Alternative splicing as a source of phenotypic diversity

Charlotte J. Wright, Christopher W. J. Smith and Chris D. Jiggins

Abstract | A major goal of evolutionary genetics is to understand the genetic processes that give rise to phenotypic diversity in multicellular organisms. Alternative splicing generates multiple transcripts from a single gene, enriching the diversity of proteins and phenotypic traits. It is well established that alternative splicing contributes to key innovations over long evolutionary timescales, such as brain development in bilaterians. However, recent developments in long-read sequencing and the generation of high-quality genome assemblies for diverse organisms has facilitated comparisons of splicing profiles between closely related species, providing insights into how alternative splicing evolves over shorter timescales. Although most splicing variants are probably non-functional, alternative splicing is nonetheless emerging as a dynamic, evolutionarily labile process that can facilitate adaptation and contribute to species divergence.

Animal and plant taxa are characterized by an astonishing diversity of morphological characteristics. Morphological differences between species reflect a combination of differences in their underlying DNA and input from the environment. Since the 1970s, it has been observed that phenotypically diverse species can possess very similar coding sequences, which suggests that regulatory sequences play a dominant part in causing phenotypic differences[1]. Indeed, genome-wide association studies in humans have estimated that 88% of genetic variants associated with heritable features are located within non-coding genomic regions[2].

The importance of gene regulatory variation in morphological evolution has been studied extensively[3–5]. Defined broadly, regulatory variation includes any heritable variation in control at transcriptional or post-transcriptional levels that leads to proteomic diversity. Although transcriptional regulation has received more attention, many genetic variants regulate co-transcriptional and post-transcriptional mechanisms, such as alternative splicing (AS), polyadenylation and RNA editing[6–8], which enable a single gene to encode multiple transcripts, proteins and hence divergent phenotypes. In particular, AS is a potent mechanism for expanding the coding capacity of genes, by generating different protein isoforms, and for quantitative regulation of protein expression levels, for example by altering the balance between functional and non-functional transcripts or by regulating transcript stability. However, the role of AS in generating phenotypic diversity has received relatively little attention.

The discovery of splicing in mammalian cells in 1977 (REFS.[9,10]) immediately sparked interest in the implications for eukaryotic gene and protein evolution by 'exon shuffling', a mechanism enabled by the split gene structure[11,12]. Splicing is an essential step in the expression of intron-containing genes, but the subsequent discovery of AS revealed how it also facilitates an unanticipated form of gene regulation in which individual genes could generate functionally distinct protein isoforms[13,14]. Transcriptomic studies have established that AS is widespread across eukaryotes. For example, an estimated 90–95% of human genes undergo AS[15,16].

Variant transcripts can arise from seven major types of AS event (FIG. 1a; see also REF.[17]) as well as from alternative transcription start sites and alternative polyadenylation sites on the first and last exons, respectively[18,19]. Combinations of these binary events can lead to more complex AS events, such as cassette exons with alternative 5′ splice sites. Moreover, transcripts from many genes undergo multiple AS events that, if not tightly coordinated, can generate combinatorial diversity. The relative importance of each AS event type and overall extent of AS varies substantially between eukaryotic lineages[20–22] (FIG. 1b). For example, exon skipping is the most common form of AS in metazoans, whereas intron retention is more frequent in plants and fungi[23,24].

Questions about how AS evolves have been raised since the phenomenon was first discovered[25,26]. Landmark papers presenting the first multi-species, multi-tissue overviews of the evolutionary transcriptomic landscape indicated that compared to the high proportion of conserved tissue-specific transcriptional control, a smaller proportion of co-regulated tissue-specific AS events were conserved, with the highest level of conserved AS events occurring in the brain, heart, muscle and testes[27,28]. Although this could be interpreted to mean that AS is less functionally important than transcriptional control, an interesting alternative perspective is that AS potentially provides an opportunity for rapid evolutionary innovation. AS is known to affect many important phenotypes including neural development[29], muscle function[30] and pigmentation[31]. Nonetheless, a gap remains in our understanding as the majority of AS events remain functionally uncharacterized. Moreover, until recently, inferences from transcriptomic analyses about the effects of AS on the proteome were limited by widespread use of short-read RNA-sequencing, which provides robust quantification of AS events, but not full-length transcripts. Recent advances in long-read RNA-sequencing now allow complete transcripts to be sequenced. Long-read data can reveal the full complexity of alternative transcripts and the predicted protein isoforms they encode and provide insight into their influence on phenotypic change. For example, long-read RNA-sequencing identified 7,874 complete transcript isoforms of *Dscam1*[32], thousands of which are essential for normal neural circuit development[33]. The growing availability of transcriptomic data for diverse organisms has fuelled a resurgence of interest in the role of AS in generating phenotypic diversity, which was first discussed in REF.[34].

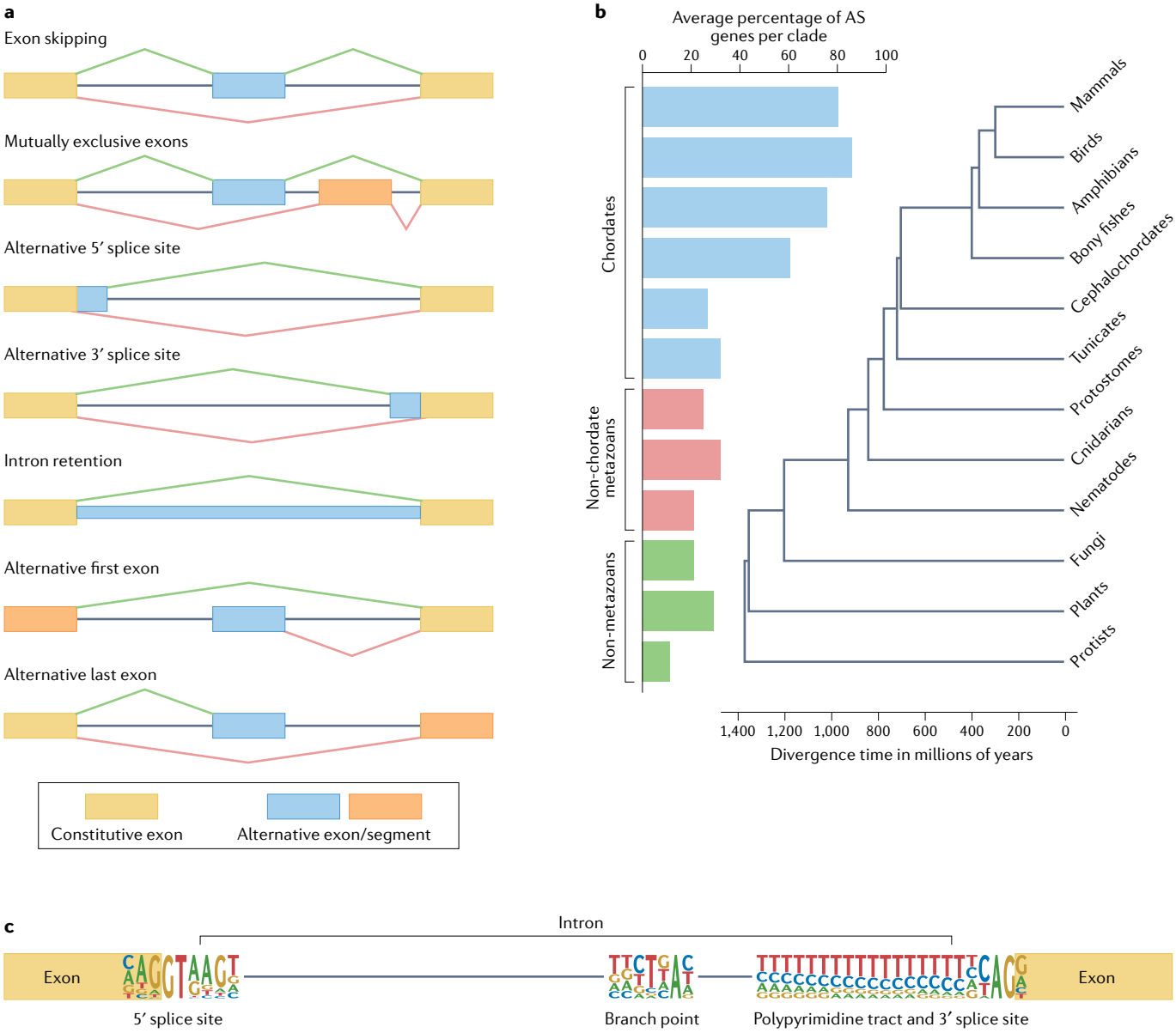


Fig. 1 | **Patterns of alternative splicing. a** | The major types of alternative splicing (AS) are exon skipping (removal of a single 'cassette' exon), mutually exclusive exons (when two or more adjacent exons are spliced such that only one exon from the group is included at a time), alternative 5′ or 3′ splice site (which affects the length of a particular exon), intron retention (an intron can be removed or retained) and alternative first or last exons. **b** | The prevalence of AS has increased during eukaryotic evolution. Bars reflect the average percentage of AS genes per clade. Phylogenetic data are from REF.[176]. **c** | The architecture of a pre-mRNA before splicing. Each intron contains a 5′ splice site and a 3′ splice site with its adjacent polypyrimidine tract and a branch point sequence (BPS) usually a short distance upstream of the 3′ splice site. The 5′ splice site, 3′ splice site and the BPS are described by consensus sequences, within which GU, AG and A, respectively, are nearly invariant. The degree to which individual splice signals match the consensus correlates with their functional strength. Consensus human sequences were made using WebLogo3[177]. Part **b** is reprinted from REF.[136], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Part **c**, image courtesy of Aishwarya Jacob.

In this Perspective, we summarize the ways in which AS occurs and the diverse functional roles it serves, namely in the context of proteomic diversity, development and phenotypic plasticity. We review the evolution of AS and the extent to which splicing variation can catalyse evolutionary change, including adaptation and speciation. Finally, we explore the evidence that AS diverges as a result of natural selection.

Mechanisms of alternative splicing

Eukaryotic pre-mRNA splicing is orchestrated by the spliceosome, which assembles stepwise across introns and then catalyses removal of the intron and splicing together of the protein-coding exons via its RNA-based catalytic centre[35]. Spliceosome assembly is precisely coordinated by splicing factors, which recognize short consensus sequences at the 5′ and 3′ ends of each intron (FIG. 1c). This process is regulated by a complex web of influences (BOX 1, and covered in detail in excellent recent reviews[36–38]). The spliceosome often assembles directly across introns, particularly when introns are short, in a process termed intron definition. However, when exons are flanked by long introns, early splicing complexes may instead first assemble between the splice sites flanking the exon, in a process termed exon definition[39]. The early exon definition complexes must subsequently

be juxtaposed and rearranged so that mature spliceosomes form across introns. Key evidence for exon definition is that mutations in a single splice site — introduced experimentally, or identified by clinical genetics — often lead to exon skipping rather than retention of just the affected intron[40]. From a regulatory standpoint, the path of splicing complex assembly determines the mode of splicing regulation; intron retention tends to be associated with intron definition, while exon definition is associated with exon skipping or inclusion and has become more prevalent during animal evolution[41]. From an evolutionary standpoint, the important corollary of exon definition is that sequence variants affecting a *cis*-splice signal frequently lead to inclusion or skipping of a complete exon and are therefore more likely to lead to viable alternative protein-coding isoforms. By contrast, intron retention — the most common form of AS in plants — more frequently leads to RNA degradation, although there is a small class of retained introns that are frame-preserving and lead to protein isoforms[42]. Despite the explanatory power of the concept of exon definition, sequencing of nascent RNAs indicates that many long introns are spliced while the downstream exon is still being transcribed, excluding a contribution for exon definition in those cases[43].

Functions of alternative splicing

AS is functionally important in many contexts. It diversifies the proteome and is a key regulatory mechanism during development. In addition, AS can facilitate responses to environmental change at all life stages.

