21 Caillat-Zucman, S. *et al.* (1992) Age-dependent HLA genetic heterogeneity of type 1 insulin-dependent diabetes mellitus. *J. Clin. Invest.* 90, 2242–2250
22 Cucca, F. *et al.* (1995) The distribution of DR4 haplotypes in Sardinia suggests a primary association of type 1 diabetes with DRB1 and DQB1 loci. *Hum. Immunol.* 43, 301–308
23 Undlien, D.E. *et al.* (1997) HLA encoded genetic predisposition in insulin-dependent diabetes mellitus (IDDM). DR4 subtypes may be associated with different degrees of protection. *Diabetes* 46, 143–149
24 Undlien, D.E. *et al.* (1999) HLA associations in type 1 diabetes among patients not carrying high-risk DR3-DQ2 or DR4-DQ8 haplotypes. *Tissue Antigens* 54, 543–551
25 Abraham, R.S. *et al.* (2000) Co-expression of HLA DR3 and DQ8 results in the development of spontaneous insulitis and loss of tolerance to GAD65 in transgenic mice. *Diabetes* 49, 548–554
26 Devlin, B. and Risch, N. (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29, 311–322
27 Caillat-Zucman, S. *et al.* (1995) Family study of linkage disequilibrium between TAP2 transporter and HLA class II genes. Absence of TAP2 contribution to association with insulin-dependent diabetes mellitus. *Hum. Immunol.* 44, 80–87
28 Pritchard, J.K. *et al.* (2000) Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181
29 Robinson, W.P. *et al.* (1993) Homozygous parent affected sib pair method for detecting disease predisposing variants: application to insulin dependent diabetes mellitus. *Genet. Epidemiol.* 10, 273–288
30 Lie, B.A. *et al.* (1999) The predisposition to type 1 diabetes linked to the HLA complex (*IDDM1*) includes at least one non-class II gene. *Am. J. Hum. Genet.* 64, 793–800
31 Spielman, R.S. *et al.* (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52, 506–516
32 Koeleman, B.P. *et al.* (2000) Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the conditional extended transmission/disequilibrium test. *Ann. Hum. Genet.* 64, 207–213
33 Herr, M. *et al.* (2000) Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of IDDM1 in the HLA region on chromosome 6p21. *Hum. Mol. Genet.* 9, 1291–1301
34 Nejentsev, S. *et al.* (2000) The effect of HLA-B allele on the IDDM risk defined by DRB1*04 subtypes and DQB1*0302. *Diabetes* 46, 1888–1892
35 Moghaddam, P.H. *et al.* (1997) TNFa microsatellite polymorphism modulates the risk of IDDM in Caucasians with the high-risk genotype HLA DQA1*0501-DQB1*0201/DQA1*0301-DQB1*0302. Belgian Diabetes Registry. *Diabetes* 46, 1514–1515
36 Lie, B.A. *et al.* (1997) HLA associations in insulin-dependent diabetes mellitus: No independent association to DP genes. *Hum. Immunol.* 55, 170–175
37 Noble, J.A. *et al.* (2000) The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. *Diabetes* 49, 121–125
38 Undlien, D.E. *et al.* (1997) No independent association of LMP2 and LMP7 polymorphisms with susceptibility to IDDM. *Diabetes* 46, 307–312
39 Hattori, M. *et al.* (1999) Cutting edge: homologous recombination of the MHC class I K region defines new MHC-linked diabetogenic susceptibility gene(s) in nonobese diabetic mice. *J. Immunol.* 163, 1721–1724
40 Mathews, C.E. *et al.* (2000) Reevaluation of the major histocompatibility complex genes of the NOD-progenitor CTS/Shi strain. *Diabetes* 49, 131–134

Alternative splicing: increasing diversity in the proteomic world

Brenton R. Graveley

How can the genome of *Drosophila melanogaster* contain fewer genes than the undoubtedly simpler organism *Caenorhabditis elegans*? The answer must lie within their proteomes. It is becoming clear that alternative splicing has an extremely important role in expanding protein diversity and might therefore partially underlie the apparent discrepancy between gene number and organismal complexity. Alternative splicing can generate more transcripts from a single gene than the number of genes in an entire genome. However, for the vast majority of alternative splicing events, the functional significance is unknown. Developing a full catalog of alternatively spliced transcripts and determining each of their functions will be a major challenge of the upcoming proteomic era.

