CellPress

Trends in Genetics

Review

The molecular basis of phenotypic evolution: beyond the usual suspects

Rong-Chien Lin[1], Bianca T. Ferreira[1], and Yao-Wu Yuan[1,*]

It has been well documented that mutations in coding DNA or *cis*-regulatory elements underlie natural phenotypic variation in many organisms. However, the development of sophisticated functional tools in recent years in a wide range of traditionally non-model systems have revealed many 'unusual suspects' in the molecular bases of phenotypic evolution, including upstream open reading frames (uORFs), cryptic splice sites, and small RNAs. Furthermore, large-scale genome sequencing, especially long-read sequencing, has identified a cornucopia of structural variation underlying phenotypic divergence and elucidated the composition of supergenes that control complex multi-trait polymorphisms. In this review article we highlight recent studies that demonstrate this great diversity of molecular mechanisms producing adaptive genetic variation and the panoply of evolutionary paths leading to the 'grandeur of life'.

Highlights

Many previously unsuspected single-nucleotide polymorphisms (SNPs) can cause phenotypic variation by affecting transcript splicing and protein translation.

Small RNAs are more challenging to study than protein-coding genes but are important contributors to phenotypic evolution.

Structural variation (e.g., transposable element insertion, gene copy number variation, chromosomal inversion) is an extremely common source of phenotypic divergence.

The sequence composition of supergenes that control complex multi-trait polymorphisms has been elucidated at an unprecedented pace by long-read sequencing.

Taxon-specific genes, including non-coding RNAs, can have a profound impact on lineage-specific phenotypic diversification and adaptation.

Diversity of molecular mechanisms underlying phenotypic variation

Understanding the genetic and molecular mechanisms that give rise to phenotypic diversity has been a long-standing goal in biology. Extensive research in the past three decades has led to a well-established tenet: phenotypic evolution is usually caused by changes in coding DNA that affect protein function or mutations in ***cis*-regulatory elements** (see Glossary) that alter gene transcription [1–5]. In principle, a much wider spectrum of mutations can cause phenotypic changes by altering any of the gene expression processes, from **chromatin loop** formation to RNA splicing, from post-transcriptional silencing to protein translation. However, these mutations were often technically difficult to pinpoint until large-scale genome, transcriptome, and proteome sequencing became routine and sophisticated functional tools were developed for a broad range of non-model systems. In addition, the explosive growth of whole-genome sequencing across the tree of life has revealed that virtually all organisms contain a large set of genes that are found only in a specific group of organisms. Such taxon-specific genes (Box 1) can have profound effects on phenotypic diversification and innovation, and sometimes form supergenes that control co-adapted sets of traits. Herein we review recent studies pinpointing the specific genetic causes of phenotypic variation, particularly in natural organisms, highlighting the diversity and intricacy of the molecular mechanisms of organismal evolution.

Underappreciated single-nucleotide polymorphisms (SNPs)

There is no shortage of empirical examples of SNPs underlying phenotypic variation [4]. Most of the examples involve SNPs in coding DNA, *cis*-regulatory elements, or exon–intron junctions, leading to changes in protein function, patterns of gene transcription, or RNA splicing, respectively. However, recent studies have revealed some intriguing examples that SNPs beyond these three categories can also contribute to phenotypic diversification.

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

*Correspondence: yaowu.yuan@uconn.edu (Y.-W. Yuan).

 https://doi.org/10.1016/j.tig.2024.04.010

uORFs and uATGs

uORFs are short ORFs translated from AUG start codons within 5′ leader sequences upstream of the main ORF (mORF), and are widespread in eukaryotic mRNAs [6]. For example, about 35% of vertebrate transcripts and 37% of angiosperm mRNAs contain at least one uAUG in their 5′ untranslated regions (UTRs) [6]. It has been known for decades that these uAUGs and associated uORFs are important regulatory elements of protein translation, often substantially reducing the translation efficiency of the downstream mORF by blocking the scanning ribosome and/or **ribosome stalling** [7–11]. However, because uATGs usually do not affect gene transcript levels or amino acid sequences, they have been largely overlooked as a potential source of phenotypic evolution.

A recent study in monkeyflowers (*Mimulus* spp.) showed that the presence versus absence of uATG can cause dramatic phenotypic differences between natural species [12]. The hummingbird-pollinated *Mimulus cardinalis* produces bright red flowers with a high concentration of anthocyanin pigments, whereas the self-pollinated *M. parishii* bears pale pink flowers with a low anthocyanin concentration. Fine-scale genetic mapping and functional interrogation identified an anthocyanin-activator, *Petal Lobe Anthocyanin* (*PELAN*), as the causal gene responsible for the anthocyanin content variation between the two species. Liang *et al.* [12] identified an SNP located at 10 bp upstream of the mORF of *PELAN*; the causal mutation creates a uATG start codon (AGG → ATG) in the *M. parishii* allele, leading to attenuated protein translation and reduced coloration in the self-pollinated species (Figure 1Ai).

Mutations that affect the length of uORFs can also contribute to phenotypic variation. A uORF of the soybean (*Glycine max*) *Phosphate Transporter Traffic Facilitator 1* (*GmPHF1*) gene is required for mORF translation and phosphorus uptake through the root [13]. A common SNP within the 5′-UTR of *GmPHF1* introduces a stop codon that shortens the uORF and inhibits translation of the mORF, causing variation in phosphorus acquisition efficiency among soybean accessions (Figure 1Aii).

Reports on uORF-mediated phenotypic variation between animal species are still scarce, although segregating polymorphisms that alter uORF presence are prevalent within species [8,14], some of which are associated with human disease [15]. One noteworthy case is the uORF in the 5′-UTR of *PPARGC1A*, which encodes the PGC1α protein, a key regulator of mitochondrial biogenesis and oxidative metabolism. This uORF represses PGC1α translation and is conserved among vertebrates, but it is absent in the Atlantic bluefin tuna (*Thunnus thynnus*), an animal with exceptionally high abundance of mitochondria and red oxidative muscle [16] (Figure 1Aiii).

With the benefit of hindsight, one would expect that mutations causing the gain or loss of uATGs and associated uORFs should be a common source for phenotypic evolution in nature, as it takes only a single nucleotide change in many sequence contexts (e.g., AGG, ACG, ATA, and ATC). The examples discussed here are likely just the first of many similar cases that remain to be discovered.