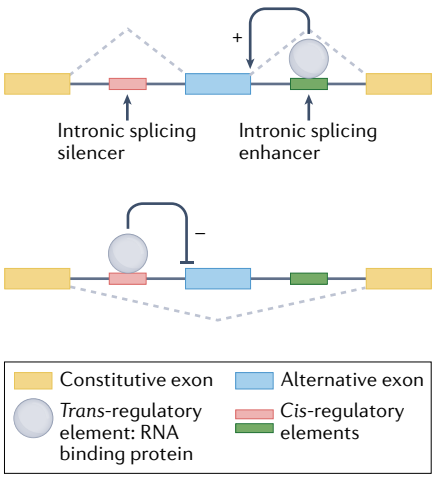
Contribution of AS to the proteome

The extent to which AS events translate into functional protein variation is the subject of intense debate (BOX 2). A number of lines of evidence suggest that the majority of observed AS events reflect splicing errors, and are neither conserved nor functional[44–48]. However, there is clear evidence that a subset of AS events contribute to functional protein diversity and to regulation of protein expression levels[49–52]. Moreover, the preponderance of non-functional, noisy AS events provides the potential for subsequent evolution of new function. Alternative transcripts can contain different open reading frames, leading to protein isoforms varying in domain architecture, binding sites, stability, activity or localization[53–55]. For example, AS of the *Drosophila melanogaster* transcription factor *lola* generates a family of 19 transcription factors (which differ in their DNA-binding properties) that regulate neural wiring[56]. Overall, AS increases both the total protein number and the degree of interconnectivity within protein networks (reviewed in REF.[34]). Indeed, AS may be used in different ways by different organisms. For example, systematic analysis of AS in multiple tissues and stress conditions in *Arabidopsis thaliana* and comparison with three animal species suggested that plants use AS mainly for quantitative gene regulation in response to stresses, whereas animals use it mainly to generate tissue-specific proteomes[57].

Tissue-specific cassette exons often encode intrinsically disordered protein regions, which typically have greater functional flexibility and are enriched in post-translational modification sites and conserved protein binding motifs[58–60]. Mutually exclusive exons frequently affect globular domains, including enzymatic domains, and structural modelling indicates how protein–protein interactions can be fine-tuned by isoform switching[61]. For example, a highly conserved AS event regulates DNA-binding preference and hence the function of transcription factor FOXP1[62]. In human embryonic stem cells, FOXP1 transcripts include exon 18b, resulting in isoforms that upregulate key pluripotency genes including *OCT4* and *NANOG*. By contrast, cell differentiation is promoted when exon 18 is included in *FOXP1* transcripts[62]. Cross-species comparisons revealed that exons 18 and 18b of human FOXP1 are highly conserved across 46 vertebrate species, indicating conservation of AS-mediated pluripotency maintenance[62].

In addition to switching protein isoforms, AS also has quantitative effects on protein expression levels. AS can affect mRNA translation efficiency by altering untranslated region sequences[63,64] and transcript stability by coupling to mRNA degradation mechanisms[65]. Intron retention

Box 1 | Components of the splicing code



The core splicing signals are described by consensus sequences, within which a limited number of positions are invariant (FIG. 1c). The degree to which individual splice signals match the consensus correlates with their functional strength. Core splicing signals alone are insufficient to distinguish authentic splice sites unambiguously from the many similar sequences within introns (except in unicellular eukaryotes, which have shorter introns, tighter consensus splice sites and display limited alternative splicing (AS))[185]. A large array of auxiliary *cis*-regulatory elements — termed exonic or intronic splicing enhancers and silencers, depending on their location and mode of action — provide the additional specificity that distinguishes authentic from cryptic splice sites. These elements, which are associated with both constitutive and AS exons, act as binding sites for splicing regulatory RNA-binding proteins (RBPs) — *trans*-regulatory factors that can either activate or repress splicing to nearby splice sites, and provide the opportunity for regulation of AS via the availability of the cognate RBPs. Auxiliary elements are extremely diverse, reflecting the large number of RBPs, each with a specific binding sequence preference. Most RBPs are functionally versatile, acting as repressors or activators depending upon the location of their binding sites relative to regulated splice sites[38,186] (see figure). For example, the vertebrate neuronal proteins Nova1/Nova2 bind specifically to YCAY motifs within transcripts and, depending on the motif location, can promote or prevent exon inclusion[187–191].

AS exons typically have weaker splice sites than constitutive exons[99], relying upon additional activating elements. Indeed, specific combinations of enhancer and silencer motifs constitute a large part of tissue-specific splicing codes, with the code being interpreted and executed by the cognate RBPs. In addition to cellular complements of RBPs, many other factors can determine AS patterns, including: variations in the levels of core splicing factors, such as snRNP proteins; RNA secondary structure, which can 'hide' splice site or regulatory elements, optimally display these elements in loops, or can bring distantly separated elements into close juxtaposition[192,193]; epitranscriptomic RNA modifications, such as adenosine N6 methylation, adenosine deamination to inosine or pseudouridylation[194,195]; changes in RNA Pol II processivity, sometimes brought about by chromatin modifications, can alter the 'window of opportunity' within which an exon with weak splice sites can be recognized[36]; and finally both RNA Pol II and chromatin marks can help to recruit core and regulatory splicing factors[36,169].

Box 2 | **To what extent does AS contribute to functional protein diversity?**

Deep-sequencing of transcriptomes identifies vast numbers of alternative transcripts[196]. However, it is unclear how many lead to functionally relevant protein isoforms. Mass-spectrometry-based proteomic analyses have frequently failed to detect evidence of the protein isoforms corresponding to mRNA variants, leading to the suggestion that a substantial fraction of alternative splicing (AS) may be non-functional[197]. The discrepancy between proteome and transcriptome data may be due, at least in part, to technical reasons[48,198]. Approximately 75% of human exon-skipping events detected by RNA-sequencing were also identified by ribosome profiling, suggesting that these events result in protein isoforms[64,199]. Similarly, 40% of the splicing changes at RNA level were concordant with changes in translation detected by ribosome profiling in human glia and glioma[200]. Nevertheless, ribosome occupancy alone is not sufficient to demonstrate functionality, because not all translation events result in stable, functional proteins[45,201]. Accumulating evidence suggests that AS may be noisy, with a large fraction of AS events reflecting erroneous splice site usage. For example, the number of isoforms increases with gene expression level and number of introns, consistent with a stochastic noise model[202]. Rarely used splice sites also show little evidence of conservation, which has been interpreted as a lack of function[46]. Recent genome-wide studies also indicate that much of the diversity of transcripts generated by alternative polyadenylation, RNA editing and transcription initiation is also driven by deleterious, molecular errors[203–205].

Although a large proportion of AS events are unlikely to contribute to functional protein production, a subset of AS events do have important functions. For example, of the one-third of human-tissue-specific AS isoforms that were detected in proteomic datasets, 95% were highly conserved across vertebrates[206]. The fact that networks of AS that are co-regulated during phenotypic transitions or in response to splicing factor knockdowns affect gene sets showing clear enrichments for biological function also strongly argues for the functional importance of much regulated AS[161]. Transcriptomic studies suggest AS enhances the complexity of regulatory networks; genes with higher levels of AS tend to exhibit a higher number of protein–protein interactions[207], and genes with tissue-specific AS events typically occupy a more central position in protein–protein interaction networks[59]. For example, genes that are differentially spliced between ecotypes of salmonid fish that differ in swimming ability tended to be highly central in regulatory networks[160]. Isoforms produced by AS have different interaction profiles and these differences can be as great as between proteins encoded by different genes[208].

Further work is required to distinguish which transcripts produce biologically important isoforms and to characterize their functions. Machine learning methods, trained with proteomics data as a proxy for functionality, have proved capable of predicting biologically important isoforms with high accuracy[209]. Functional characterization of AS events remains challenging because many splice variants are likely to have important biological effects but be restricted to specific cell types or conditions, and until recently the technologies were not available for directed manipulation of splicing events in organisms[210]. CRISPR–Cas9 genome editing and variants that target RNA now allow experimental tests for functional effects of isoforms. For instance, blocking the AS of a key exon of *titin* in mice causes a dystrophic phenotype in tissues that express the isoform[211]. Systematic approaches that integrate multiple sources of data are the most promising approach to cataloguing AS diversity and functionality, as demonstrated by detailed analysis of human G protein-coupled receptor isoform function[212].

in particular frequently leads to nuclear retention or the introduction of premature termination codons (PTCs) and cytoplasmic nonsense mediated decay (NMD). Although primarily considered as a mechanism that deals with aberrant transcripts, NMD also acts upon numerous mRNAs that contain PTCs as a result of regulated AS[66]. Retention of introns containing PTCs, inclusion of poison exons with PTCs[67,68], and frameshift-inducing exon-skipping leading to PTC incorporation[69–71] can all lead to NMD. Coupled AS–NMD is frequently used in autoregulation and cross-regulation between splicing factors. For example, PTBP1 induces partial skipping of a frame-preserving exon in *PTBP1* mRNA in an autoregulatory loop that limits PTBP1 levels, and near complete skipping of the equivalent exon in the paralogue *PTBP2* to keep PTBP2 expression switched off[65]. Strikingly, the entire family of human splicing regulatory SR proteins contain poison exons within ultraconserved regions (originally defined as >200-base pair (bp) regions with complete sequence identity between human, mouse and rat[67,68]) that mediate extensive cross-regulation between SR family members. These regulatory motifs are ancient; the 'unproductive splicing' of SRSF5 is conserved among all animals and even in fungi[72]. The functional importance of regulated poison exons has been demonstrated directly by high-throughput CRISPR–Cas9-based targeted deletion studies[73]. Remarkably, many poison exons are essential for cell fitness, but deletion of a subset of poison exons also led to enhanced tumorigenesis in mouse xenograft models, underscoring the importance of AS in quantitative gene regulation. It should be noted that not all predicted AS–NMD isoforms are observed to undergo NMD[74], and the rules governing NMD and the transcript features involved are still to be fully elucidated[75,76]. Nonetheless, there are many clearly documented examples of functional AS–NMD in the regulation of gene families, particularly of splicing factors[77].

AS regulates development

The consequences of AS at the molecular level can manifest as major effects at the physiological system level. One of the most fundamental developmental processes that can be controlled by AS is sex determination. The cascade of AS events that regulate sex determination is best understood in *Drosophila melanogaster* (FIG. 2a). Remarkably, sex-specific splicing of *dsx* — a key sex-determining gene in *Drosophila* spp. — is conserved across diverse insects, including the silkmoth *Bombyx mori*, olive fruitflies *Bactrocera oleae* and red flour beetle *Tribolium castaneum*, suggesting an ancient origin and role in sex determination[78–80] (reviewed in REF.[81]). Recent work has also revealed the role of AS in sex determination in poikilothermic reptiles in response to small variations in temperature[82] (FIG. 2b). In turtles, at temperatures below 26 °C, the regulatory kinase CLK1/4 phosphorylates SR proteins that regulate splicing[82]. This leads to nuclear localization of the SR proteins, which in turn switches a set of regulated AS events, ultimately resulting in male development. Above 31 °C, CLK1/4 is inactive, leading to AS events associated with female development. At intermediate temperatures, a mixture of both sexes develop[83]. Remarkably, the activation temperature of CLK1/4 is finely tuned to respond to relevant temperature ranges across a variety of poikilotherms, and to physiological cold- and heat-shocks and circadian-regulated AS in homeotherms[82,84].

AS mediates phenotypic plasticity

Temperature-dependent sex determination is just one example of phenotypic plasticity, whereby organisms alter their phenotype in response to environmental cues. AS can facilitate phenotypic plasticity by changing the expression of alternative transcripts of genes influencing developmental trajectories, or modulating their overall expression levels. For example, AS has a key role in modulating flowering time in *A. thaliana* in response to temperature, by regulating expression of the transcription

factor FLOWERING LOCUS M (FLM). The underlying mechanism was originally suggested to involve a mutually exclusive switch from the functional FLM-ß isoform, which is produced at low temperatures (16 °C) and interacts with the floral repressor SVP, to a non-functional FLM-δ isoform at high temperatures (27 °C)[85]. More recent data suggest that an AS switch reduces FLM-ß levels through the generation of a variety of non-productive mRNA isoforms rather than the FLM-δ isoform[86–88].

