology*, **7**, 437–448.

McGregor AP, Orgogozo V, Delon I *et al.* (2007) Morphological evolution through multiple cis-regulatory mutations at a single. *Nature*, **448**, 587–U586.

Mehta RB, Nonaka MI, Nonaka M (2009) Comparative genomic analysis of the major histocompatibility complex class I region in the teleost genus Oryzias. *Immunogenetics*, **61**, 385–399.

Merrill RM, Dasmahapatra KK, Davey J *et al.* (2015) The diversification of Heliconius butterflies: what have we learned in 150 years ? *Journal of Evolutionary Biology*, **28**, 1417–1438.

Mokkonen M, Kokko H, Koskela E *et al.* (2011) Negative frequency-dependent selection of sexually antagonistic alleles in *Myodes glareolus*. *Science*, **334**, 972–974.

Nadeau NJ, Pardo-Diaz C, Whibley A *et al.* (2016) The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature* **534**, 106–110.

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, **39**, 121–152.

Nicholson AJ (1927) A new theory of mimicry. *Australian Zoologist*, **5**, 10–104.

Nielsen R, Williamson S, Kim Y *et al.* (2005) Genomic scans for selective sweeps using SNP data. *Genome Research*, **15**, 1566–1575.

Nishikawa H, Iijima T, Kajitani R *et al.* (2015) A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nature Genetics*, **47**, 405–409.

Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. (2015) The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biology*, **16**, doi:10.1186/s13059-014-0567-z.

Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society B-Biological Sciences*, **276**, 1119–1128.

van Oosterhout C (2009a) A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 657–665.

van Oosterhout C (2009b) Transposons in the MHC: the yin and the yang of the vertebrate immune system. *Heredity* **103**, 190–191.

Otto SP, Bourguet D (1999) Balanced polymorphisms and the evolution of dominance. *American Naturalist*, **153**, 561–574.

Pavey SA, Gaudin J, Normandeau E *et al.* (2015) RAD sequencing highlights polygenic discrimination of habitat ecotypes in the Panmictic American Eel. *Current Biology*, **25**, 1666–1671.

Petanidou T, Godfree RC, Song DS *et al.* (2012) Self-compatibility and plant invasiveness: comparing species in native and invasive ranges. *Perspectives in Plant Ecology Evolution and Systematics*, **14**, 3–12.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Poulton EB (1912) Mimicry, mutation and Mendelism. *Bedrock*, **2**, 42–56.

Pruitt JN, Riechert SE, Jones TC (2008) Behavioural syndromes and their fitness consequences in a socially polymorphic spider, *Anelosimus studiosus*. *Animal Behaviour*, **76**, 871–879.

Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M (2014) Convergent genetic architecture underlies social organization in ants. *Current Biology*, **24**, 2728–2732.

Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA (2015) Genomic evidence for role of inversion 3RP of *Drosophila melanogaster* in facilitating climate change adaptation. *Molecular Ecology*, **24**, 2423–2432.

Reed RD, Papa R, Martin A *et al.* (2011) optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, **333**, 1137–1141.

Rendueles O, Amherd M, Velicer GJ (2015) Positively frequency-dependent interference competition maintains diversity and pervades a natural population of cooperative microbes. *Current Biology: CB*, **25**, 1673–1681.

Richards PM, Liu MM, Lowe N *et al.* (2013) RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, **22**, 3077–3089.

Rosenblum EB, Roempler H, Schoeneberg T, Hoekstra HE (2010) Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 2113–2117.

Rostant WG, Kay C, Wedell N, Hosken DJ (2015) Sexual conflict maintains variation at an insecticide resistance locus. *BMC Biology*, **13**, 1.

Roux C, Pauwels M, Ruggiero MV, Charlesworth D, Castric V, Vekemans X (2013) Recent and ancient signature of balancing selection around the S-locus in Arabidopsis halleri and A. lyrata. *Molecular Biology and Evolution*, **30**, 435–447.

Rueffler C, Van Dooren TJM, Leimar O, Abrams PA (2006) Disruptive selection and then what? *Trends in Ecology & Evolution*, **21**, 238–245.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics*, **147**, 835–846.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research Cambridge*, **76**, 51–62.

Schoen DJ, Busch JW (2009) The evolution of dominance in sporophytic self-incompatibility systems. II. Mate availability and recombination. *Evolution*, **63**, 2099–2113.

Schwander T, Libbrecht R, Keller L (2014) Supergenes and Complex Phenotypes. *Current Biology*, **24**, R288–R294.

Segurel L, Thompson EE, Flutre T *et al.* (2012) The ABO blood group is a trans-species polymorphism in primates. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 18493–18498.