Cryptic splice variants

While mutations in conserved exon–intron junctions are readily predicted to affect RNA splicing and potentially cause phenotypic change, emerging examples suggest that exonic and intronic SNPs that are not located in normal splice sites could also trigger **alternative splicing** or impact splicing efficiency. For example, the *MSX2A* gene, encoding a homeodomain transcription factor associated with osteoblast differentiation, is responsible for the difference in dorsal spine length between marine three-spined sticklebacks (*Gasterosteus aculeatus*) and their freshwater

Glossary

Adaptive radiation: the rapid evolution of taxonomic and phenotypic diversity as a consequence of adaptation to novel ecological niches.
Alternative splicing: a post-transcriptional mechanism that generates different transcripts from the same nascent RNA molecules.
Batesian mimicry: a palatable species mimics the warning signals of an unpalatable species to gain protection from the shared predators.
Chromatin loop: the ring-like structure of DNA sequences that allows interactions between different regions of the genome, from the same or different chromosomes, by bringing them closer to each other.
Chromosomal inversion: a DNA structural rearrangement in which a chromosomal fragment breaks at two points and reattaches to the original chromosome in the reversed orientation.
***Cis*-regulatory elements:** DNA segments that regulate transcription of the nearby coding DNA; they contain transcription factor binding sites.
Clade: a group of organisms that consists of a common ancestor and all of its descendants, representing a distinct branch on an evolutionary tree.
Cline: a spatial gradient in a genetic or phenotypic character across the geographic range of a species.
Haplotype: a stretch of DNA on a single DNA molecule (e.g., chromosome or mitochondrial DNA); alleles on a haplotype are inherited together from one of the parents.
Hemizygosity: only a single copy at a genetic locus is present, instead of the customary two copies, in diploid organisms.
Histone modifications: covalent addition of chemical groups (e.g., acetyl, methyl, or ubiquitin groups) on histone proteins, which can alter their interactions with DNA and other proteins.
Müllerian mimicry: two or more species with effective defenses mimic each other's warning signals and resemble each other in appearance.
Nascent transcript: newly produced RNA that is physically associated with the RNA polymerase during the transcription process.
Phased siRNAs: small interfering RNAs that are produced through miRNA-mediated cleavage of RNA transcripts derived from protein-coding

counterparts. Full-length *MSX2A* transcripts are produced in the former, whereas a shorter, non-functional transcript is produced in the latter (Figure 1Bi). Howes *et al.* [17] showed that this non-functional transcript is due to a single nucleotide substitution in the first exon, converting the sequence 'GGAGG' to a poly-G tract 'GGGGG' in the freshwater fish. Such poly-G tracts can serve as splicing enhancers that promote selection of nearby cryptic splice sites [18].

Similarly, a single nucleotide substitution in the middle of an intron of *FLOWERING LOCUS M* (*FLM*) creates a new 3' splice site that outcompetes the normal splice site, producing a non-functional transcript in some natural accessions of arabidopsis (*Arabidopsis thaliana*) [19] (Figure 1Bii). *FLM* encodes a MADS-box transcription factor inhibiting flowering at low temperatures, and the non-functional *FLM* transcript affects plant growth by conferring faster vegetative growth, earlier flowering, and decreased responsiveness to ambient temperature fluctuations. Thus, this intronic substitution may be advantageous in areas where monthly ambient temperatures vary unpredictably by preventing inappropriate induction of flowering under fluctuating temperatures [19].

An even more intriguing case comes from the flowering plant *Capsella rubella*, a selfing species with much smaller flowers than its outbreeding ancestors. One of the causal genes for petal size reduction is *CYP724A1*, encoding a brassinosteroid biosynthesis enzyme [20]. Two exonic nucleotide substitutions in the derived *C. rubella* allele do not trigger alternative splicing *per se*, but substantially enhance splicing efficiency of the functional *CYP724A1* transcript, which leads to higher-than-optimal levels of brassinosteroids and decrease in petal cell proliferation (Figure 1Biii).

The diversity of these examples, in terms of both the study organism and the mode of action, suggest that exonic and intronic SNPs that alter splicing patterns might be a common source of phenotypic variation. It is difficult to establish a causal link between these cryptic variants and phenotypic variation, as they are not obvious candidates considered to alter splicing. However, with the accumulation of more empirical examples through painstaking genetic mapping and functional characterization, we are hopeful that machine-learning-based methods can be developed in the future to predict which exonic or intronic SNPs have a high likelihood of altering RNA splicing and the ultimate phenotype (see Outstanding questions).

Small RNAs (sRNAs)

sRNAs are non-coding RNA molecules 20–30 nucleotides in length, and are involved in the regulation of diverse biological processes by targeting transcripts or chromatin based on sequence complementarity. The two most widespread sRNA types are microRNAs (miRNAs) and small interfering RNAs (siRNAs), found in virtually all eukaryotic organisms, whereas PIWI-interacting

genes or non-coding loci; these siRNAs are produced in a phased pattern defined by the miRNA cleavage sites.

Ribosome stalling: ribosomes pile up at specific positions while translating an ORF, which can lead to various biological consequences such as repressing the translation of downstream ORFs or activating mRNA decay pathways.

Short interspersed nuclear elements (SINEs): one subgroup of class I TEs. SINEs are derived from small cellular RNAs (e.g., tRNAs) and are non-autonomous elements, relying on the machinery of other TEs to replicate.

Standing variation: the presence of more than one allele at a locus in a population.

Sympatry: the existence of two or more species in the same geographical area at the same time.

Transposable elements (TEs): DNA segments that can move from one position to another non-homologous position within the genome. TEs are classified into two major groups: class I elements propagate via 'copy-and-paste' mechanisms, while class II elements move through 'cut-and-paste' mechanisms.

Box 1. Taxon-specific genes

With the explosive growth of whole-genome sequencing across the tree of life, it has become abundantly clear that new genes evolve in specific groups of organisms all the time and become 'taxonomically-restricted' as they are not shared across clades [79–84]. Although it has been widely recognized that phenotypic diversity is often generated by changes in function or expression patterns of toolkit genes that are shared among organisms [1–4,85], empirical evidence supporting the role of taxon-specific genes in phenotypic diversification and adaptation are accumulating rapidly [29,54,86–89]. For example, the flower color supergene *YUP-SOLAR-PELAN* is restricted to only a subclade of monkeyflowers (*Mimulus* spp.) and, since its origin ~5 million years ago, has played a critical role in flower color diversification and adaptation to different pollination modes in this group of *Mimulus* species (Figure IA) [29]. The rattlesnake (*Cortalus* sp.) genome contains a massive number of snake venom metalloproteinase (SVMP) genes in a tandem array that encode secreted SVMP toxins for subduing prey. This SVMP gene array resulted from multiple gene duplications and intragenic deletions in the rattlesnake lineage from a single ancestral disintegrin and metalloproteinase gene, *adam28* [88] (Figure IB), enabling rattlesnakes to employ these novel biochemical weapons. A similar mechanism involving intragenic deletions and complex duplications and modifications from an ancestral trypsinogen gene led to the massive expansion of the antifreeze glycoprotein gene array in the Antarctic notothenioid fish, enabling lineage-specific adaptation to sub-zero temperatures in the Antarctic Ocean [89] (Figure IC).