Remarkable phenotypic plasticity is also seen in the caste system of eusocial insects, in which various adult morphologies are generated from the same genome. Comparative analyses suggest that AS is a contributing factor to caste differences in termites[89], honeybees[90] and potentially ants[91]. In the buff-tailed bumble bee, *Bombus terrestris*, 40% of genes express multiple isoforms, many of which are caste-specific[92]. AS has also been suggested to have a complementary role in gene expression changes in seasonal plasticity in the butterfly *Bicyclus anynana*[93]. However, functional studies into the causal role of AS are still to be carried out in all these systems.

How novel splice patterns arise

The growing evidence for functional roles of AS in shaping phenotypic traits makes understanding its molecular evolution and response to selective pressures a key goal. Novel splice patterns arise when mutations occur that affect splice site selection and spliceosome assembly. These mutations can be divided into *cis*-effects, where the mutation is within the gene undergoing splicing, and *trans*-effects, where a core splicing or regulatory gene is mutated[94–96] (FIG. 3). The main *cis*-effects are: duplication of existing exons; 'transition', whereby a constitutive exon is converted to an alternative exon, or vice versa; exonization, whereby mutations convert an intronic sequence into an exon; or conversely intronization, whereby mutations lead to complete loss of splicing and the previously exonic sequence becomes part of an intron[97–100]. Mechanistically, intronization can be seen as an extreme form of transition whereby the exon becomes so weak that it is always skipped. The direct impacts of *trans*-effects can be either qualitative (altering the coding sequence of the *trans*-factor) or quantitative (changing its expression levels), resulting in a wave of quantitative changes in expression level of AS transcripts.

Exon duplication

Exon duplication, coupled to mechanisms to prevent inclusion of both exons, is the source of mutually exclusive exons (MXEs) (FIG. 3a). One copy of the exon retains ancestral function while the other can undergo change, leading to a new function. For example, in both

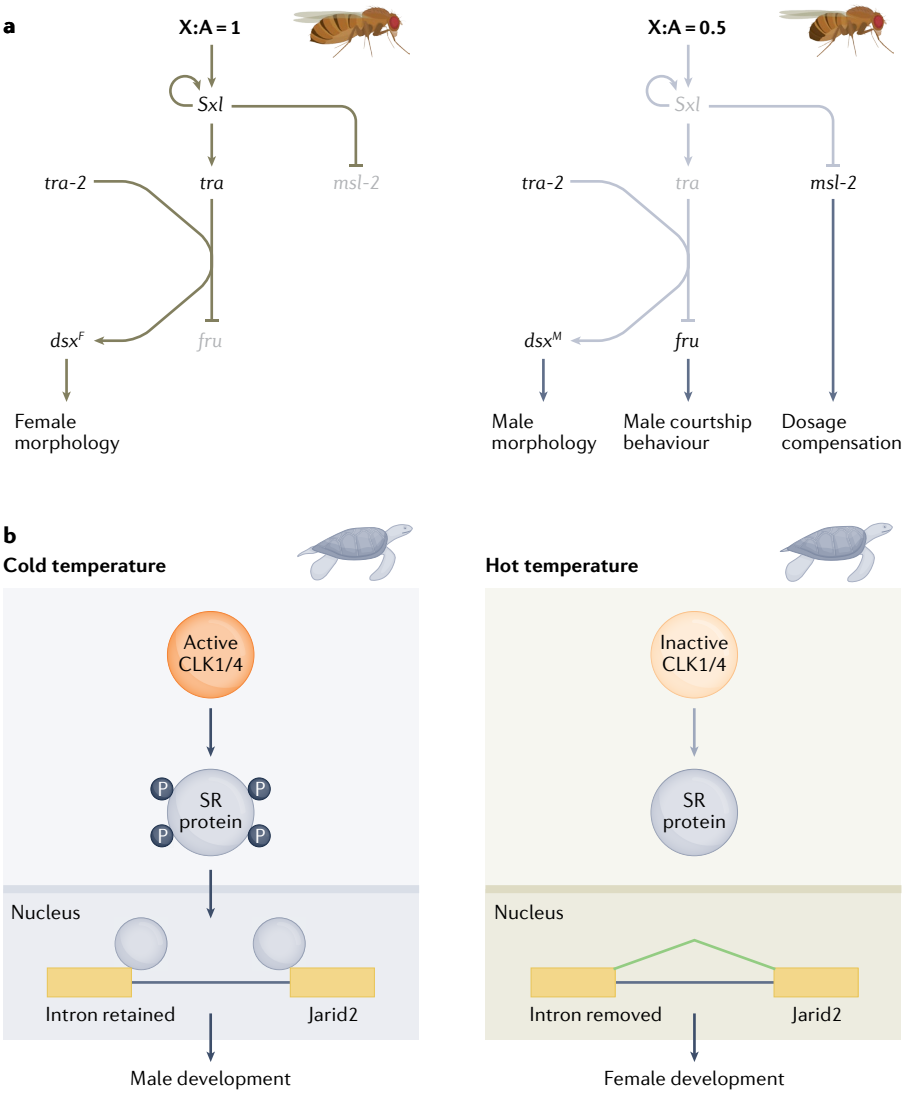


Fig. 2 | **Role of alternative splicing in sex determination. a** | In *Drosophila melanogaster*, the presence of two X chromosomes, resulting in a 1:1 ratio of X chromosomes to autosomes (X:A = 1), leads to expression of the negative splicing regulator sex-lethal (sxl)[178] and to female development (left); in the presence of one X-chromosome, and an X:A ratio of 0.5, sxl is not expressed and male development occurs (right). Sxl promotes the use of a distal 3′ splice site during splicing of *transformer* (*tra*) transcripts[179]. The resulting *tra* transcripts lack a premature termination codon, leading to female-specific expression of Tra (REFS.[180,181]). Tra in conjunction with Tra-2 positively regulates splicing of *doublesex* (*dsx*), leading to inclusion of a female-specific exon and the isoform that orchestrates female development[182]. Tra also interacts with Tra-2 to inhibit expression of Fru, a protein that establishes male courtship behaviour. Sxl ensures that dosage compensation is restricted to males by promoting degradation via intron retention in *male-specific lethal 2* (*msl2*), which upregulates transcription of the X chromosome[183]). Sxl maintains its expression through a splicing-mediated positive feedback loop that is self-sustaining even after the initiating X:A signal has been erased by dosage compensation, thus ensuring that sex fate choice is maintained. Light grey lines and text reflect proteins and processes that are not expressed. **b** | Sex determination in turtles is governed by temperature-dependent changes in alternative splicing. At cooler temperatures (<26 °C), active CLK1/4 phosphorylates SR proteins, causing nuclear localization. SR proteins then regulate splicing events, resulting in protein isoforms that orchestrate male development (left). For example, SR proteins promote intron retention in the chromatin modifier Jarid2. At warmer temperatures (>31 °C), CLK1/4 is inactive and so SR proteins are not phosphorylated and remain cytoplasmic. This prevents SR proteins from regulating splicing of targets — the resulting protein isoforms mediate chromatin modifications associated with female development (right).

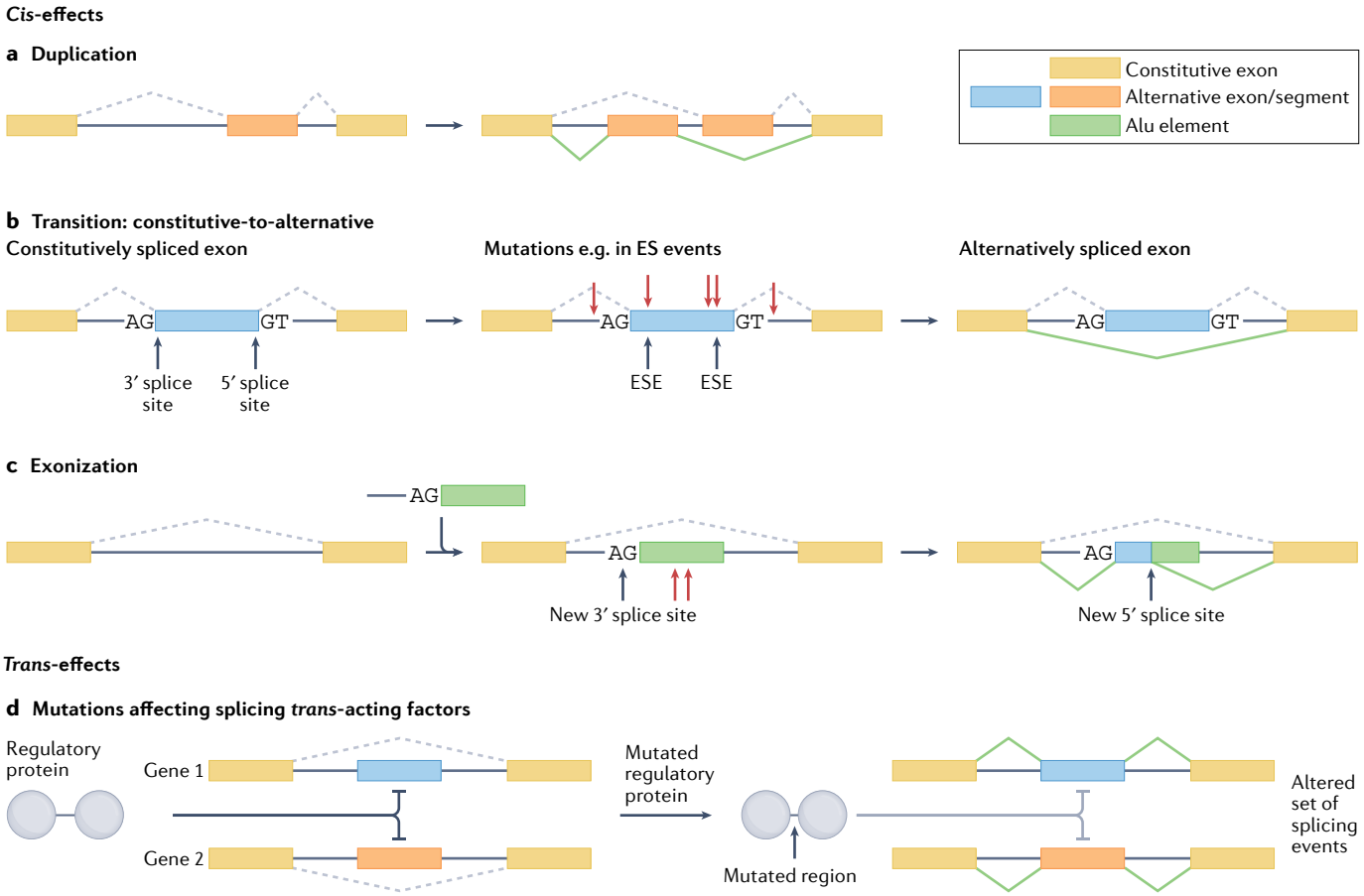


Fig. 3 | **Genetic basis of novel splice forms and patterns. a–c** | Novel splice variants result from *cis*-effects (which occur in or near the gene undergoing splicing). *Cis*-effects include: duplication of an existing exon, often resulting in mutually exclusive exon pairs (part **a**); transition, caused by mutations that weaken the splice sites of a constitutive exon, such as in exonic splicing enhancers (ESEs), leading to exon skipping (part **b**); and exonization, which occurs through insertion of a transposable element that subsequently acquires splice sites through mutations (part **c**). Taken to the extreme, this can result in intronization of a previously exonic sequence. **d** | *Trans*-effects occur when loci involved in regulating splicing are mutated, and often lead to changes in the expression levels of an array of existing splice variants. Mutations to *trans*-acting regulatory proteins can lead to an altered set of splicing events. Parts **a–c** adapted with permission from REF.[184], Wiley.