B.R. Graveley
Dept of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030, USA.
e-mail: graveley@neuron.uchc.edu

A major undertaking of the post-genomic era will be the description and functional characterization of the full complement of proteins (i.e. the proteome) expressed by an organism. DNA recombination, RNA editing and alternative splicing make this task more difficult than it first appears as these processes increase the number of proteins that can be synthesized from each gene. As a result, the number of proteins in the proteome is by no means equivalent to the number of genes, but can exceed it by literally orders of magnitude. The mechanism most widely used to enhance protein diversity, with regard to the number of genes affected and the breadth of organisms it occurs in, is alternative splicing (Box 1). Alternative splicing can generate multiple transcripts encoding proteins with subtle or opposing functional differences that can have profound biological consequences. In this article, I will discuss the prevalence of alternative splicing and the amount of diversity it can create, provide examples of several complex alternative splicing events and, finally, discuss the issue of how much of the observed alternative splicing actually represents errors or 'noise' in the system. Readers should refer to an excellent recent review by Smith and Valcárcel[1] for more information about the biochemical mechanisms of alternative splicing.

It's all around us

All eukaryotes contain INTRONS (see Glossary) in at least some of their genes although the number can vary considerably from organism to organism. For example, only 250 out of the 6000 genes in *Saccharomyces cerevisiae* contain introns[2], whereas most of the estimated 35 000 human genes[3,4] are thought to contain introns. But how many of these genes encode transcripts that are alternatively

  PII: S0168-9525(00)02176-4

Box 1. Types of alternative splicing

Every conceivable pattern of alternative splicing is found in Nature. Exons can have multiple 5¢- or 3¢-splice sites that are alternatively used (Fig. Ia,b). Cassette exons are fully contained exons that are alternatively used. Single cassette exons can reside between two constitutive exons such that the alternative exon is either included or skipped (Fig. Ic). Alternatively, multiple cassette exons can reside between two constitutive exons such that the splicing machinery must choose between them (Fig. Id). Finally, introns can be retained in the mRNA and become translated (Fig. Ie). The constitutive exons are depicted as open boxes and alternative exons are shaded. The lines above and below the boxes show possible alternative splicing events.