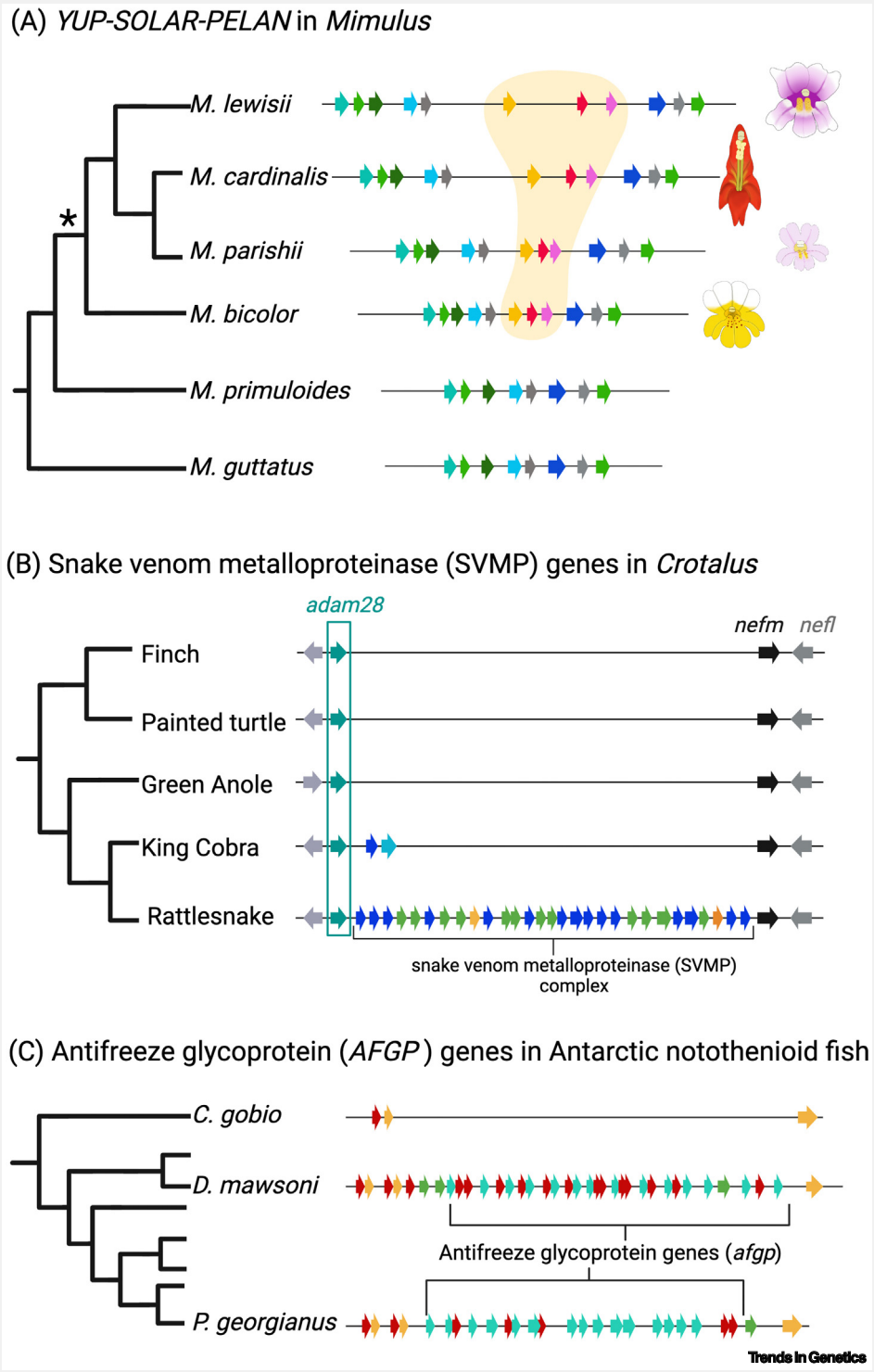

Figure I. Taxon-specific genes. (A) The *YUP-SOLAR-PELAN* supergene evolved in the common ancestor of a subclade of *Mimulus* (marked by the asterisk) and has played a critical role in flower color diversification in the descendant species. (B) The snake venom metalloproteinase (SVMP) gene family has been massively expanded specifically in the rattlesnake lineage. (C) The antifreeze glycoprotein gene array is unique to the Antarctic notothenioid fish genomes. Abbreviations: *C. gobio*, *Cottoperca gobio*; *D. mawsoni*, *Dissostichus mawsoni*; *P. georgianus*, *Pseudochaenichthys georgianus; PELAN, Petal Lobe Anthocyanin*; *YUP, YELLOW UPPER*.

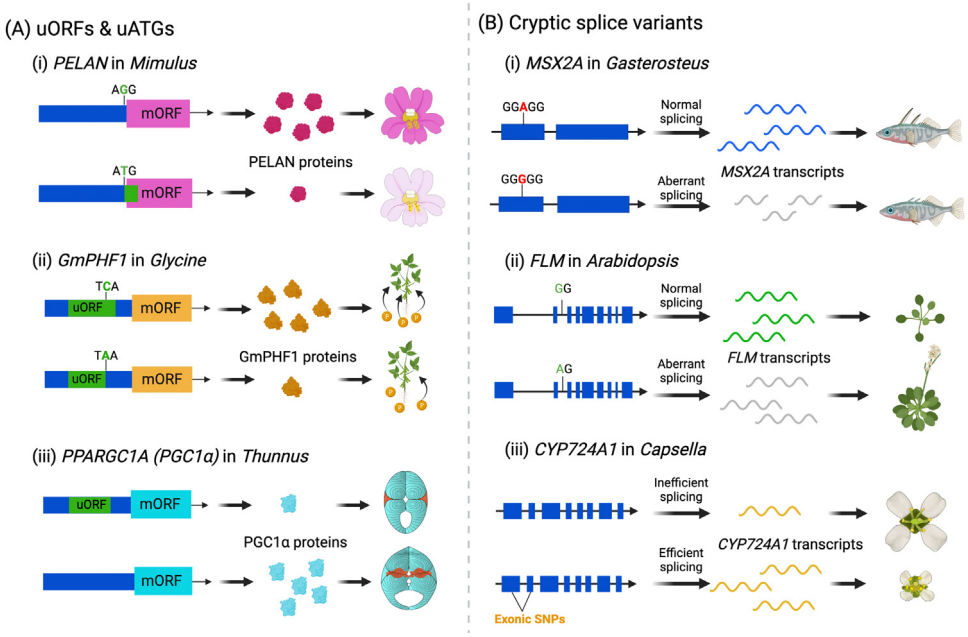

Figure 1. Underappreciated single-nucleotide polymorphisms (SNPs). (A) Changes in upstream open reading frames (uORFs) affect protein translation efficiency of the corresponding main ORFs (mORFs). In monkeyflowers (*Mimulus* spp.), a mutation in the 5′-UTR of a regulator of pigmentation (*Petal Lobe Anthocyanin, PELAN*) generates a new uORF that attenuates protein translation of the mORF, leading to reduced flower coloration (i). In soybean (*Glycine max*), a uORF upstream of *GmPHF1* is required for efficient protein translation and phosphorus uptake. A mutation truncates the uORF, leading to decreased protein abundance and phosphorus uptake (ii). In the Atlantic bluefin tuna (*Thunnus thynnus*), the lack of a highly conserved uORF that represses PGC1α protein translation results in exceptionally high abundance of mitochondria and red oxidative muscle (iii). (B) An exonic (i) and intronic (ii) SNP causes aberrant splicing, resulting in non-functional transcripts, and two exonic SNPs together enhance splicing efficiency of the functional transcript (iii).

RNAs (piRNAs) occur only in animals. The biogenesis of miRNAs and siRNAs involves the production of sRNA duplexes, processed by RNase III-like enzymes from longer precursors that are either double-stranded RNAs or single-stranded RNAs with hairpin structures [21,22].