α-actinins and oxoglutarate dehydrogenase, MXE exons affect calcium binding and responsiveness[101,102]. Calcium binding by the ancestral isoform is lost in the more recent isoform, allowing muscle actinin to form stable structures suitable for muscle contraction in response to activating calcium. Many MXEs are ancient and highly conserved across lineages[103]. However, in the extensive arrays of MXEs in insect *Dscam* genes, variants within individual arrays are more closely related to each other than to the equivalent arrays in other insect species, indicating that these arrays have arisen by multiple independent duplication events[104].

Transition

Transition events can lead to modular loss of protein sequence and thus function (FIG. 3b). In principle, transition could occur via any mutation in the consensus sequences or regulatory splicing elements (BOX 1) that impairs splicing of the exon, although sufficient levels of exon inclusion will be needed to maintain ancestral function. An interesting example of transition during primate evolution is provided by exon 6 of *FAS*. The ancestral FAS isoform is a membrane-bound receptor that promotes apoptosis. A primate-specific isoform lacking exon 6 is a soluble circulating isoform that acts as an antagonistic decoy, competing with the ancestral isoform for binding to the Fas ligand. Analysis of the contribution to exon skipping of the 12 exonic sequence variants that separate the human alternative exon from the constitutively spliced primate ancestor mathematically defined a complex nonlinear relationship between the effects of individual mutations on exon inclusion that was context-dependent, with the nonlinearity arising from competition between splice sites[105]. Similarly, recent machine learning approaches found that only a small minority of exons in the human exome undergo large changes in inclusion level in response to exonic variants. Instead, most gradually change their inclusion level with subsequent exonic mutations[106]. Taken together, these studies indicate that transition may occur by a series of stepwise changes in response to sequence variants affecting splicing enhancers or silencers. Compared to the more catastrophic effects of mutations in consensus splice site sequences, which account for 10% of human inherited disease[107,108], this progression via incremental changes might be more amenable to maintaining ancestral isoform function while sampling the potential function of new isoforms.

Exonization

Many exonization events arise through creation of splice sites in short retrotransposons or by gradual loss of silencer sequences in long retrotransposons[109–112] (FIG. 3c). For example, a large number of

lineage-specific exons have arisen from *Alu* elements in primate genomes. Exons arising by exonization generally have low inclusion levels, consistent with maintenance of ancestral function, and are thought to be generally restricted to regulatory functions such as downregulation via AS–NMD[113–117]. Systematic analysis of proteomic data has identified up to 33 human *Alu*-derived exons that contribute to protein isoforms[118,119], some of which are included at high levels and/or are dynamically regulated, but unambiguous evidence for function is lacking.

Although repetitive elements are well appreciated as a source of new exons, less than 20% of exonization events in mice and humans arise from short interspersed elements (SINEs), and it has been suggested that the majority of new exons occur from unique, rather than repetitive, sequences[120]. A common class of lineage-specific new exons occurs as the second exon in their transcripts and affect 5′ untranslated regions. Inclusion of these exons is associated with higher transcription levels, highlighting a mechanism by which promoter-proximal splicing can increase transcription levels and even activate new promoters, further enhancing isoform diversity[120,121].

Balance of cis- and trans-effects

The evolutionary consequences of changes to AS directed by *cis*- and *trans*-effects are fundamentally different. Alterations to the *cis* inputs generally lead to localized changes in an individual splicing event, whereas perturbations in the *trans* components are likely to lead to widespread changes in splicing[122] (FIG. 3d). Despite this distinction, the relative importance of *cis* versus *trans* regulatory mutations to AS evolution is unclear.

Mutations at *trans*-acting regulatory loci could be an efficient way of altering a suite of transcripts (and hence proteins) in a coordinated manner[123]. In support of this idea, various functional differences in *trans*-acting regulatory factors have been identified. For example, exon 9 of splicing regulator PTBP1 is sometimes skipped in mammals, whereas it is always included in chickens[27]. Exon skipping removes 26 amino acids of a linker between the second and third RNA binding domains and reduces PTBP1 repressor activity, which in early neuronal differentiation leads to a wave of AS changes[124]. Consistent with this mechanism, deletion of the orthologous exon in chicken cells resulted in mammalian-like AS changes. This example highlights how a *cis*-driven AS change that directly modulates the activity of a splicing regulator can lead to a wider set of splicing changes in *trans*. However, mutations affecting the coding sequence of *trans*-acting regulators are likely to be deleterious, owing to their pleiotropic nature. For example, mutations in the human splicing factors SF3B1, U2AF1 and SRSF2 are associated with myelodysplasias featuring widespread splicing changes[125]. Likewise, experimental depletion of splicing repressors such as hnRNPC leads to widespread and disruptive exonization of Alu elements[126]. Nevertheless, it is also clear that major evolutionary innovations in AS can be driven in *trans*. A notable example is acquisition in early bilaterians of a novel 'enhancer of micro-exon' (eMIC) domain in the SRRM2/3/4 splicing factor that facilitated the subsequent evolution of AS programmes of neuron-specific micro-exons, which are among the most conserved and switch-like AS events in mammals and are needed for proper brain development and function[127].

Overall, although functional changes to splicing regulators do occur between species, *cis*-acting effects seem to be more common[27,128,129]. For example, 77% of strain-specific AS quantitative trait loci (QTL) in *Caenorhabditis elegans* were locally regulated, suggesting a *cis*-regulatory basis[130]. Similarly, most vertebrate species-specific splicing patterns were found to be under *cis* control[27]. However, further work is required to fully understand the basis of variation in AS patterns, because some studies find a larger effect of *trans*-regulatory variation (for example, during sunflower domestication[123]).

Finally, evolution does not always have to wait for de novo mutations. Instead, novel isoforms can arise from either introgression or from standing genetic variation. This is equally true for novel splice forms. For example, domestication of sunflowers has resulted in part from introgression of domesticate-associated splicing patterns and from changes in frequency of splicing patterns that existed among standing genetic variation in close wild relatives[123]. Overall, changes to AS can originate in a plethora of ways, suggesting that AS is a relatively easy target for mutation and has substantial scope for evolutionary change.

AS as a substrate of evolution

Eukaryotes show enormous variation in phenotypic diversity and complexity. The functional roles of AS, together with the many ways in which AS can evolve, make it probable that it has played a part in the evolution of both complexity and diversity.

The evolution of complexity

The genetic basis of complexity — defined as the number of distinct cell types in an organism — has been debated since comparative studies found that the total number of protein-coding genes cannot account for the increased cellular diversity observed in more complex eukaryotes. For example, both the human genome and the genome of the roundworm *C. elegans* have about 20,000 protein-coding genes[131,132]. Numerous genomic features have been proposed to account for the poor correlation between organism complexity and total gene content ('the G-value paradox'[133]), including AS, microRNAs, long non-coding RNAs and non-coding DNA[34,134,135]. Of these features, AS is a particularly attractive candidate as, by definition, it allows multiple transcripts and thus proteins to stem from a single gene.

Comparative transcriptomic studies revealed extensive differences in the extent of AS between eukaryotes. Direct comparison of levels of AS across 47 diverse eukaryotic species (spanning protists, fungi, plants and animals[136]) revealed that the prevalence of AS (defined as the proportion of AS in multiexonic genes) was strongly correlated with organism complexity, with the highest levels in vertebrates (FIG. 1c). Importantly, this analysis accounted for differences in transcript coverage and found AS to be a strong predictor of organism complexity regardless of effective population size[136] (see REF.[34]). However, it should be noted that this finding is based on just 12 species and so requires further study. In another study, analyses across 65 eukaryotic species found that exon skipping frequencies have increased during evolution of bilaterians and this increase is associated with a gene architecture expected of exon definition during spliceosome assembly[41]. However, further work is needed to understand the importance of AS scaling with complexity, especially given that a large proportion of AS events is expected to be non-functional.

AS in rapid evolutionary change

Understanding the genetic basis of morphological change remains a central challenge in evolutionary developmental biology. As just a few mutations can alter AS events and thus phenotype, divergence in AS has the potential to facilitate rapid phenotypic changes during evolution. Important questions include whether loci underpinning splicing show evidence of accelerated evolution relative to other loci, and to what extent changes in AS profiles

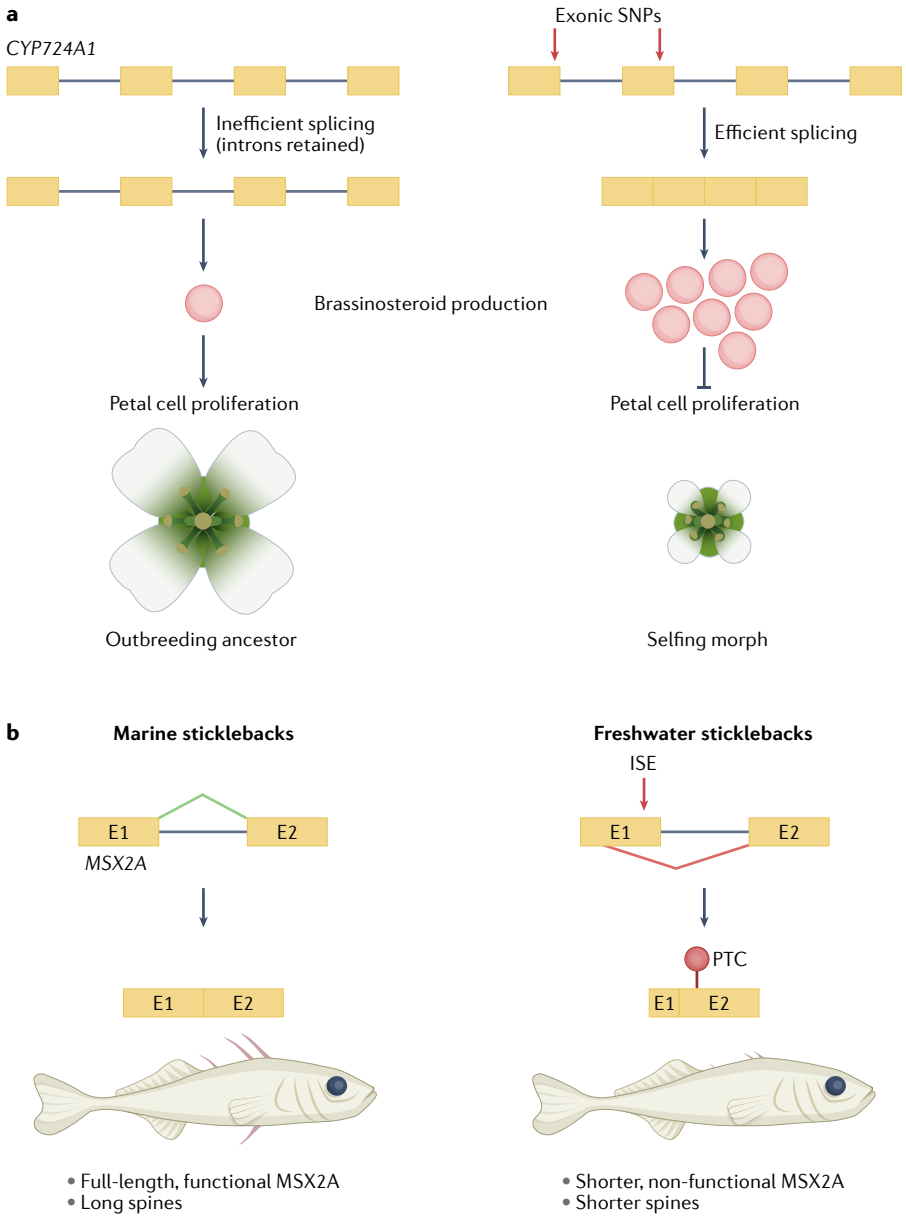


Fig. 4 | **Alternative splicing can underpin morphological evolution. a** | Evolution of the selfing morph of *Capsella*. A key phenotypic change in the evolutionary transition from outbreeding to selfing is a reduction in petal size. In the ancestral outbreeding morph, an optimal level of brassinosteroids facilitate the development of large petal sizes. Levels of brassinosteroids are low because the gene encoding the brassinosteroid-biosynthesis enzyme, CYP724A1, is inefficiently spliced. However, two SNPs in the selfing morph have led to efficient splicing of introns in *CYP724A1*, resulting in higher brassinosteroid levels that inhibit petal-cell proliferation and lead to reduced petal size. **b** | Evolution of spine length in threespine sticklebacks. A change in splicing in *MSX2A* underpins the adaptive shift from long spines in marine environments to shorter spines in freshwater environments[147]. Marine sticklebacks have long spines and express full-length MSX2A, a homeodomain transcription factor that regulates spine length. By contrast, freshwater sticklebacks produce a high proportion of truncated, non-functional MSX2A, resulting in shorter spine lengths. This change in splicing is caused by a single mutation within exon 1 of the freshwater *MSX2A* allele, which creates an intronic splicing enhancer (ISE) that activates an alternative 5′ splice site in exon 1. Use of this splice site results in a frameshift in exon 2 and the introduction of a premature termination codon (PTC). Part **a** is adapted with permission from REF.[141], Elsevier.

account for inter-specific phenotypic divergence. We shall use the term 'divergent splicing' to refer to divergence in profiles of splicing between lineages[123].