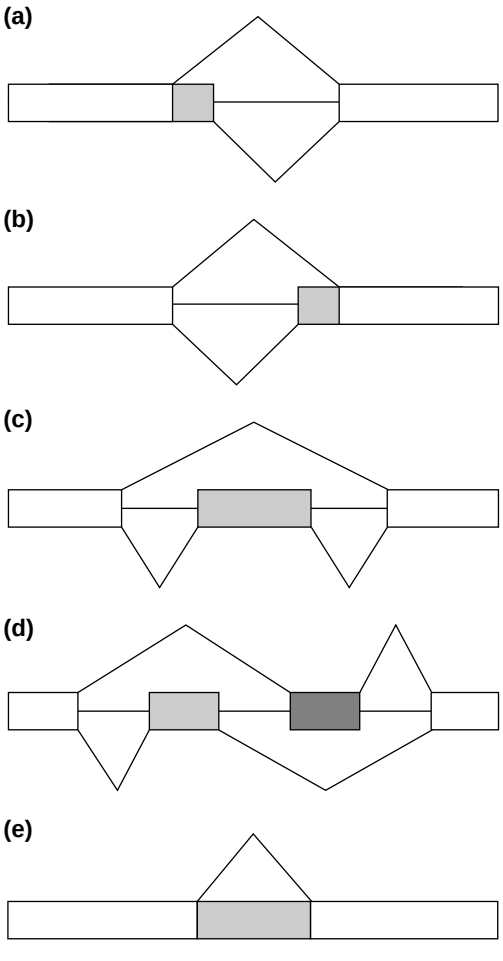


Fig. I TRENDS in Genetics

Glossary

Exon: A portion of a pre-mRNA that is retained in a fully processed mRNA. Exons can either be constitutive (included in all mRNAs) or alternative (included in only some mRNAs).
Expressed sequence tag (EST): Short sequences (100–500 nucleotides) typically from one end of a cloned cDNA. ESTs are useful in gene identification, determining the tissue distribution of transcripts and identification of alternatively spliced transcripts.
Intron: An RNA segment that is excised from a pre-mRNA by the spliceosome.
messenger RNA (mRNA): A mature transcript typically generated by post-transcriptional processing of a larger precursor molecule, the pre-mRNA. mRNAs are typically exported from the nucleus to the cytoplasm where they are translated.
Pre-mRNA: A primary transcript that is a precursor to an mRNA.
Ribosome: A macromolecular machine that facilitates mRNA translation. Like the spliceosome, the ribosome is composed of both protein and RNA.
RNA editing: A post-transcriptional RNA processing event in which specific nucleotides are chemically modified. For example, the conversion of adenosine to inosine is catalyzed by a family of enzymes called adenosine deaminases. This type of RNA processing event can alter the identity of a single amino acid in the encoded protein.
RNA-mediated interference (RNAi): The phenomenon whereby double-stranded RNA identical to exonic sequences of a given gene acts post-transcriptionally to interfere with gene expression. This is mediated by the specific degradation of transcripts containing sequences identical to the double-stranded RNA.
Spliceosome: The macromolecular machine that catalyzes the splicing reaction. The spliceosome consists of five small nuclear RNAs and between 50 and 100 proteins. Throughout this review, the terms ‘spliceosome’ and ‘splicing machinery’ are used interchangeably.
Splice sites: The junctions between exons and introns at which the chemical reactions that mediate intron removal occur. The 5¢ splice site is the junction between the 5¢ end of the intron and the 3¢ end of the exon. The 3¢ splice site is the junction between the 3¢ end of the intron and the 5¢ end of the exon.

spliced? Only three *S. cerevisiae* genes are currently known to encode alternatively spliced transcripts[5,6]. The fact that all of these were identified experimentally and not predicted by *in silico* methods, four years after completion of the yeast genome sequence, underscores the tremendous difficulty of predicting whether genes encode alternatively spliced transcripts.

Recent estimates, based on analyses of EXPRESSED SEQUENCE TAGS (ESTs), suggest that the transcripts from 35% of human genes are alternatively spliced[7,8]. However, this number is likely to be an