Many miRNAs play pivotal roles in organism development, and not surprisingly, mutations in miRNA loci can facilitate evolutionary changes. For example, Todesco *et al*. [23] reported a naturally occurring SNP in the miRNA gene *MIR164A* in *A. thaliana*, which alters the stability of the miRNA:miRNA* duplex and reduces the accumulation of mature miR164. miR164 targets and represses *CUP-SHAPED COTYLEDONS2* (*CUC2*), encoding a transcription factor required for the serration on leaf margins. Natural *A. thaliana* strains carrying this miR164 mutation develop deeper leaf serrations than other strains (Figure 2Ai). In *Drosophila melanogaster*, the trichome-free area on the femur of the second leg (i.e., naked valley) exhibits considerable size variation among populations. This intraspecific variation was mapped to *mir-92a*, an miRNA gene targeting a positive regulator of trichome development, *shavenoid* (Figure 2Aii). Changes in *mir-92a* expression level underlie the naked valley size variation [24].

siRNAs can be produced from genomic loci that contain inverted duplicated sequences, as the resulting transcripts fold into stem–loop structures that are substrates for the sRNA biogenesis machinery. These inverted duplication loci are abundant in eukaryotic genomes [25] and often show taxon-specific distributions (Box 1). Recent studies showed that such taxon-specific

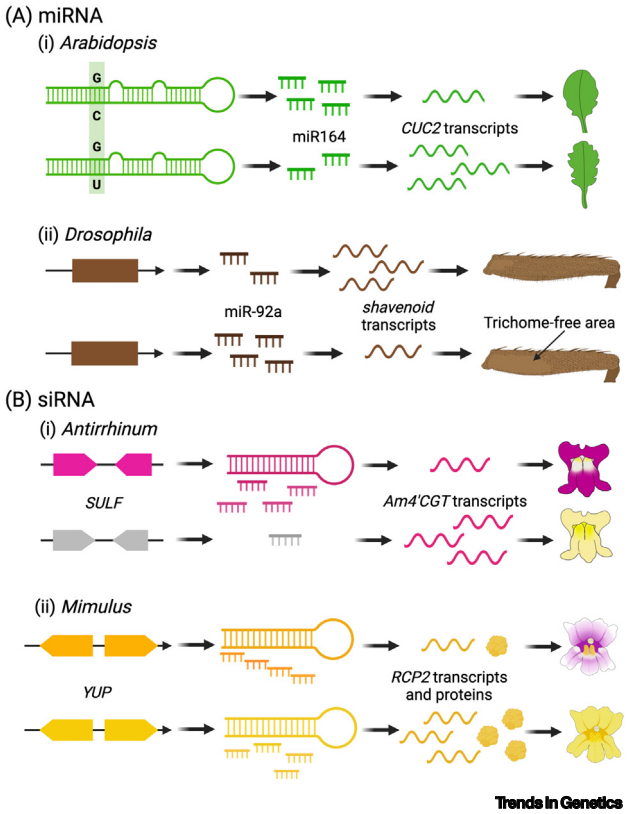

Figure 2. Small RNAs. (A) Mutations that alter the stability of the miRNA: miRNA* duplex (i) or transcriptional level of the primary microRNA (miRNA) (ii) cause intraspecific phenotypic variation. (B) Small interfering RNAs (siRNAs) derived from inverted duplications target flower color genes in *Antirrhinum* (i) and *Mimulus* (ii) species.

siRNAs play important roles in lineage-specific adaptation and speciation. For example, two subspecies of snapdragon, *Antirrhinum majus pseudomajus* and *A. m. striatum*, have alternative flower pigmentation patterns. The former has magenta flowers with a yellow patch near the corolla throat (Figure 2Bi) signaling the entry point for pollinating bees, whereas the latter has yellow flowers with magenta veins at the entry point as nectar guides. The distribution of the yellow aurone pigments is determined by *SULF* [25,26], a dominant suppressor for aurone deposition in the magenta flower. Although the genetic function of *SULF* has been known for >60 years [26], the molecular identity remained unknown until cost-effective genome sequencing became available. Bradley *et al.* [25] found that the *SULF* locus contains an inverted duplication of a gene fragment encoding chalcone 4′-O-glucosyltransferase (Am4′CGT), an enzyme required for aurone biosynthesis. The dominant *SULF* allele produces multiple siRNAs that target the original *Am4′CGT* (i.e., the source gene for *SULF*), whereas the recessive *sulf* allele produces few siRNAs. *SULF* is present at higher frequency in *A. m. pseudomajus* and shows steep **clines** in allele frequency in a natural hybrid zone between the two subspecies, suggesting strong selection acting on this *Antirrhinum*-specific siRNA locus [25].

The *YELLOW UPPER* (*YUP*) locus in monkeyflowers (*Mimulus* spp.) represents a parallel case. *YUP* is a major contributor to pollinator choice in the natural habitat of the bee-pollinated *M. lewisii* and the hummingbird-pollinated *M. cardinalis*, promoting reproductive isolation between these two species in **sympatry** [27]. *Mimulus lewisii* carries the dominant *YUP* allele that suppresses the accumulation of yellow carotenoid pigments in the pink corolla (Figure 2Bii). The recessive *yup* allele in *M. cardinalis* allows carotenoid accumulation on a pink background,

resulting in the bright red flower color (Figure I in Box 1). Substitution of the *YUP* allele from each species to the other produces a pollinator shift [28]. Similar to *SULF*, *YUP* has been known for >50 years as an important genetic locus, but its molecular identity remained mysterious until very recently. Liang *et al.* [29] reported that *YUP* is a non-coding locus producing **phased siRNAs**, one of which targets a master regulator of carotenoid accumulation, *Reduced Carotenoid Pigmentation 2* (*RCP2*) [30], for both transcript cleavage and translational inhibition. The recessive *yup* allele in *M. cardinalis* is due to mutations that disrupted the phased pattern of siRNA production, leading to low abundance of the specific siRNA that targets *RCP2*. *YUP* originated in the common ancestor of a subclade of *Mimulus* spp. through a partial inverted duplication of a *CYP450* gene [29]. Intriguingly, this *CYP450* source gene has no phylogenetic affinity with *RCP2*, highlighting the idiosyncrasy and intricacy of the molecular bases of phenotypic variation in nature.

Structural variation

Transposable elements

A common source of structural variation in eukaryotic genomes are **transposable elements (TEs)**. Although TEs have been postulated as a driving force of adaptive evolution for decades [31–33], and TE-induced phenotypic variation in domesticated organisms has been extensively documented [34–36], convincing evidence for TE-induced adaptive evolution in nature has been slow in coming. However, the increasing ease of large-scale genome sequencing, especially long-read sequencing, of virtually any organism has started to uncover fascinating examples of TE-mediated phenotypic evolution in nature.

The British peppered moth (*Biston betularia*) is a graphic example of real-time morphological adaptation to an altered environment. During the Industrial Revolution, the common pale form of the moth was rapidly replaced by a novel black form across the UK. Using a combination of linkage and association mapping in conjunction with high-quality, **haplotype-**resolved reference assemblies, van't Hof *et al*. [37] pinpointed the industrial melanism mutation to a large, tandemly repeated TE insertion at the *cortex* locus (Figure 3Ai), which has been linked to wing color patterning in numerous moth and butterfly species [38–41]. Although it was initially thought that the *cortex* gene itself controls wing color polymorphisms, three recent studies independently pointed the actual cause to a long non-coding RNA (lncRNA) and an associated miRNA at the *cortex* locus [42–44]. As such, the TE-induced melanism in the peppered moth may not be mediated through increased *cortex* transcript level, as originally suggested (Figure 3Ai) [37], but is more likely through the lncRNA and associated miRNA.

Another remarkable example of a TE insertion contributing to morphological innovation and **adaptive radiation** is the 'haplochromines' **clade** of cichlid fishes [45]. Adult males of ~1500 species in this clade produce a series of vibrantly colored circular markings in their anal fins, the 'egg-spots', which are signals for the mating behavior of the female fish. The expression of a pigmentation gene, *fhl2b* (*four and a half LIM domain protein 2b*), was found to be strongly associated with the formation of egg-spots. A **short interspersed nuclear element (SINE)** insertion in the *cis*-regulatory region of *fhl2b* was found in all surveyed haplochromine species bearing egg-spots, but was absent in all non-haplochromine species. Furthermore, transgenic experiments in zebrafish demonstrated that this TE insertion drives specific gene expression in pigment cells, supporting the critical role of this TE insertion in egg-spot formation [45].