Divergent splicing and speciation. Divergent splicing has been identified between morphs of the same species and between closely related species, for example between mice subspecies that diverged around 0.5 million years ago[137] and between the domesticated sunflower and its wild ancestor since their divergence around 5,000 years ago[123]. Divergent splicing is also associated with the recent divergence between human head and body lice, *Pediculus humanus*, which inhabit distinct ecological niches and vary in their ability to transmit disease[138]. Although genome sequencing and expressed sequence tag data found few differences between the two types[139,140], deep transcriptomics revealed over 3,500 AS events specific to either head or body lice[138]. Differences in AS were associated with salivary gland processes that could influence vector competence between the two lice species, although functional studies remain to be carried out[138]. Patterns of AS can therefore diverge rapidly in association with incipient speciation, but in all these cases, it remains to be seen how much of this divergence is functionally relevant to the process of speciation.

One direct way in which divergent splicing can contribute to rapid species divergence is by causing phenotypic changes that reduce gene flow. For example, the flowering plant *Capsella rubella* evolved into a selfing species from its outbreeding ancestor, *Capsella grandiflora*[141] (FIG. 4a). The defining phenotypic change was a reduction in petal size in the selfing morph. QTL analysis identified the brassinosteroid-biosynthesis enzyme *CYP724A1* as the causal gene. Two SNPs in *CYP724A1* of the selfing-morph caused more efficient removal of introns, leading to higher CYP724A1 levels and therefore brassinosteroid levels. Higher brassinosteroid levels restrict petal-cell proliferation, leading to reduced petal size. These SNPs were shown to be sufficient for petal reduction, given that when the *C. rubella* allele was expressed in *A. thaliana*, petal size was reduced whereas the *C. grandiflora* allele had no effect. Consequently, a single change in splicing efficiency, which results in smaller petals in *C. rubella*, mediated the transition from outbreeding to selfing.

Divergence in splicing can drive adaptation. In some cases, the differential splicing of individual large-effect genes has been linked to the evolution of adaptive traits and life history strategies. For example, the genetic basis of the parasitic Cape honeybee workers, *Apis mellifera capensis*, has been linked to a change in AS of the transcription factor-encoding gene, *gemini*[142,143]. Typically, only mated queens can produce daughters

because ovary activation in workers is inhibited by pheromones[144]. However, workers occasionally evade this control, leading to offspring production that threatens the queen's position in the hive[145]. A 9-bp deletion of a potential intronic splicing regulatory element in *gemini* in parasitic workers correlates with the increased usage of an alternative 5′ splice site in exon 7. The resulting in-frame loss of 8 amino acids in the CP2 DNA-binding domain is predicted to lead to altered gemini activity and hence higher expression of genes involved in ovary activation. In support of exon 7 splicing influencing reproductive status, queens possess a similar ratio of exon 7 splice site selection to parasitic workers. The inclusion level of *gemini* exon 5 is also thought to alter ovary activation in the closely related subspecies *A. mellifera carnica*.

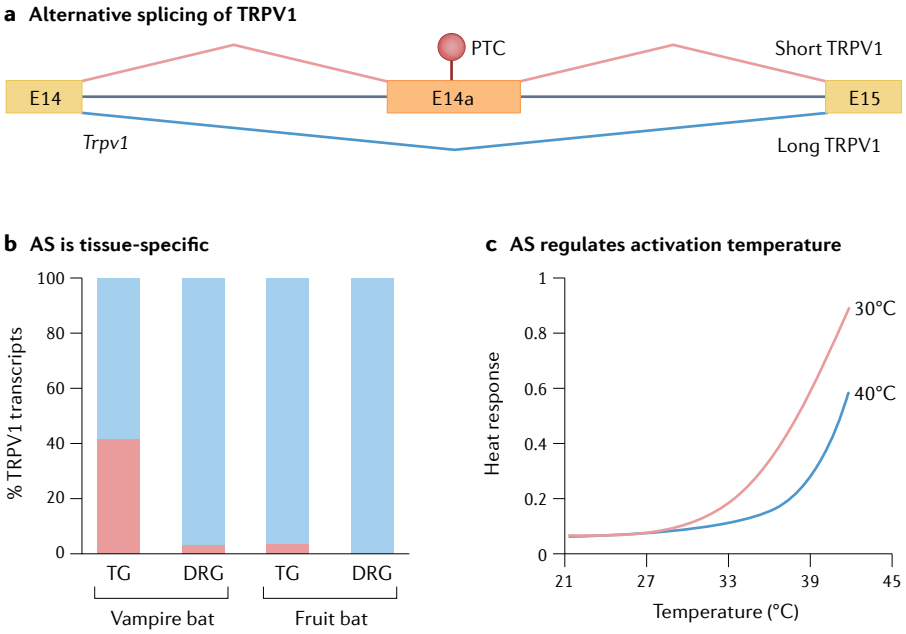
Often, changes in splicing regulate traits by producing non-functional proteins, rather than by altering or acquiring a new function. For example, predator-driven divergent selection has resulted in the threespine sticklebacks *(Gasterosteus aculeatus)* repeatedly evolving shorter spines than their marine counterparts, owing to changes in splicing of *MSX2A*, which encodes a homeodomain transcription factor[146,147] (FIG. 4b). In marine sticklebacks, full-length MSX2A protein is produced, whereas in freshwater sticklebacks, a high proportion of *MSX2A* transcripts encode non-functional truncated protein owing to the absence of the DNA-binding domain encoded by exon 2. The freshwater allele contains a non-synonymous (Glu to Gly) A to G transition, generating a poly-G tract splicing enhancer element that activates an alternative 5′ splice site within exon 1. The role of splicing differences in regulating MSX2A and hence driving rapid adaptive divergence in spine production was confirmed by transgenic expression of stable full-length MSX2A in freshwater sticklebacks, which led to increased spine size[147]. Similarly, the genetic basis of adaptation to food scarcity in cavefish has been linked to a change in AS of *per2*, resulting in reduced activity of the Per2 protein[148]. Cave-dwelling populations of *Astyanax maxicanus* tolerate periods of low food availability by upregulating the expression of lipogenesis genes when food is abundant, leading to increased body fat accumulation compared to river-dwelling populations. Upregulation of lipogenesis genes is achieved through increased expression of the lipid metabolism regulator Pparγ in the liver. Per2 normally acts to repress Pparγ activity. However, skipping of exon 21 and introduction of a PTC in *per2* mRNA leads to the loss of 160 amino acids adjacent to the C-terminal Pparγ-binding domain of Per2 in the Pachón cavefish population, suggesting that a simple change in splicing mediates adaptation to nutrient-limited environments. Interestingly, different AS events leading to similar C-terminally truncated Per2 isoforms have also been found in a second independent cavefish population of *Astyanax maxicanus* as well as in *Phreatichthys andruzzi*, another cave-dwelling fish, raising the possibility of convergent evolution of Per2 in cave adaptation[149].

Interestingly, recent work raises the possibility that AS also contributes to the evolution of sex-specific adaptations — particularly in cases where changes to gene expression patterns are limited by functional constraints[150–153]. Through AS, transcripts that confer a sex-specific benefit can be exclusively expressed in the sex it benefits. Comparison of AS patterns across sexes in multiple bird species revealed hundreds of genes with sex-specific splicing patterns associated with phenotypic sexual dimophism[154]. For example, in wild turkeys, dominant males have sexually selected traits such as distinct plumage, whereas subordinate males have an intermediate phenotype between that of dominant males and females, and also had an intermediate splicing profile[154]. Moreover, a population genomics approach showed that sex-specific splicing patterns were driven by sexual selection and have sex-specific effects. This study indicates that sex-specific AS events could facilitate sex-specific adaptations.

Overall, divergent splicing has been implicated in evolutionary, adaptive changes. However, in many cases, the precise mechanism by which changes in splicing affect a given trait is incompletely understood. Going forwards, dissecting the underlying mechanisms will be necessary to illuminate the most common avenues of phenotypic change, for example through loss of protein function or through altered protein function via switching of functional domains.

Interplay between splicing and transcriptional control. Genome-wide studies suggest that splicing and transcriptional control are largely regulated by different genetic loci[37,155–157]. This implies that the processes can evolve independently, for example by affecting different genes or biological processes[158–160]. Indeed, it has been observed that AS and transcriptional control programmes tend to affect different, complementary sets of genes within a given tissue[74,161]. Studies of closely related populations show that even over short evolutionary timescales, divergence in AS and mRNA abundance already affects different sets of genes. For example, genes that were differentially spliced and differentially expressed between ecotypes of Arctic charr (*Salvelinus alpinus*) were involved in different processes and were largely independent, with less than 6% overlap[160]. It has been suggested that AS may be more evolutionarily flexible than gene expression as it can alter the balance of isoforms of constrained genes, such as pleiotropic genes, while maintaining expression of essential isoforms[160,162].

Studies indicate that splicing diverges faster than gene expression, reinforcing the view that the two processes provide alternative, complementary routes to rapid adaptation. Comparisons of organ transcriptomes from multiple vertebrate species found that AS patterns have rapidly diverged during vertebrate evolution, with splicing variation between species exceeding within-species variation across tissues[27,28]. By contrast, tissue-level gene expression profiles are strongly conserved[163]. However, these patterns are based on evolution across many millions of years. Equivalent studies spanning shorter evolutionary distances have proved more variable. For example, whereas three times more differences in AS than in gene expression were detected between recently radiated cichlid species that diverged around 0.7–2.8 million years ago[158], similar numbers of AS and gene expression differences were found between ecotypes of Arctic charr that diverged 10,000 to 15,000 years ago[160]. Moreover, it should be acknowledged that relative rates of divergence of splicing and gene expression alone cannot be used to ascertain the relative contribution of each process to evolution. Splicing and transcript levels are also inherently interconnected, as splicing is largely co-transcriptional and so is related to transcription levels, and in turn AS can influence transcript stability. Overall, evidence suggests AS that complements other regulatory changes and can play an important part during early stages of divergence and ecological speciation, although more studies, particularly over shorter evolutionary timescales, are needed to dissect their relationships.