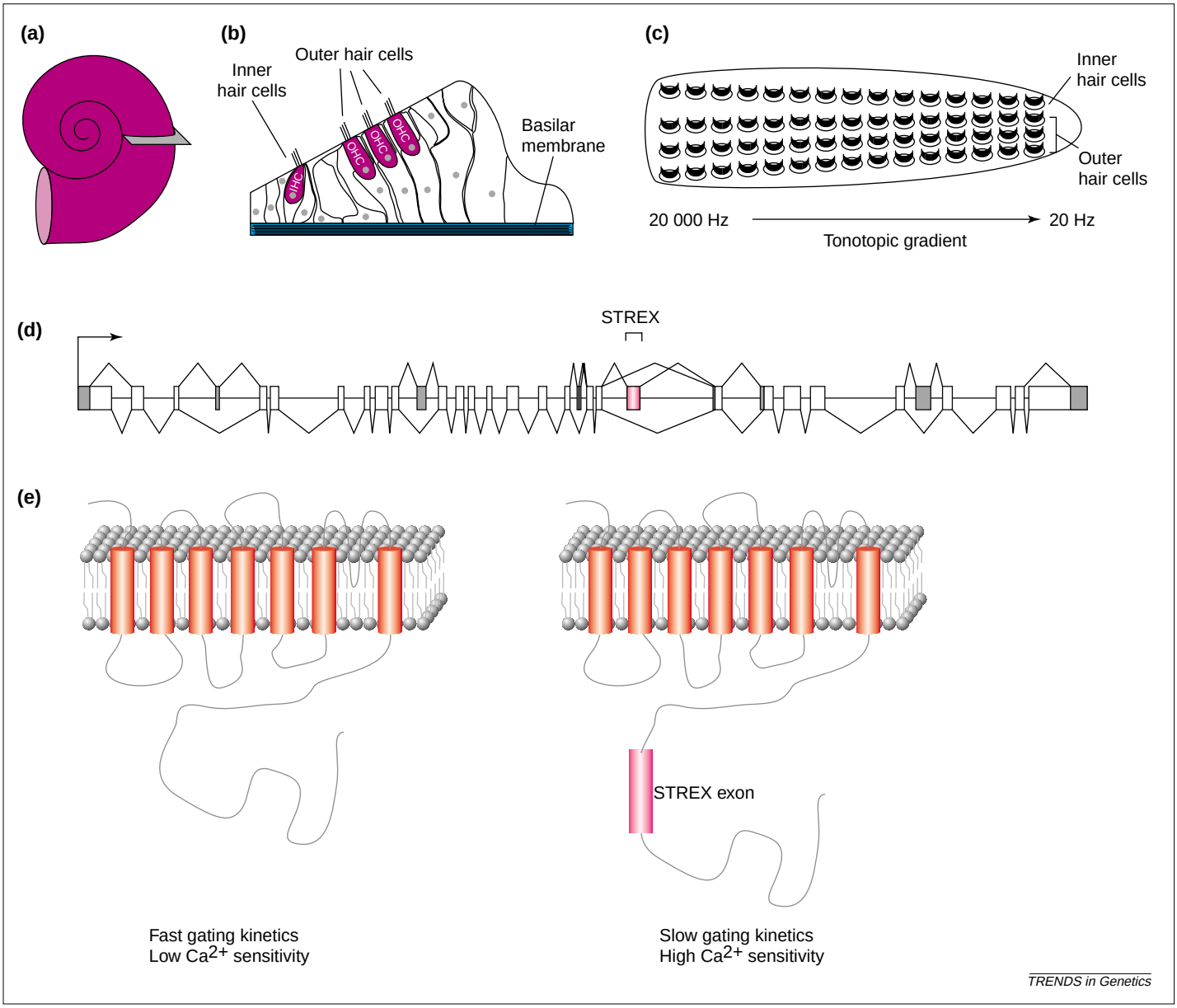


Fig. 1. Alternative splicing of the *slo* gene. (a) The mammalian cochlea. The cochlea is a snail-like structure of the inner ear that contains hair cells organized along a basilar membrane. The basilar membrane traverses the length of the curled-up cochlea. (b) The cochlea is sliced transversely as shown in (a) and the section of the cochlea containing the basilar membrane and the hair cells depicted. There are four rows of hair cells, one inner hair cell and three outer hair cells, situated above the basilar membrane. (c) The cochlea is unrolled to reveal the basilar membrane viewed from above. The four hair cells are arranged in rows along the length of the basilar membrane. The hair cells are tuned to unique narrow sound frequencies along the basilar membrane creating a tonotopic gradient. At one end of the membrane, hair cells are tuned to respond to a frequency of 20 Hz, where as hair cells at the other end respond to 20 000 Hz. (d) Organization of the human *slo* gene. The exon–intron organization of the *slo* gene (determined by an analysis of draft sequence of the human genome) is depicted. The constitutive splicing events are indicated below the gene and alternative splicing events are depicted above the gene. The constitutive exons are white and the alternative exons are shaded. The STREX exon is purple. (e) Isoforms of the Slo protein lacking sequences encoded by the STREX exon have fast deactivation kinetics and low Ca^{2+} sensitivity, where as isoforms containing STREX-encoded sequences have slower deactivation kinetics and higher Ca^{2+} sensitivity.