Segurel L, Gao ZY, Przeworski M (2013) Ancestry runs deeper than blood: the evolutionary history of ABO points to cryptic variation of functional importance. *BioEssays*, **35**, 862–867.

Shiina T, Ota M, Shimizu S *et al.* (2006) Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. *Genetics*, **173**, 1555–1570.

Shuster SM, Wade MJ (1991) Equal mating success among male reproductive strategies in a marine isopod. *Nature*, **350**, 608–610.

Silva SE, Rodrigues ASB, Marabuto E *et al.* (2015) Differential survival and reproduction in colour forms of *Philaenus spumarius* give new insights to the study of its balanced polymorphism. *Ecological Entomology*, **40**, 759–766.

Sinervo B, Lively CM (1996) The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, **380**, 240–243.

Smithson A, Macnair MR (1997) Negative frequency-dependent selection by pollinators on artificial flowers without rewards. *Evolution*, **51**, 715–723.

Spottiswoode CN, Stevens M (2012) Host-parasite arms races and rapid changes in bird egg appearance. *American Naturalist*, **179**, 633–648.

Stift M, Hunter BD, Shaw B *et al.* (2013) Inbreeding depression in self-incompatible North-American *Arabidopsis lyrata*: disentangling genomic and S-locus-specific genetic load. *Heredity*, **110**, 19–28.

Stoeckel S, Castric V, Mariette S, Vekemans X (2008) Unequal allelic frequencies at the self-incompatibility locus within local populations of *Prunus avium* L.: an effect of population structure? *Journal of Evolutionary Biology*, **21**, 889–899.

Sugimoto M (2014) Developmental genetics of the mouse t-complex. *Genes & Genetic Systems*, **89**, 109–120.

Takahashi Y, Yoshimura J, Morita S, Watanabe M (2010) negative frequency-dependent selection in female color polymorphism of a damselfly. *Evolution*, **64**, 3620–3628.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967–978.

Thomas JW, Caceres M, Lowman JJ *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.

Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity*, **113**, 1–8.

Thompson MJ, Timmermans MJTN, Jiggins CD, Vogler AP (2014) The evolutionary genetics of highly divergent alleles of the mimicry locus in *Papilio dardanus*. *BMC Evolutionary Biology*, **14**, 140. doi:10.1186/1471-2148-14-140.

Thorneycroft HB (1975) Cytogenetic study of white-throated sparrow, *Zonotrichia-albicollis* (Gmelin). *Evolution*, **29**, 611–621.

Timmermans M, Baxter SW, Clark R *et al.* (2014) Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proceedings of the Royal Society B-Biological Sciences*, **281**, 20140465.

Turelli M, Barton N (2004) Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G× E interactions. *Genetics*, **166**, 1053–1079.

Turner JRG (1987) The evolutionary dynamics of batesian and muellerian mimicry - similarities and differences. *Ecological Entomology*, **12**, 81–95.

Tuttle EM, Bergland AO, Korody ML *et al.* (2016) Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, **26**, 344–350.

Uyenoyama MK (1997) Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics*, **147**, 1389–1400.

Uyenoyama MK (2003) Genealogy-dependent variation in viability among self-incompatibility genotypes. *Theoretical Population Biology*, **63**, 281–293.

Van Valen L (1973) A new Evolutionary law. *Evolutionary Theory*, **1**, 1–30.

Vekemans X, Schierup MH, Christiansen FB (1998) Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. *Evolution*, **52**, 19–29.

Wallbank RWR, Baxter SW, Pardo-Diaz C *et al.* (2016) Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *Plos Biology*, **14**, e1002353.

Wang J, Wurm Y, Nipitwattanaphon M *et al.* (2013) A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, **493**, 664–668.

Weedall GD, Conway DJ (2010) Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends in Parasitology*, **26**, 363–369.

Weill M, Lutfalla G, Mogensen K *et al.* (2003) Insecticide resistance in mosquito vectors. *Nature*, **423**, 136–137.

Wilkins AS, Wrangham RW, Fitch WT (2014) The "Domestication Syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, **197**, 795–808.

Yassin A, Bastide H, Chung H, Veuille M, David JR, Pool JE (2016a) Ancient balancing selection at tan underlies female colour dimorphism in *Drosophila erecta*. *Nature Communications*, **7**, 10400. doi:10.1038/ncomms10400.

Yassin A, Delaney EK, Reddiex AJ *et al.* (2016b) The pdm3 locus is a hotspot for recurrent evolution of female-limited color dimorphism in Drosophila. *Current Biology*, **26**, 2412–2422.

Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E1743–E1751.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, **65**, 1897–1911.

All authors wrote the manuscript together