Short TRPV1 expressed in 'leaf pits' around the nose

Detection of infrared radiation from heat spots of prey

Fig. 5 | **Divergent alternative splicing can mediate lineage-specific adaptations.** Alternative splicing (AS) permits infrared sensation in vampire bats, *Desmodus rotundus*. **a** | Skipping of exon 14a (blue lines) gives rise to transcripts that encode the full-length TRPV1 isoform, whereas inclusion of exon 14a (red lines), which contains a premature termination codon (PTC), generates transcripts that encode the short isoform of TRPV1. **b** | Transcripts encoding full-length TRPV1 are expressed in the dorsal root ganglia (DRG) of vampire bats and fruit bats. Transcripts encoding the short TRVP1 isoform constitute around 40% of total transcripts in the trigeminal ganglia (TG) of vampire bats, whereas this form is expressed at physiologically non-relevant levels in the TG of fruit bats. **c** | Relative heat response profiles of full-length and truncated TRPV1 isoforms show that truncated TRPV1 has a lower activation threshold. **d** | The lower activation threshold of truncated TRPV1 enables activation in response to temperatures of around 30 °C, which corresponds to infrared radiation from the thermal hotspots of warm-blooded prey. Parts **a** and **c** adapted from REF.[166], Springer Nature Limited. Credit for photograph of *Desmodus rotundus*: Mendesbio/Alamy Stock Photo, and credit for photgraph of cow: Maximilian Buzun/Alamy Stock Photo.

What drives divergence in AS?

Divergence in AS profiles could reflect relaxed selection against AS regulation and thus non-functional transcriptome noise, or lineage-specific innovations. Determining the AS events that are adaptive versus neutral or even mildly deleterious is a key challenge for the field. One approach is through mapping the genetic variants that control AS (splicing QTLs), which can provide insights into their evolutionary history[164]. For example, in sunflowers, several *trans*-splicing QTLs, a large proportion of which affect spliceosomal proteins, were identified that differ between domesticated and wild populations and controlled genes involved in domestication traits, notably seed development[123]. The association of AS with traits subject to artificial selection supports a role for selection in driving rapid AS divergence.

Comparisons between species generally indicate that rapid divergence in AS is primarily driven by genetic drift. For example, conservation of alternative exons seems to be limited in primates and many differences in exons are predicted to have minor effects[165]. Such limited conservation tends to suggest a faster evolution rate associated with weak selective pressure and lack of function. Consistent with weak purifying selection, human genome-wide genetic variation data suggest that most alternative exons have a higher non-synonymous to synonymous substitution ratio than do constitutive exons[99]. One potential explanation may be that alternative exons tend to have lower expression levels and so experience lower effective selection than constitutive exons[117]. It has also been observed that AS rates are higher in genes with lower expression levels and in genes with fewer introns[44]. This suggests that most variation in AS reflects splicing errors due to weak selective pressure as the cost of such errors increases with the amount of resources used in translation. Thus, overall, while it could be that some alternative transcripts may still perform important functional roles under certain environmental conditions, it seems likely that the majority of divergence in AS profiles is driven by drift.

It has been suggested that, by largely evolving under neutral conditions, AS can rapidly evolve and provide a route for existing genes to acquire new functions and thus adaptive benefits[99,117]. In brief, when alternative exons first arise they are likely to be included only in a small proportion of transcripts owing to weak splice sites, and so are unlikely to disrupt normal protein function. Over time, the alternative exon may acquire additional mutations and thus evolve quickly under relaxed selection pressures. In support of this scenario, exons with low inclusion levels undergo increased evolutionary change[117]. This leads to standing variation in alternative transcripts at low expression levels, which are largely non-functional. However, if a particular alternative transcript gains a useful function, for example with a change in environment, selection might then act to increase its expression by strengthening splice sites. One case where this process may have occurred is the heat-sensitive ion channel, TRPV1, in common vampire bats (*Desmodus rotundus*) (FIG. 5). In both vampire bats and the closely related fruit

bats, *Carollia brevicauda*, full-length TRPV1 is expressed in dorsal root ganglia and activated at temperatures higher than 38 °C, allowing detection of noxious heat[166]. Although the *TRPV1* gene contains exon 14a in both species, it is only included at physiologically relevant levels in the trigeminal nerve fibres that innervate specialized pit organs on the face of the vampire bat. Inclusion of exon 14a results in a PTC and truncation of the C-terminal domain of TRPV1, resulting in its activation at lower temperatures (around 30 °C), thereby permitting infrared detection — a key adaptation that allows the bats to locate blood vessels in prey[167]. Therefore, it seems that vampire bats adapted to their bloodthirsty lifestyle by altering the splicing pattern of *TRPV1* specifically in the trigeminal ganglia. Remarkably, the newly acquired exon is dynamically regulated, so the ancestral splicing pattern and thus heat-sensor function of *TRPV1* is retained in the dorsal root ganglia and other cells. In this way, the rapid evolution of alternative exons, which were previously encoded but dormant, can facilitate lineage-specific adaptations.

Recently, it has been hypothesized that NMD could serve as a buffer against the effects of AS variation on phenotype[168,169]. By reducing the expression of highly expressed or otherwise aberrant transcripts, NMD may allow organisms to explore the landscape of AS, while being protected from potential deleterious effects of AS transcripts. The resulting accumulation of concealed genetic variation in AS may potentiate future adaptation, for example when a mutation frees the transcript from NMD-mediated buffering. However, this theory, which has been compared to the theory of evolutionary capacitance, requires experimental evidence and the specific mechanism of buffering requires characterization[168].

Occasionally, changes in splicing can have both adaptive and deleterious effects, leading to trade-offs. For example, a recent preprint paper has attributed the genetic basis of tail loss during the evolution of hominoids to a novel exon-skipping event in transcripts encoding the transcription factor, TBXT[170]. Insertion of an *Alu* element into intron 6 of *TBXT* in the hominoid ancestor results in skipping of exon 6 in about 50% of transcripts. The new *Alu* element is hypothesized to form an inverted repeat pair with a simian-specific *Alu* element in intron 5, forming a stem-loop structure around exon 6, driving exon-skipping. Strikingly, mouse models heterozygous for exon-skipped *TBXT* have a reduced or absent tail. It has been suggested that the selective advantage of tail loss, coincident with the evolution of bipedalism and transition to a non-arboreal lifestyle, was strong, because neural tube defects have also been linked to changes in *TBXT* splicing[170]. Thus, evolutionary trade-offs associated with changes in splicing may continue to have consequences for human health today.

Conclusions and future directions

Since the discovery of introns in 1977, the potential of AS to accelerate evolutionary change has been hypothesized. The biochemist Walter Gilbert theorized the existence of alternative transcripts and that hence proteins could evolve through sequential mutations near splice sites, resulting in multiple transcripts without disrupting gene function, so that "evolution can seek new solutions without destroying the old"[11].

Recent studies support the idea that AS is an important contributor to adaptive evolutionary change, interacting with other forms of genetic variation, such as transcriptional regulation. Much like mutations in the genome more broadly, the majority of splicing variation is probably neutral or mildly deleterious, representing biological noise rather than functional variation. However, this noise might represent useful standing genetic variation that could later be harnessed by selection to produce functional variants. Understanding how drift and selection interact to shape patterns of AS will be fundamental to establishing the role of alternative splicing in driving both phenotypic variation and complexity.

The challenge is now to combine genome-wide approaches with molecular studies to dissect the functional consequences of AS at the proteomic and phenotypic level. With the recent development of long-read RNA-sequencing, combined with growing use of CRISPR–Cas9 genome editing, the tools to do so are now available[171,172]. Moreover, the increasing number of high-quality chromosomal assemblies becoming available for a wide range of species will enable exploration of AS across the tree of life and lead to a better understanding of the extent to which variation in AS is evolutionarily functional[171,172].

As alternative transcripts are often tissue- or cell-type-specific, it is expected that single-cell transcriptomics will provide major insights into the roles and dynamics of AS, particularly in tissues with complex AS, including the brain[173]. Currently, quantitative profiling of AS in single cells is technically and computationally challenging (as highlighted by REF.[174]), although computational methods that account for technical limitations, such as low recovery of mRNAs, are under development[175].

In summary, AS is an effective mechanism with which to shape the evolutionary landscape of multicellular eukaryotes because of its twofold flexibility: both in the form of regulation — many mechanisms can contribute to splicing profiles — and evolution, as a few mutations can have a dramatic effect. We expect that

Glossary

Alternative splicing
The process of selecting different combinations of splice sites within the pre-mRNAs of a gene, resulting in mature mRNA transcripts that differ in composition. These transcripts may result in different protein isoforms with potentially divergent functions.

Cassette exons
Alternative exons that can be included or skipped from the spliced mRNA.

***Cis*-regulatory elements**
Genetic regions that regulate the expression of a coding sequence on the same DNA strand.

Effective population size
The number of individuals in an idealized population that would show the same amount of genetic drift as observed in the real population.

Exon definition
A process in which pairs of splice sites are first recognized, and splicing complexes assembled, across exons.

Introgression
The transfer of genetic material between two species owing to hybridization and subsequent backcrossing with one of the species.

Intron definition
A process in which pairs of splice sites are recognized and then splicing complexes are assembled directly across the intron to be spliced out.

Poison exons
Exons that introduce in-frame premature termination codons, frequently in all three reading frames.

Purifying selection
The removal of deleterious alleles from a population by natural selection.

Quantitative trait loci
(QTL). Regions of the genome whose variation is associated with a particular quantitative trait.

Spliceosome
The large macromolecular complex composed of proteins and small nuclear RNAs that functions as the basal splicing machinery.

***Trans*-regulatory factors**
Cellular factors (RNA or protein) that regulate the expression of a coding sequence on a different DNA strand to which it is encoded.

future studies on AS will shed further light on the many biological processes it can influence and deepen our understanding of transcriptional regulation more broadly.

Charlotte J. Wright[1,2]✉, *Christopher W. J. Smith*[3]✉ *and Chris D. Jiggins*[2]✉