In addition to altering gene transcription, TE insertions can also cause phenotypic variation through modulation of mRNA stability. The aforementioned *Capsella rubella* is an annual and inbreeding plant species with limited genetic variation due to an extreme genetic bottleneck during its speciation [46]. Niu *et al.* [47] found frequent TE insertions at the *FLOWERING LOCUS C* (*FLC*)

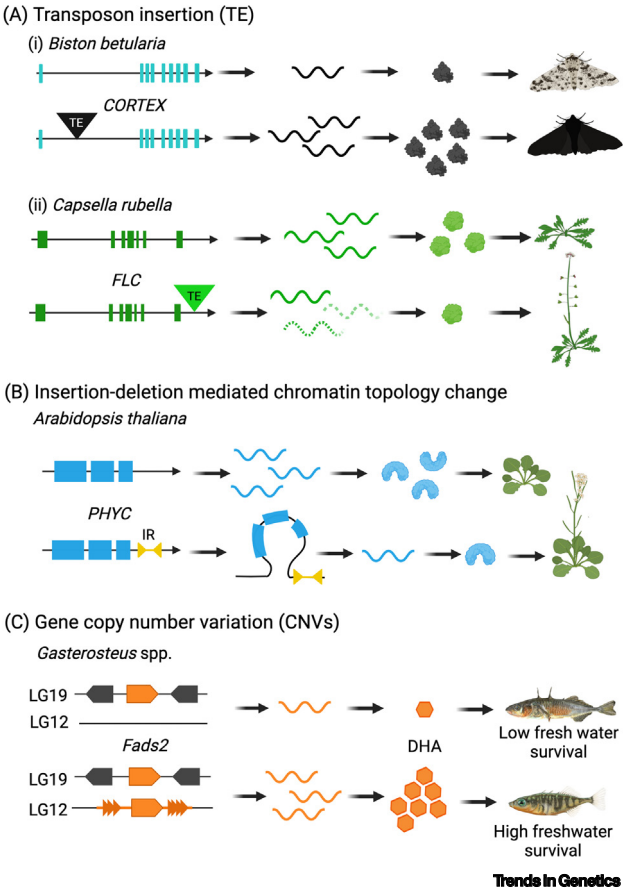

Figure 3. Structural variation. (A) A transposable element (TE) insertion at the *cortex* locus increases melanin production in the British peppered moth (*Biston betularia*) (i), initially thought to be mediated through increased transcription level of *cortex* but is more likely mediated through a long non-coding RNA (lncRNA) and the associated microRNA (miRNA) based on more recent evidence; and a TE insertion in the 3′-UTR of *FLC* in *Capsella rubella* decreases mRNA stability and triggers early flowering (ii). (B) Insertion of an inverted repeat (IR) downstream of the *PHYC* gene triggers chromatin loop formation that represses *PHYC* transcription in some *Arabidopsis thaliana* accessions. (C) Increased copy number of the *Fads2* gene in some three-spined stickleback lineages leads to higher docosahexaenoic acid (DHA) production and increased survival rate in DHA-deficient freshwater environments.

locus in natural populations of this species, explaining 12.5% of the natural variation in flowering time, a key life history trait correlated with fitness. FLC represses flowering. A recent TE insertion in the 3′-UTR was shown to decrease the steady-state mRNA level of *FLC*, thereby promoting early flowering. However, this insertion does not affect the abundance of unspliced **nascent transcripts** but reduces mRNA stability (Figure 3Aii). This study suggests that TE activities might be particularly important to quickly generate genetic and phenotypic change in species with limited **standing variation**, facilitating rapid adaptation to changing environments.

Insertion/deletion-mediated chromatin topology change

The considerable variation in flowering time among the natural accessions of *C. rubella* revealed that insertions/deletions can also cause phenotypic variation by mediating chromatin conformation change [48]. The causal mutations of some early flowering accessions were mapped to two distinct but overlapping deletions (32 bp and 54 bp, respectively) in the 5′-UTR region of *FLC*. These deletions reduce *FLC* expression and promote early flowering. Transgenic experiments showed that these deletions do not affect any functional *cis*-regulatory elements, but are tightly correlated with a more compact chromatin conformation and corresponding **histone modifications** [48].

Perhaps a more common type of insertion/deletion polymorphism underlying chromatin topology changes is the inverted repeat (IR), which can result from inverted duplications or form part

of the long terminal IRs of certain TE superfamilies (e.g., TEs in the MULE superfamily often have long terminal IRs [49]). Arce *et al.* [50] identified many insertional polymorphisms of IRs near protein-coding genes among natural accessions of *A. thaliana*. For example, in some accessions an IR is located ~500 bp downstream of *PHYTOCHROME C* (*PHYC*), a photoreceptor gene for sensing red and far-red light, promoting the formation of a repressive chromatin loop encompassing the entire *PHYC* gene (Figure 3B). By contrast, in the accessions lacking this IR, the chromatin loop is undetectable and *PHYC* expression level is higher with related developmental changes such as delayed flowering and shortened hypocotyls. Furthermore, clustered regularly interspaced short palindromic repeats (CRISPR) mutants with part of the IR deleted closely resembled natural accessions without the IR in chromatin topology, gene expression, and developmental phenotype. This confirms the causal role of the IR insertional polymorphism in chromatin loop formation and phenotypic variation.

Gene copy number variations (CNVs)

Another common type of structural variation is the difference in gene copy number between individuals. Changes in copy number often affect gene expression levels (gene dosage effects), which can generate phenotypic variation for adaptation to new environments. For example, Ishikawa *et al*. [51] demonstrated that copy number of the *Fatty acid desaturase 2* (*Fads2*) gene plays a key role in freshwater colonization by some stickleback fish lineages (Figure 3C). The food chain in freshwater ecosystems is usually deficient in docosahexaenoic acid (DHA), which imposes a nutritional constraint for freshwater colonization by marine animals. *Fads2* encodes an enzyme that is crucial for DHA biosynthesis. Through genetic examination, DHA measurements, and transgenic manipulation, Ishikawa *et al*. [51] showed that sticklebacks with higher *Fads2* copy numbers express *Fads2* at higher levels, leading to increased DHA synthesis and higher survival rate in the DHA-deficient freshwater environments.

Gene presence versus absence

An extreme case of gene copy number variations is when a gene of interest is present in one genotype but absent in another. A well-known example is the male sex-determining gene *SRY* in mammals [52]. Another example is the *S*-locus underlying heterostyly in primroses (*Primula* spp.) [53,54]. Individual primula plants produce either S-morph flowers with short styles and high anthers or L-morph flowers with long styles and low anthers (Figure 4A). This reciprocal positioning of sexual organs and the associated self-incompatibility is an effective way to promote outcrossing. The dominant *S*-locus haplotype carries a 280-kb DNA segment that is absent from the recessive *s* haplotype [55]. This genomic segment comprises five predicted genes, including *CYP734A50* and *GLOBOSA2*. *CYP734A50* encodes a cytochrome P450 mono-oxygenase degrading brassinosteroids, plant hormones that promote cell elongation. *GLOBOSA2* encodes a homoeotic MADS-box transcription factor that positively regulates stamen length (and hence anther height). Expression of *CYP734A50* and *GLOBOSA2* in the dominant *S* haplotype results in short styles (due to reduced brassinosteroid level) and long stamens, respectively; absence of these genes in the recessive *s* haplotype leads to the opposite phenotype (i.e., long styles and short stamens) [53,54]. Phylogenetic analysis revealed that both *CYP734A50* and *GLOBOSA2* evolved via gene duplication events specific to the primula lineage, once again demonstrating the importance of taxon-specific genes (Box 1) in phenotypic evolution.