underestimate. First, the human EST collection (http://www.ncbi.nlm.nih.gov/dbEST) does not represent all protein coding sequences and for most genes, ESTs only cover a portion of the transcript. For example, a recent study found that only about half of the annotated genes on human chromosome 22 are represented in the current EST database[9]. Moreover, 26% and 65% of the ESTs correspond to the extreme 5¢ or 3¢ end, respectively, of any given transcript[10]. Because much of the functionally significant alternative splicing will occur in the coding region of the transcript, many alternative splicing events might be overlooked by EST comparisons.

Second, and perhaps more importantly, many alternative splicing events are very rare and occur only in a specific tissue at a specific time in development and/or under certain physiological conditions. These types of splicing events will probably not be well represented in EST collections. For these reasons, the current EST collection might not provide an accurate estimate of the number of alternatively spliced genes and as a result, the number of alternatively spliced genes will undoubtedly be inaccurate for many years. Taking these issues into consideration, the prediction that 35% of human genes

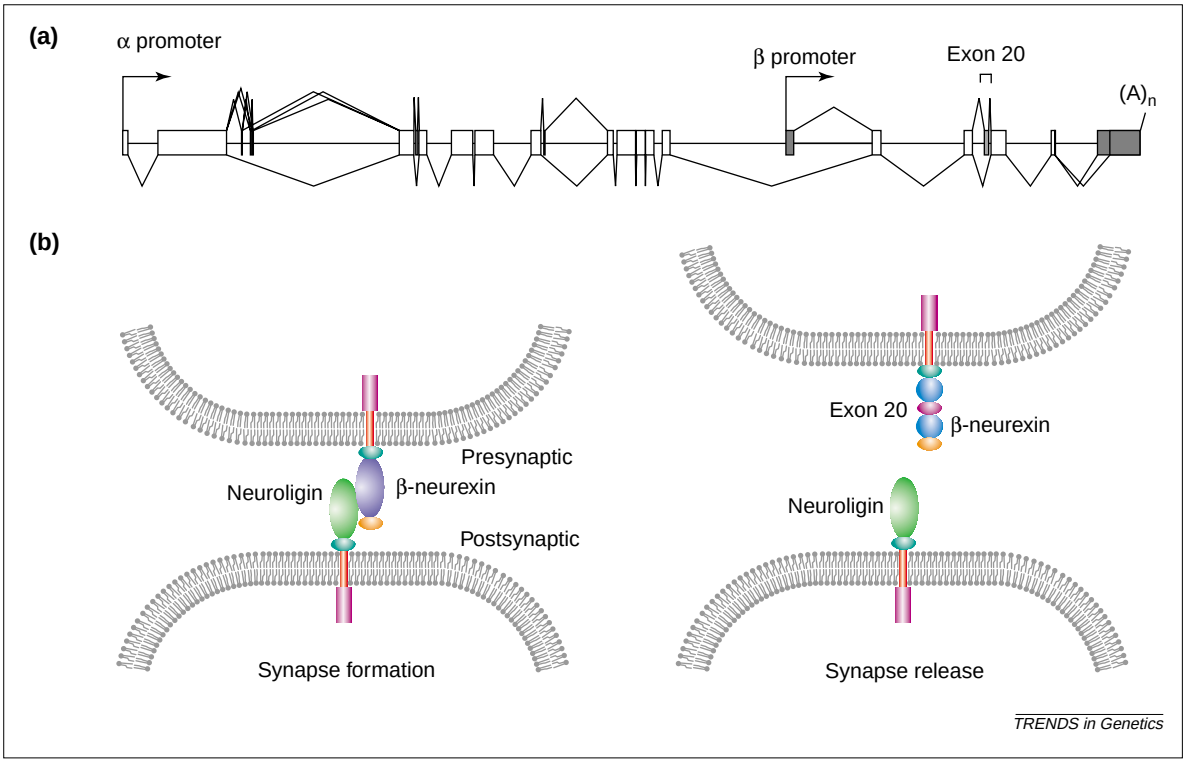


Fig. 2. Alternative splicing of the neurexin genes. (a) Organization of the human gene encoding neurexin I. The exon–intron structure of the human neurexin I gene is depicted (L. Rowen and B. Graveley, unpublished). The constitutive splicing events are indicated below the gene and alternative splicing events are depicted above the gene. The constitutive exons are white and the alternative exons are shaded. Exon 20 is indicated. Human neurexins II and III have a very similar exon-intron organization (L. Rowen and B. Graveley, unpublished). (b) Model for the function of the alternative splicing of exon 20 in b-neurexin I. b-neurexin I (present in the presynaptic cell) lacking sequences encoded by exon 20 can interact with neuroligin present in the postsynaptic cell, and thus function to initiate synaptogenesis. In contrast, b-neurexin I containing exon 20 encoded sequences can not interact with neuroligins. This form of b-neurexin I might indirectly function in releasing synapses.

encode alternatively spliced transcripts is likely to be a significant underestimate[8,11], and we might, in fact, find that the majority of transcripts expressed from human genes are alternatively spliced.

Diversity, diversity, diversity