[1]*Tree of Life, Wellcome Sanger Institute, Cambridge, UK.*

[2]*Department of Zoology, University of Cambridge, Cambridge, UK.*

[3]*Department of Biochemistry, University of Cambridge, Cambridge, UK.*

✉*e-mail: cw22@sanger.ac.uk; cwjs1@cam.ac.uk; c.jiggins@zoo.cam.ac.uk*

https://doi.org/10.1038/s41576-022-00514-4

Published online 12 July 2022

1. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
2. Manning, K. S. & Cooper, T. A. The roles of RNA processing in translating genotype to phenotype. *Nat. Rev. Mol. Cell Biol.* **18**, 102–114 (2017).
3. Stern, D. L. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
4. Carroll, S. B. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**, 577–580 (2000).
5. Doebley, J. & Lukens, L. Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**, 1075–1082 (1998).
6. Monlong, J., Calvo, M., Ferreira, P. G. & Guigó, R. Identification of genetic variants associated with alternative splicing using sQTLseekeR. *Nat. Commun.* **5**, 4698 (2014).
7. Yang, Y. et al. SNP2APA: a database for evaluating effects of genetic variants on alternative polyadenylation in human cancers. *Nucleic Acids Res.* **48**, D226–D232 (2020).
8. Belkadi, A. et al. Identification of genetic variants controlling RNA editing and their effect on RNA structure stabilization. *Eur. J. Hum. Genet.* **28**, 1753–1762 (2020).
9. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA* **74**, 3171–3175 (1977).
10. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
11. Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).
12. Gilbert, W. The exon theory of genes. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 901–905 (1987).
13. Rogers, J. et al. Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin mu chain. *Cell* **20**, 303–312 (1980).
14. Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. M. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* **298**, 240–244 (1982).
15. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
16. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
17. Hertel, K. J. *Spliceosomal Pre-mRNA Splicing: Methods and Protocols* (Humana Press, 2014).
18. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **18**, 18–30 (2017).
19. Pal, S., Gupta, R. & Davuluri, R. V. Alternative transcription and alternative splicing in cancer. *Pharmacol. Ther.* **136**, 283–294 (2012).
20. Zhang, C., Yang, H. & Yang, H. Evolutionary character of alternative splicing in plants. *Bioinform. Biol. Insights* **9**, 47–52 (2015).
21. Grützmann, K. et al. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA Res.* **21**, 27–39 (2014).
22. Schwartz, S. H. et al. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* **18**, 88–103 (2008).
23. Gonzalez-Hilarion, S. et al. Intron retention-dependent gene regulation in *Cryptococcus neoformans*. *Sci. Rep.* **6**, 32252 (2016).
24. Kalyna, M. et al. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* **40**, 2454–2469 (2012).
25. Smith, C. W., Patton, J. G. & Nadal-Ginard, B. Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23**, 527–577 (1989).
26. Sharp, P. A. Split genes and RNA splicing (Nobel lecture). *Angew. Chem. Int. Edn Engl.* **33**, 1229–1240 (1994).
27. Barbosa-Morais, N. L. et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
28. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
29. Li, Q. et al. The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. *eLife* **3**, e01201 (2014).
30. Gallagher, T. L. et al. Rbfox-regulated alternative splicing is critical for zebrafish cardiac and skeletal muscle functions. *Dev. Biol.* **359**, 251–261 (2011).
31. Terai, Y., Morikawa, N., Kawakami, K. & Okada, N. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proc. Natl Acad. Sci. USA* **100**, 12798–12803 (2003).
32. Bolisetty, M. T., Rajadinakaran, G. & Graveley, B. R. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* **16**, 204 (2015).
33. Hattori, D. et al. Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* **461**, 644–648 (2009).
34. Bush, S. J., Chen, L., Tovar-Corona, J. M. & Urrutia, A. O. Alternative splicing and the evolution of phenotypic novelty. *Phil. Trans. R. Soc. Lond. B* **372**, 20150474 (2017).
35. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
36. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
37. Lee, Y. & Rio, D. C. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).
38. Fu, X.-D. & Ares, M. Jr Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* **15**, 689–701 (2014).
39. Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94 (1990).
40. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414 (1995).
41. Grau-Bové, X., Ruiz-Trillo, I. & Irimia, M. Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. *Genome Biol.* **19**, 135 (2018).
42. Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A. & Kalyna, M. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res.* **25**, 995–1007 (2015).
43. Sousa-Luís, R. et al. POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol. Cell* **81**, 1935–1950.e6 (2021).
44. Saudemont, B. et al. The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. *Genome Biol.* **18**, 208 (2017).
45. Melamud, E. & Moult, J. Structural implication of splicing stochastics. *Nucleic Acids Res.* **37**, 4862–4872 (2009).
46. Pickrell, J. K., Pai, A. A., Gilad, Y. & Pritchard, J. K. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **6**, e1001236 (2010).
47. Stepankiw, N., Raghavan, M., Fogarty, E. A., Grimson, A. & Pleiss, J. A. Widespread alternative and aberrant splicing revealed by lariat sequencing. *Nucleic Acids Res.* **43**, 8488–8501 (2015).
48. Reixachs-Solé, M. & Eyras, E. Uncovering the impacts of alternative splicing on the proteome with current omics techniques. *Wiley Interdiscip. Rev. RNA* https://doi.org/10.1002/wrna.1707 (2022).
49. Kelemen, O. et al. Function of alternative splicing. *Gene* **514**, 1–30 (2013).
50. McGlincy, N. J. & Smith, C. W. J. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* **33**, 385–393 (2008).
51. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
52. Bhuiyan, S. A. et al. Systematic evaluation of isoform function in literature reports of alternative splicing. *BMC Genom.* **19**, 637 (2018).
53. Stamm, S. et al. Function of alternative splicing. *Gene* **344**, 1–20 (2005).
54. Resch, A. et al. Assessing the impact of alternative splicing on domain interactions in the human proteome. *J. Proteome Res.* **3**, 76–83 (2004).
55. Xing, Y., Xu, Q. & Lee, C. Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS Lett.* **555**, 572–578 (2003).
56. Goeke, S. et al. Alternative splicing of *lola* generates 19 transcription factors controlling axon guidance in *Drosophila*. *Nat. Neurosci.* **6**, 917–924 (2003).
57. Martín, G., Márquez, Y., Mantica, F., Duque, P. & Irimia, M. Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biol.* **22**, 35 (2021).
58. Buljan, M. et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* **46**, 871–883 (2012).
59. Ellis, J. D. et al. Tissue-specific alternative splicing remodels protein–protein interaction networks. *Mol. Cell* **46**, 884–892 (2012).
60. Romero, P. R. et al. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl Acad. Sci. USA* **103**, 8390–8395 (2006).
61. Lam, S. D., Babu, M. M., Lees, J. & Orengo, C. A. Biological impact of mutually exclusive exon switching. *PLoS Comput. Biol.* **17**, e1008708 (2021).
62. Gabut, M. et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146 (2011).
63. Floor, S. N. & Doudna, J. A. Tunable protein synthesis by transcript isoforms in human cells. *eLife* **5**, e10921 (2016).
64. Weatheritt, R. J., Sterne-Weiler, T. & Blencowe, B. J. The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* **23**, 1117–1123 (2016).
65. Mockenhaupt, S. & Makeyev, E. V. Non-coding functions of alternative pre-mRNA splicing in development. *Semin. Cell Dev. Biol.* **47–48**, 32–39 (2015).
66. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
67. Ni, J. Z. et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
68. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
69. da Costa, P. J., Menezes, J. & Romão, L. The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. *Int. J. Biochem. Cell Biol.* **91**, 168–175 (2017).
70. Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. J. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**, 91–100 (2004).
71. Uzor, S. et al. Autoregulation of the human splice factor kinase CLK1 through exon skipping and intron retention. *Gene* **670**, 46–54 (2018).
72. Leclair, N. K. et al. Poison exon splicing regulates a coordinated network of SR protein expression during differentiation and tumorigenesis. *Mol. Cell* **80**, 648–665.e9 (2020).
73. Thomas, J. D. et al. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nat. Genet.* **52**, 84–94 (2020).
74. Pan, Q. et al. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* **20**, 153–158 (2006).
75. Rivas, M. A. et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666–669 (2015).

76. Lindeboom, R. G. H., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48**, 1112–1118 (2016).
77. García-Moreno, J. F. & Romão, L. Perspective in alternative splicing coupled to nonsense-mediated mRNA decay. *Int. J. Mol. Sci.* **21**, 9424 (2020).
78. Zheng, Z.-Z. et al. Alternative splicing regulation of doublesex gene by RNA-binding proteins in the silkworm *Bombyx mori*. *RNA Biol.* **16**, 809–820 (2019).
79. Lagos, D., Koukidou, M., Savakis, C. & Komitopoulou, K. The transformer gene in *Bactrocera oleae*: the genetic switch that determines its sex fate. *Insect Mol. Biol.* **16**, 221–230 (2007).
80. Shukla, J. N. & Palli, S. R. Sex determination in beetles: production of all male progeny by parental RNAi knockdown of transformer. *Sci. Rep.* **2**, 602 (2012).
81. Salz, H. K. Sex determination in insects: a binary decision based on alternative splicing. *Curr. Opin. Genet. Dev.* **21**, 395–400 (2011).
82. Haltenhof, T. et al. A conserved kinase-based body-temperature sensor globally controls alternative splicing and gene expression. *Mol. Cell* **78**, 57–69.e4 (2020).
83. Rhen, T. & Lang, J. W. Among-family variation for environmental sex determination in reptiles. *Evolution* **52**, 1514–1520 (1998).
84. Preußner, M. et al. Body temperature cycles control rhythmic alternative splicing in mammals. *Mol. Cell* **67**, 433–446.e4 (2017).
85. Posé, D. et al. Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* **503**, 414–417 (2013).
86. Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C. & Balasubramanian, S. Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in *Arabidopsis*. *Nat. Plants* **2**, 16055 (2016).
87. Capovilla, G., Symeonidi, E., Wu, R. & Schmid, M. Contribution of major FLM isoforms to temperature-dependent flowering in *Arabidopsis thaliana*. *J. Exp. Bot.* **68**, 5117–5127 (2017).
88. Lutz, U. et al. Natural haplotypes of FLM non-coding sequences fine-tune flowering time in ambient spring temperatures in *Arabidopsis*. *eLife* **6**, e22114 (2017).
89. Terrapon, N. et al. Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* **5**, 3636 (2014).
90. Foret, S. et al. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl Acad. Sci. USA* **109**, 4968–4973 (2012).
91. Bonasio, R. et al. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764 (2012).
92. Price, J. et al. Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*. *Mol. Ecol.* **27**, 1036–1043 (2018).
93. Steward, R. A., de Jong, M. A., Oostra, V. & Wheat, C. W. Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nat. Commun.* **13**, 755 (2022).
94. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
95. Lemos, B., Araripe, L. O., Fontanillas, P. & Hartl, D. L. Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. *Proc. Natl Acad. Sci. USA* **105**, 14471–14476 (2008).
96. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88 (2004).
97. Kondrashov, F. A. & Koonin, E. V. Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. *Trends Genet.* **19**, 115–119 (2003).
98. Ast, G. How did alternative splicing evolve? *Nat. Rev. Genet.* **5**, 773–782 (2004).
99. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* **11**, 345–355 (2010).
100. Hynes, R. O. The evolution of metazoan extracellular matrix. *J. Cell Biol.* **196**, 671–679 (2012).
101. Královičová, J. et al. Restriction of an intron size en route to endothermy. *Nucleic Acids Res.* **49**, 2460–2487 (2021).
102. Waites, G. T. et al. Mutually exclusive splicing of calcium-binding domain exons in chick alpha-actinin. *J. Biol. Chem.* **267**, 6263–6271 (1992).
103. Hatje, K. et al. The landscape of human mutually exclusive splicing. *Mol. Syst. Biol.* **13**, 959 (2017).
104. Graveley, B. R. et al. The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA* **10**, 1499–1506 (2004).
105. Baeza-Centurion, P., Miñana, B., Schmiedel, J. M., Valcárcel, J. & Lehner, B. Combinatorial genetics reveals a scaling law for the effects of mutations on splicing. *Cell* **176**, 549–563.e23 (2019).
106. Glidden, D. T., Buerer, J. L., Saueressig, C. F. & Fairbrother, W. G. Hotspot exons are common targets of splicing perturbations. *Nat. Commun.* **12**, 2756 (2021).
107. Krawczak, M. et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* **28**, 150–158 (2007).
108. Wang, G.-S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* **8**, 749–761 (2007).
109. Schmitz, J. & Brosius, J. Exonization of transposed elements: a challenge and opportunity for evolution. *Biochimie* **93**, 1928–1934 (2011).
110. Attig, J. et al. Heteromeric RNP assembly at LINEs controls lineage-specific RNA processing. *Cell* **174**, 1067–1081.e17 (2018).
111. Sorek, R. et al. Minimal conditions for exonization of intronic sequences: 5′ splice site formation in alu exons. *Mol. Cell* **14**, 221–231 (2004).
112. Attig, J. et al. Splicing repression allows the gradual emergence of new Alu-exons in primate evolution. *eLife* **5**, e19545 (2016).
113. Shen, S. et al. Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl Acad. Sci. USA* **108**, 2837–2842 (2011).
114. Lin, L. et al. Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet.* **4**, e1000225 (2008).
115. Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067 (2002).
116. Alekseyenko, A. V., Kim, N. & Lee, C. J. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* **13**, 661–670 (2007).
117. Modrek, B. & Lee, C. J. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* **34**, 177–180 (2003).
118. Martinez-Gomez, L. et al. Few SINEs of life: Alu elements have little evidence for biological relevance despite elevated translation. *Nar. Genom. Bioinform* **2**, lqz023 (2020).
119. Lin, L. et al. The contribution of Alu exons to the human proteome. *Genome Biol.* **17**, 15 (2016).
120. Merkin, J. J., Chen, P., Alexis, M. S., Hautaniemi, S. K. & Burge, C. B. Origins and impacts of new mammalian exons. *Cell Rep.* **10**, 1992–2005 (2015).
121. Fiszbein, A., Krick, K. S., Begg, B. E. & Burge, C. B. Exon-mediated activation of transcription starts. *Cell* **179**, 1551–1565.e17 (2019).
122. Wong, E. S. et al. Interplay of *cis* and *trans* mechanisms driving transcription factor binding and gene expression evolution. *Nat. Commun.* **8**, 1092 (2017).
123. Smith, C. C. R. et al. Genetics of alternative splicing evolution during sunflower domestication. *Proc. Natl Acad. Sci. USA* **115**, 6768–6773 (2018).
124. Gueroussov, S. et al. An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349**, 868–873 (2015).
125. Liang, Y. et al. SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. *Leukemia* **32**, 2659–2671 (2018).
126. Zarnack, K. et al. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453–466 (2013).
127. Torres-Méndez, A. et al. A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nat. Ecol. Evol.* **3**, 691–701 (2019).
128. Thatcher, S. R. et al. Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell* **26**, 3472–3487 (2014).
129. Jelen, N., Ule, J., Zivin, M. & Darnell, R. B. Evolution of Nova-dependent splicing regulation in the brain. *PLoS Genet.* **3**, 1838–1847 (2007).
130. Li, Y. et al. Global genetic robustness of the alternative splicing machinery in *Caenorhabditis elegans*. *Genetics* **186**, 405–410 (2010).
131. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
132. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
133. Hahn, M. W. & Wray, G. A. The G-value paradox. *Evol. Dev.* **4**, 73–75 (2002).
134. Taft, R. J. & Mattick, J. S. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biol.* **5**, P1 (2003).
135. Peterson, K. J., Dietrich, M. R. & McPeek, M. A. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* **31**, 736–747 (2009).
136. Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A. & Urrutia, A. O. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.* **31**, 1402–1413 (2014).
137. Harr, B. & Turner, L. M. Genome-wide analysis of alternative splicing evolution among *Mus* subspecies. *Mol. Ecol.* **19**, 228–239 (2010).
138. Tovar-Corona, J. M. et al. Alternative splice in alternative lice. *Mol. Biol. Evol.* **32**, 2749–2759 (2015).
139. Olds, B. P. et al. Comparison of the transcriptional profiles of head and body lice. *Insect Mol. Biol.* **21**, 257–268 (2012).
140. Drali, R., Boutellis, A., Raoult, D., Rolain, J. M. & Brouqui, P. Distinguishing body lice from head lice by multiplex real-time PCR analysis of the Phum_PHUM540560 gene. *PLoS ONE* **8**, e58088 (2013).
141. Fujikura, U. et al. Variation in splicing efficiency underlies morphological evolution in *Capsella*. *Dev. Cell* **44**, 192–203.e5 (2018).
142. Jarosch, A., Stolle, E., Crewe, R. M. & Moritz, R. F. A. Alternative splicing of a single transcription factor drives selfish reproductive behavior in honeybee workers (*Apis mellifera*). *Proc. Natl Acad. Sci. USA* **108**, 15282–15287 (2011).
143. Jarosch-Perlow, A., Yusuf, A. A., Pirk, C. W. W., Crewe, R. M. & Moritz, R. F. A. Control of mandibular gland pheromone synthesis by alternative splicing of the CP-2 transcription factor gemini in honeybees (*Apis mellifera carnica*). *Apidologie* **49**, 450–458 (2018).
144. Winston, M. L. & Slessor, K. N. Honey bee primer pheromones and colony organization: gaps in our knowledge. *Apidologie* **29**, 81–95 (1998).
145. Sakagami, S. F. The false-queen: fourth adjustive response in dequeened honeybee colonies. *Behaviour* **13**, 280–295 (1958).
146. Marchinko, K. B. Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution* **63**, 127–138 (2009).
147. Howes, T. R., Summers, B. R. & Kingsley, D. M. Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. *BMC Biol.* **15**, 115 (2017).
148. Xiong, S. et al. Enhanced lipogenesis through Pparγ helps cavefish adapt to food scarcity. *Curr. Biol.* https://doi.org/10.1016/j.cub.2022.03.038 (2022).
149. Ceinos, R. M. et al. Mutations in blind cavefish target the light-regulated circadian clock gene, period 2. *Sci. Rep.* **8**, 8754 (2018).
150. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
151. Brown, J. B. et al. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399 (2014).
152. Gibilisco, L., Zhou, Q., Mahajan, S. & Bachtrog, D. Alternative splicing within and between *Drosophila* species, sexes, tissues, and developmental stages. *PLoS Genet.* **12**, e1006464 (2016).
153. Graveley, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
154. Rogers, T. F., Palmer, D. H. & Wright, A. E. Sex-specific selection drives the evolution of alternative splicing in birds. *Mol. Biol. Evol.* **38**, 519–530 (2021).
155. Li, Y. I. et al. RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).
156. Gao, Q., Sun, W., Ballegeer, M., Libert, C. & Chen, W. Predominant contribution of *cis*-regulatory divergence in the evolution of mouse alternative splicing. *Mol. Syst. Biol.* **11**, 816 (2015).
157. Wang, X. et al. *Cis*-regulated alternative splicing divergence and its potential contribution to environmental responses in *Arabidopsis*. *Plant. J.* **97**, 555–570 (2019).