In addition, because of the **hemizygosity** of the *S*-locus, linkage among the five genes in this locus cannot be broken by recombination; as such, they are inherited together as a single unit: a so-called 'supergene' whose importance in generating complex multi-trait polymorphisms has become inescapably evident in recent years [56], as highlighted in the following section.

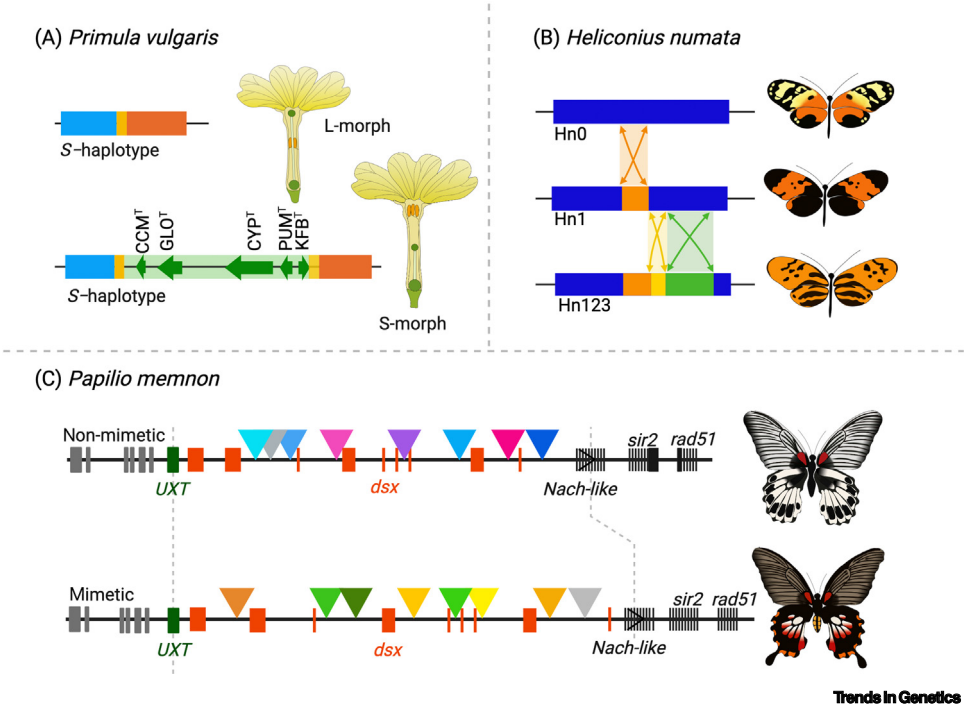

Figure 4. Supergenes. (A) The *S*-locus in *Primula vulgaris* controls floral dimorphism. The dominant *S*-haplotype contains five genes, including *CYP734A50* (*CYP*^{T}) and *GLOBOSA2* (*GLO*^{T}) that control style length and anther height, respectively. (B) Structure of the mimicry supergene locus *P* in *Heliconius numata* and a representative phenotype of each haplotype. The Hn0 haplotype carries the ancestral gene order, while the derived Hn1 and Hn123 haplotypes carry one and three inversions, respectively. (C) Structure of the mimicry supergene locus *A* in *Papilio memnon*. The vertical dashed lines mark the supergene boundaries. There are no chromosomal inversions in or near this supergene locus. Recombination suppression between the mimetic and non-mimetic haplotypes is most likely due to overaccumulation of transposable elements (TEs) (colored triangles) in this region [60].

Supergenes

A supergene is a chromosomal region with two or more tightly linked genes that control a co-adapted set of phenotypes. These linked genes segregate within or between populations as a single mendelian locus due to lack of recombination between haplotypes [56,57]. Recombination suppression can result from physical proximity of the individual genes, structural variation between homologous chromosomes such as hemizygosity and **chromosomal inversions**, or genomic context (e.g., centromeres, regions with abundant TEs) [57,58].

The aforementioned siRNA locus *YUP* in *Mimulus spp.* is part of a supergene composed of three tightly linked genes (*YUP*, *SOLAR*, and *PELAN*) in an ~11 kb chromosomal region [29]. While *YUP* represses accumulation of the carotenoid pigments, both *SOLAR* and *PELAN* are activators of anthocyanin pigmentation. The flower color output is a composite phenotype of both pigment types. The hummingbird-pollinated *M. cardinalis* carries a haplotype with recessive *yup* and dominant *PELAN* allele, resulting in the accumulation of both yellow carotenoids and purple anthocyanins, and thus the red color typical for hummingbird-pollinated flowers (Figure I in Box 1). The self-pollinated *M. parishii* carries a haplotype with dominant *YUP* and recessive *pelan* allele, resulting in the lack of carotenoids and low concentration of anthocyanins, and thus a very pale color as observed in many self-pollinated flowers. Physical proximity protects the haplotype and the co-adapted carotenoid and anthocyanin traits in each species from being shuffled by recombination.

Heterostyly in primroses represents the classic example of supergene maintenance by hemizygosity (Figure 4A). Strikingly, the same mechanism seems to be responsible for several independent origins of heterostyly across angiosperms, although the specific composition of the supergene differs from case to case [59,60]. Together with recent revelations of the genomic bases of color polymorphisms in *Timema* stick insects [61] and male wing dimorphism in the pea aphid [62], these examples suggest that hemizygous supergenes might be more common than previously thought.

Mounting evidence from a wide range of organisms indicates that chromosomal inversion is extremely prevalent in supergene formation and maintenance [54,63–72]. For example, the well-known **Müllerian mimicry** in the butterfly species *Heliconius numata* is controlled by a supergene, the *P* locus. Recombination at the *P* locus is suppressed by three chromosomal inversions (Figure 4B), which together cover a 1.75 Mb region including the *cortex* locus and several other genes with differential expression between mimetic forms [70]. However, the specific genes and molecular variants responsible for the different wing patterns in this system have yet to be determined.

Another butterfly species, *Papilio memnon*, provides a potential example of recombination suppression due to genomic context. *P. memnon* displays a female-limited **Batesian mimicry** polymorphism controlled by a supergene locus *A* (Figure 4C), containing the sex-determining transcription factor gene *doublesex* [66]. Functional interrogation using RNAi in a related species *P. polytes* showed that *doublesex* indeed plays a central role in switching the mimetic and non-mimetic wing patterns, and that genes flanking *doublesex* contribute to the refinement of the mimicry [72]. This supergene locus is highly divergent between the mimetic (*A*) and non-mimetic haplotype (*a*) compared with the rest of the genome due to lack of recombination within the supergene. However, this divergence is maintained without a chromosomal inversion. Instead, overaccumulation of repetitive sequences like TEs in this region may have provided the genomic context for recombination suppression [66].