Another important question is how many alternatively spliced transcripts can be synthesized from a single gene? If each gene can produce two mRNA isoforms, the proteome would be twice as complex as the genome. However, the complexity of the proteome would be significantly higher if alternative splicing can create a much greater extent of transcript diversity. It turns out that several genes do, in fact, encode transcripts that are alternatively spliced to produce not only hundreds or thousands, but even tens of thousands of different mRNAs. Below I will describe three genes, in order of increasing complexity, for which alternative splicing is used to produce extraordinary molecular diversity.

Tuning hair cells

To experience the sounds of a symphony fully, our ears must have the ability to detect a remarkable range of frequencies. Within the cochlea of the inner ear individual hair cells are arranged in four rows along the length of the basilar membrane (Fig. 1a–c). Each of these hair cells is tuned to respond to a unique and narrow range of sound frequencies in such a way that there is a tonotopic gradient across the length of the basilar membrane (Fig. 1c). The collective response of activated hair cells is what we interpret as sound. Hair cell tuning is thought to be mediated, in part, by the alternative splicing of transcripts expressed from the calcium-activated potassium channel gene, *slo*. This leads to the production of multiple protein isoforms that have subtle functional differences, translating into profound biological consequences[12].

There are at least eight sites of alternative splicing in the *slo* transcript that are conserved among vertebrate species[13,14] (Fig. 1d). At most of these sites there is a single alternative exon that can either be included or excluded, or a constitutive exon containing alternative 5¢ or 3¢ SPLICE SITES. Because each alternative splicing event appears to be controlled independently, it is estimated that >500 different *slo* mRNAs can be synthesized[12]. Importantly, this alternative splicing is functionally relevant. Some of the Slo protein isoforms have been shown to vary in their physiological properties. For example, Slo isoforms encoded by mRNAs containing the STREX EXON (stress axis-regulated exon) have slower deactivation kinetics and higher Ca^{2+} sensitivity than isoforms expressed from mRNAs lacking this exon[15] (Fig. 1e). Unfortunately, very little is known about how the alternative splicing of *slo* transcripts is regulated.

Neuronal signaling and synaptogenesis

The neurexins are a family of neural proteins present in vertebrates that have important functions as receptors for neuropeptides[16] and as adhesion molecules that participate in synaptogenesis[17]. The isolation of multiple cDNAs from rat and cow revealed that over 1000 different neurexin mRNAs could potentially be synthesized from three genes by virtue of alternative promoter usage and alternative splicing[18] (Fig. 2a).