158. Singh, P., Börger, C., More, H. & Sturmbauer, C. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biol. Evol.* **9**, 2764–2781 (2017).
159. Healy, T. M. & Schulte, P. M. Patterns of alternative splicing in response to cold acclimation in fish. *J. Exp. Biol.* **222**, jeb193516 (2019).
160. Jacobs, A. & Elmer, K. R. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol. Ecol.* https://doi.org/10.1111/mec.15817 (2021).
161. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
162. Papakostas, S. et al. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat. Commun.* **5**, 4071 (2014).
163. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
164. Battle, A. et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* **24**, 14–24 (2014).
165. Reyes, A. et al. Drift and conservation of differential exon usage across tissues in primate species. *Proc. Natl Acad. Sci. USA* **110**, 15377–15382 (2013).
166. Gracheva, E. O. et al. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* **476**, 88–91 (2011).
167. Kürten, L. & Schmidt, U. Thermoperception in the common vampire bat (*Desmodus rotundus*). *J. Comp. Physiol. A* **146**, 223–228 (1982).
168. Verta, J.-P. & Jacobs, A. The role of alternative splicing in adaptation and evolution. *Trends Ecol. Evol.* https://doi.org/10.1016/j.tree.2021.11.010 (2021).
169. Gehring, N. H. & Roignant, J.-Y. Anything but ordinary — emerging splicing mechanisms in eukaryotic gene regulation. *Trends Genet.* **37**, 355–372 (2021).
170. Xia, B. et al. The genetic basis of tail-loss evolution in humans and apes. Preprint at *bioRxiv* https://doi.org/10.1101/2021.09.14.460388 (2021).
171. Rhie, A. et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
172. Lewin, H. A. et al. Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl Acad. Sci. USA* **115**, 4325–4333 (2018).
173. Arzalluz-Luque, Á. & Conesa, A. Single-cell RNAseq for the study of isoforms — how is that possible? *Genome Biol.* **19**, 110 (2018).
174. Buen Abad Najar, C. F., Yosef, N. & Lareau, L. F. Coverage-dependent bias creates the appearance of binary splicing in single cells. *eLife* **9**, e54603 (2020).
175. Buen Abad Najar, C., Burra, P., Yosef, N. & Lareau, L. F. Identifying cell-state associated alternative splicing events and their co-regulation. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.23.453605 (2021).
176. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
177. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
178. Cline, T. W. Two closely linked mutations in *Drosophila melanogaster* that are lethal to opposite sexes and interact with daughterless. *Genetics* **90**, 683–698 (1978).
179. Inoue, K., Hoshijima, K., Sakamoto, H. & Shimura, Y. Binding of the *Drosophila* sex-lethal gene product to the alternative splice site of transformer primary transcript. *Nature* **344**, 461–463 (1990).
180. Belote, J. M., McKeown, M., Boggs, R. T., Ohkawa, R. & Sosnowski, B. A. Molecular genetics of transformer, a genetic switch controlling sexual differentiation in *Drosophila*. *Dev. Genet.* **10**, 143–154 (1989).
181. Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. & McKeown, M. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the transformer gene. *Cell* **50**, 739–747 (1987).
182. Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H. & Shimura, Y. Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. *Science* **252**, 833–836 (1991).
183. Bashaw, G. J. & Baker, B. S. The regulation of the *Drosophila msl-2* gene reveals a function for sex-lethal in translational control. *Cell* **89**, 789–798 (1997).
184. Kim, E., Goren, A. & Ast, G. Alternative splicing: current perspectives. *Bioessays* **30**, 38–47 (2008).
185. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).
186. Pandit, S. et al. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell* **50**, 223–235 (2013).
187. Buckanovich, R. J., Yang, Y. Y. & Darnell, R. B. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* **16**, 1114–1122 (1996).
188. Jensen, K. B. et al. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).
189. Yang, Y. Y., Yin, G. L. & Darnell, R. B. The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl Acad. Sci. USA* **95**, 13254–13259 (1998).
190. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol. Cell. Biol.* **17**, 3194–3201 (1997).
191. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K. & Darnell, R. B. The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl Acad. Sci. USA* **97**, 5740–5745 (2000).
192. Warf, M. B. & Berglund, J. A. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* **35**, 169–178 (2010).
193. Graveley, B. R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65–73 (2005).
194. Dominissini, D. et al. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485**, 201–206 (2012).
195. Lewis, C. J. T., Pan, T. & Kalsotra, A. RNA modifications and structures cooperate to guide RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* **18**, 202–210 (2017).
196. Sultan, M. et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
197. Tress, M. L., Abascal, F. & Valencia, A. Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* **42**, 98–110 (2017).
198. Blencowe, B. J. The relationship between alternative splicing and proteomic complexity. *Trends Biochem. Sci.* **42**, 407–408 (2017).
199. Ule, J. & Blencowe, B. J. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* **76**, 329–345 (2019).
200. Reixachs-Solé, M., Ruiz-Orera, J., Albà, M. M. & Eyras, E. Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nat. Commun.* **11**, 1768 (2020).
201. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013).
202. Melamud, E. & Moult, J. Stochastic noise in splicing machinery. *Nucleic Acids Res.* **37**, 4873–4886 (2009).
203. Xu, C. & Zhang, J. Alternative polyadenylation of mammalian transcripts is generally deleterious, not. adaptive. *Cell Syst.* **6**, 734–742.e4 (2018).
204. Xu, G. & Zhang, J. Human coding RNA editing is generally nonadaptive. *Proc. Natl Acad. Sci. USA* **111**, 3769–3774 (2014).
205. Xu, C., Park, J.-K. & Zhang, J. Evidence that alternative transcriptional initiation is largely nonadaptive. *PLoS Biol.* **17**, e3000197 (2019).
206. Rodriguez, J. M., Pozo, F., di Domenico, T., Vazquez, J. & Tress, M. L. An analysis of tissue-specific alternative splicing at the protein level. *PLoS Comput. Biol.* **16**, e1008287 (2020).
207. Neverov, A. D. et al. Alternative splicing and protein function. *BMC Bioinform.* **6**, 266 (2005).
208. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).
209. Pozo, F. et al. Assessing the functional relevance of splice isoforms. *NAR Genom. Bioinform* **3**, lqab044 (2021).
210. Yuan, J. et al. Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell* **72**, 380–394.e7 (2018).
211. Charton, K. et al. Exploiting the CRISPR/Cas9 system to study alternative splicing in vivo: application to titin. *Hum. Mol. Genet.* **25**, 4518–4532 (2016).
212. Marti-Solano, M. et al. Combinatorial expression of GPCR isoforms affects signalling and drug responses. *Nature* **587**, 650–656 (2020).

Acknowledgements
The authors thank everyone at the Insect Evolution and Genomics Group at the University of Cambridge for useful discussion and comments on the manuscript, and G. Simpson and J. Valcárcel for helpful discussions. This research was funded in part by the Wellcome Trust (grants 206194 and 218328). C.D.J. was supported by a European Research Council (grant 339873 Speciation Genetics) and the UK Biotechnology and Biological Sciences Research Council (grant BB/R007500/1). C.W.J.S. received funding from a Wellcome Trust Investigator award (209368/Z/17/Z). We thank A. Jacob for generating the consensus sequences in Fig. 1c.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Genetics thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.