Concluding remarks and future perspectives

The examples discussed herein clearly demonstrate that the molecular mechanisms producing adaptive genetic variation are much more diverse than the two usual suspects: coding DNA mutations that change protein function or *cis*-regulatory mutations that alter gene transcription. While these two variant types are undoubtedly important in generating phenotypic variation, an overemphasis on them in the past three decades was perhaps partly due to the 'streetlight effect', as they are relatively easier to pin down than the molecular variants discussed in this review. In our opinion, more emphasis on the following areas will likely lead to fruitful investigations in the coming years. (i) Comparative proteomics: given that most phenotypic changes are ultimately mediated through protein action, we expect that comparative proteomics between and within species will gradually come to center stage and will help unmask molecular variation at the translational and post-translational levels. (ii) Multigenic traits: most of the well characterized examples so far involve phenotypic variation caused by single or few genes of large effect. However, much of the phenotypic variation observed in nature (e.g., size, weight, shape) are polygenic. More research on multigenic traits is badly needed. (iii) Functional tool development: the case studies discussed in this review highlight the importance of functional tools, such as stable genetic transformation and CRISPR, in the discovery of causal molecular variants that would otherwise not have been foreseen. The continuing development of functional tools in understudied organisms will likely uncover additional (perhaps even surprising) molecular variants. (iv) Phenotypic plasticity: developmental plasticity is considered one of the major pathways to phenotypic novelty and diversity [73]. The same genotype can produce different phenotypes in response to

Outstanding questions

How do we predict the functional consequence of SNPs in exons, introns, or 5′- and 3′-UTRs in RNA splicing, mRNA stability, protein translation, and ultimately the phenotype? With the accumulation of empirical examples through painstaking genetic mapping and functional characterization, perhaps machine-learning-based methods can be developed towards this goal.

Associations between TEs and natural phenotypic variation have been frequently reported in recent years, yet how many of them are causal, and the functional mechanisms through which these TEs affect phenotypes, remain unknown. Even in the convincing case of the British peppered moth, it is still unclear how the complex TE insertion at the *cortex* locus increases melanin production.

What are the molecular elements or functional units that control the individual traits in a complex multi-trait polymorphism? Although the genomic location and sequence composition of many supergenes have been elucidated, the actual functional units within the supergene remain unknown in the majority of cases. The lack of recombination in these supergenes limits the power of fine-scale genetic mapping, but the continuous improvement in functional tools in these systems may shed some light.

How are supergenes formed in the first place? Many supergenes contain individual genes that were duplicated from an ancestral copy somewhere else in the genome, but were these individual genes recruited to the supergene location in a stepwise or wholesale fashion?

Does polygenic adaptation employ the same spectrum of molecular variants as the more extensively characterized single-gene or oligogenic variation?

What are the molecular bases of phenotypic plasticity that may fuel phenotypic diversification and the evolution of phenotypic novelty?

environmental cues through a myriad mechanisms, including temperature-dependent alternative splicing [74,75], light- and hormone-induced protein subcellular redistribution [76,77], and cell-lineage-specific DNA methylation [78]. However, how these plastic responses on the individual level become assimilated to phenotypic variation between populations or species remains poorly understood. We believe that research in these areas will lead to many more exciting discoveries and will greatly enrich our understanding of the molecular bases of phenotypic evolution.

Acknowledgments

We thank Dr Toby Bradshaw, Dr Elizabeth Jockusch, and our laboratory members for discussions on the subject over the years. Thanks to Dr Patricia Lang for discussion of miRNAs. Our work on the molecular basis of phenotypic evolution is supported by NSF grant IOS-2319721 and NIH grant R01GM140092.

Declaration of interests

The authors have no conflicts of interest to declare.

References

1. Hoekstra, H.E. and Coyne, J.A. (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evol. Int. J. Org. Evol.* 61, 995–1016
2. Carroll, S.B. (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36
3. Stern, D.L. and Orgogozo, V. (2009) Is genetic evolution predictable? *Science* 323, 746–751
4. Courtier-Orgogozo, V. *et al.* (2020) Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes. *Nucleic Acids Res.* 48, D696–D703
5. Hill, M.S. *et al.* (2021) Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* 22, 203–215
6. Zhang, H. *et al.* (2021) Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nat. Commun.* 12, 1076
7. Morris, D.R. and Geballe, A.P. (2000) Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* 20, 8635–8642
8. Calvo, S.E. *et al.* (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci.* 106, 7507–7512
9. Zhang, H. *et al.* (2019) Function and evolution of upstream ORFs in Eukaryotes. *Trends Biochem. Sci.* 44, 782–794
10. Dever, T.E. *et al.* (2020) Conserved upstream open reading frame nascent peptides that control translation. *Annu. Rev. Genet.* 54, 237–264
11. Wu, H.-Y.L. *et al.* (2023) Improved super-resolution ribosome profiling reveals prevalent translation of upstream ORFs and small ORFs in Arabidopsis. *Plant Cell* 36, 510–539
12. Liang, M. *et al.* (2022) Lost in translation: molecular basis of reduced flower coloration in a self-pollinated monkeyflower (*Mimulus*) species. *Sci. Adv.* 8, eabo1113
13. Guo, Z. *et al.* (2022) A natural uORF variant confers phosphorus acquisition diversity in soybean. *Nat. Commun.* 13, 3796
14. Zhang, H. *et al.* (2018) Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. *PLoS Biol.* 16, e2003903
15. Barbosa, C. *et al.* (2013) Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* 9, e1003529
16. Dumesic, P.A. *et al.* (2019) An evolutionarily conserved uORF regulates PGC1α and oxidative metabolism in mice, flies, and bluefin tuna. *Cell Metab.* 30, 190–200.e6
17. Howes, T.R. *et al.* (2017) Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. *BMC Biol.* 15, 115
18. Xiao, X. *et al.* (2009) Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat. Struct. Mol. Biol.* 16, 1094–1100
19. Hanemian, M. *et al.* (2020) Natural variation at FLM splicing has pleiotropic effects modulating ecological strategies in *Arabidopsis thaliana*. *Nat. Commun.* 11, 4140
20. Fujikura, U. *et al.* (2018) Variation in splicing efficiency underlies morphological evolution in *Capsella*. *Dev. Cell* 44, 192–203.e5
21. Chen, X. and Rechavi, O. (2022) Plant and animal small RNA communications between cells and organisms. *Nat. Rev. Mol. Cell Biol.* 23, 185–203
22. Zhan, J. and Meyers, B.C. (2023) Plant small RNAs: their biogenesis, regulatory roles, and functions. *Annu. Rev. Plant Biol.* 74, 21–51
23. Todesco, M. *et al.* (2012) Natural variation in biogenesis efficiency of individual *Arabidopsis thaliana* microRNAs. *Curr. Biol.* 22, 166–170
24. Arif, S. *et al.* (2013) Evolution of mir-92a underlies natural morphological variation in *Drosophila melanogaster*. *Curr. Biol.* 23, 523–528
25. Bradley, D. *et al.* (2017) Evolution of flower color pattern through selection on regulatory small RNAs. *Science* 358, 925–928
26. Stubbe, H. (1966) *Genetik und Zytologie von Antirrhinum*, Fischer
27. Schemske, D.W. and Bradshaw, H.D. (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci. U. S. A.* 96, 11910–11915
28. Bradshaw, H.D. and Schemske, D.W. (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426, 176–178
29. Liang, M. *et al.* (2023) Taxon-specific, phased siRNAs underlie a speciation locus in monkeyflowers. *Science* 379, 576–582
30. Stanley, L.E. *et al.* (2020) A tetratricopeptide repeat protein regulates carotenoid biosynthesis and chromoplast development in monkeyflowers (*Mimulus*). *Plant Cell* 32, 1536–1555
31. McClintock, B. (1984) The significance of responses of the genome to challenge. *Science* 226, 792–801
32. Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405
33. Schrader, L. and Schmitz, J. (2019) The impact of transposable elements in adaptive evolution. *Mol. Ecol.* 28, 1537–1549
34. Clark, L.A. *et al.* (2006) Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *Proc. Natl. Acad. Sci.* 103, 1376–1381
35. Studer, A. *et al.* (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43, 1160–1163
36. Butelli, E. *et al.* (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255
37. van't Hof, A.E. *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105
38. Nadeau, N.J. *et al.* (2016) The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature* 534, 106–110
39. van der Burg, K.R.L. *et al.* (2020) Genomic architecture of a genetically assimilated seasonal color pattern. *Science* 370, 721–725
40. Livraghi, L. *et al.* (2021) Cortex cis-regulatory switches establish scale colour identity and pattern diversity in *Heliconius*. *eLife* 10, e68549

41. Wang, S. *et al.* (2022) The evolution and diversification of oakleaf butterflies. *Cell* 185, 3138–3152.e20
42. Tian, S. *et al.* (2024) A micro-RNA drives a 100-million-year adaptive evolution of melanic patterns in butterflies and moths. *bioRxiv*, Published online April 18, 2024. http://doi.org/10.1101/2024.02.09.579741
43. Fandino, R.A. *et al.* (2024) The ivory lncRNA regulates seasonal color patterns in buckeye butterflies. *bioRxiv*, Published online February 21, 2024. http://doi.org/10.1101/2024.02.09.579733
44. Livraghi, L. *et al.* (2024) A long non-coding RNA at the cortex locus controls adaptive colouration in butterflies. *bioRxiv*, Published online February 12, 2024. http://doi.org/10.1101/2024.02.09.579710
45. Santos, M.E. *et al.* (2014) The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat. Commun.* 5, 5149
46. Guo, Y.-L. *et al.* (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc. Natl. Acad. Sci.* 106, 5246–5251
47. Niu, X.-M. *et al.* (2019) Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc. Natl. Acad. Sci.* 116, 6908–6913
48. Yang, L. *et al.* (2018) Parallel evolution of common allelic variants confers flowering diversity in *Capsella rubella*. *Plant Cell* 30, 1322–1336
49. Yuan, Y.-W. and Wessler, S.R. (2011) The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci.* 108, 7884–7889
50. Arce, A.L. *et al.* (2023) Polymorphic inverted repeats near coding genes impact chromatin topology and phenotypic traits in *Arabidopsis thaliana*. *Cell Rep.* 42, 112029
51. Ishikawa, A. *et al.* (2019) A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 364, 886–889
52. Goodfellow, P.N. and Lovell-Badge, R. (1993) SRY and sex determination in mammals. *Annu. Rev. Genet.* 27, 71–92
53. Huu, C.N. *et al.* (2016) Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *eLife* 5, e17956
54. Huu, C.N. *et al.* (2020) Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proc. Natl. Acad. Sci.* 117, 23148–23157
55. Li, J. *et al.* (2016) Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nat. Plants* 2, 1–7
56. Berdan, E.L. *et al.* (2022) Genomic architecture of supergenes: connecting form and function. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20210192
57. Schwander, T. *et al.* (2014) Supergenes and complex phenotypes. *Curr. Biol.* 24, R288–R294
58. Gutiérrez-Valencia, J. *et al.* (2021) The genomic architecture and evolutionary fates of supergenes. *Genome Biol. Evol.* 13, evab057
59. Shore, J.S. *et al.* (2019) The long and short of the S-locus in *Turnera* (Passifloraceae). *New Phytol.* 224, 1316–1329
60. Yang, J. *et al.* (2023) Haplotype-resolved genome assembly provides insights into the evolution of S-locus supergene in distylous *Nymphoides indica*. *New Phytol.* 240, 2058–2071
61. Villoutreix, R. *et al.* (2020) Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science* 369, 460–466
62. Li, B. *et al.* (2020) A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *eLife* 9, e50608
63. Lowry, D.B. and Willis, J.H. (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* 8, e1000500
64. Kunte, K. *et al.* (2014) doublesex is a mimicry supergene. *Nature* 507, 229–232
65. Küpper, C. *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* 48, 79–83
66. Iijima, T. *et al.* (2018) Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Sci. Adv.* 4, eaao5416
67. Kess, T. *et al.* (2019) A migration-associated supergene reveals loss of biocomplexity in Atlantic cod. *Sci. Adv.* 5, eaav2461
68. Pearse, D.E. *et al.* (2019) Sex-dependent dominance maintains migration supergene in rainbow trout. *Nat. Ecol. Evol.* 3, 1731–1742
69. Jay, P. *et al.* (2021) Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat. Genet.* 53, 288–293
70. Jay, P. *et al.* (2022) Association mapping of colour variation in a butterfly provides evidence that a supergene locks together a cluster of adaptive loci. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20210193
71. Matschiner, M. *et al.* (2022) Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* 6, 469–481
72. Komata, S. *et al.* (2023) Functional unit of supergene in female-limited Batesian mimicry of *Papilio polytes*. *Genetics* 223, iyac177
73. Moczek, A.P. *et al.* (2011) The role of developmental plasticity in evolutionary innovation. *Proc. R. Soc. B Biol. Sci.* 278, 2705–2713
74. Posé, D. *et al.* (2013) Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* 503, 414–417
75. Haltenhof, T. *et al.* (2020) A conserved kinase-based body-temperature sensor globally controls alternative splicing and gene expression. *Mol. Cell* 78, 57–69
76. von Arnim, A.G. and Deng, X.-W. (1994) Light inactivation of *Arabidopsis* photomorphogenic repressor COP1 involves a cell-specific regulation of its nucleocytoplasmic partitioning. *Cell* 79, 1035–1045
77. Zavaliev, R. *et al.* (2020) Formation of NPR1 condensates promotes cell survival during the plant immune response. *Cell* 182, 1093–1108
78. Tang, M. *et al.* (2022) Mitotically heritable epigenetic modifications of CmMYB6 control anthocyanin biosynthesis in chrysanthemum. *New Phytol.* 236, 1075–1088
79. Khalturin, K. *et al.* (2009) More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25, 404–413
80. Kaessmann, H. (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326
81. Chen, S. *et al.* (2013) New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* 14, 645–660
82. Neme, R. and Tautz, D. (2013) Phylogenetic patterns of emergence of new genes support a model of frequent de novoevolution. *BMC Genomics* 14, 117
83. Paps, J. and Holland, P.W.H. (2018) Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nat. Commun.* 9, 1730
84. Xia, S. *et al.* (2021) Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in *Drosophila* development. *PLoS Genet.* 17, e1009654
85. King, M.-C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116
86. Deng, C. *et al.* (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc. Natl. Acad. Sci.* 107, 21593–21598
87. Santos, M.E. *et al.* (2017) Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science* 358, 386–390
88. Giorgianni, M.W. *et al.* (2020) The origin and diversification of a novel protein family in venomous snakes. *Proc. Natl. Acad. Sci.* 117, 10911–10920
89. Bista, I. *et al.* (2023) Genomics of cold adaptations in the Antarctic notothenioid fish radiation. *Nat. Commun.* 14, 3412