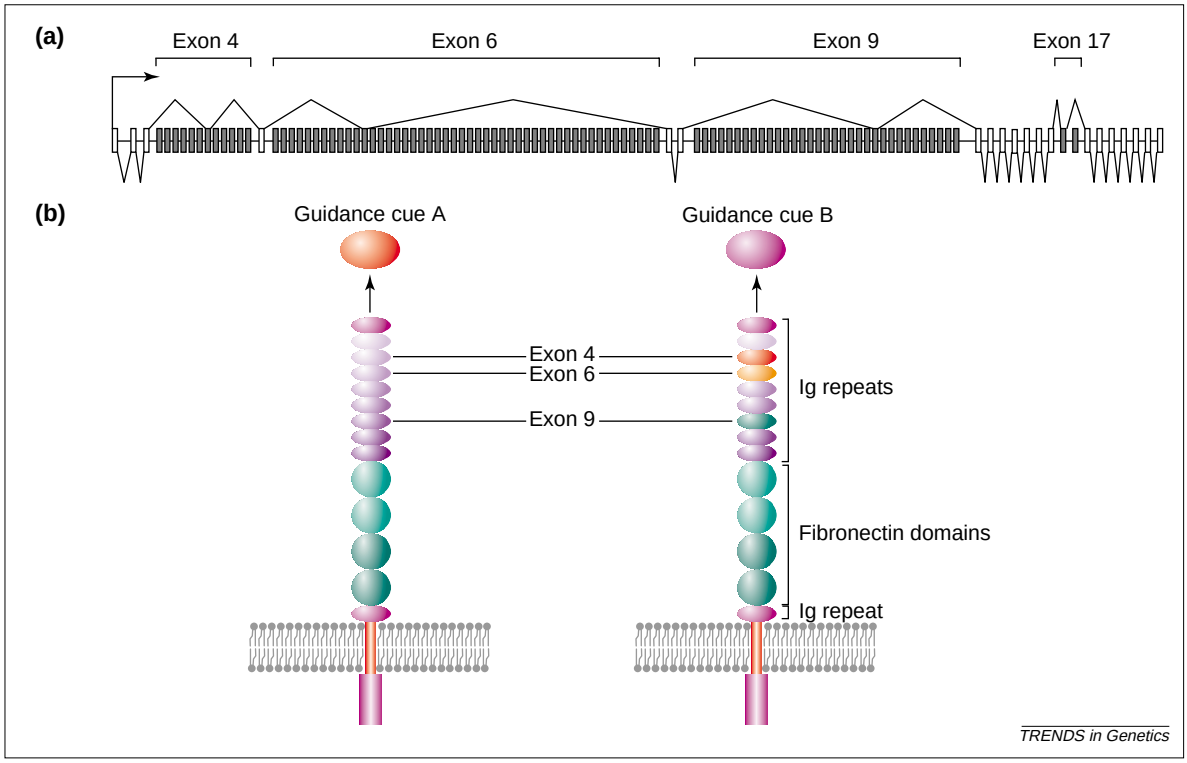


Fig. 3. Alternative splicing of the gene encoding *Drosophila Dscam*. (a) The organization of the *Dscam* gene. The constitutive splicing events are indicated below the gene and alternative splicing events are depicted above the gene. The constitutive exons are white and the alternative exons are shaded. The *Dscam* gene contains four sites of alternative splicing at exons 4, 6, 9 and 17. There are 12 variants of exon 4, 48 variants of exon 6, 33 variants of exon 9 and 2 variants of exon 17. Only one variant exon from each position is included in the *Dscam* mRNAs. Alternative exons 4, 6, and 9 encode alternative versions of immunoglobulin repeats. (b) Functional consequences of *Dscam* alternative splicing. The Dscam protein functions as an axon guidance receptor. It is thought that each Dscam variant will interact with a unique set of axon guidance cues. The form of Dscam shown on the left will interact with guidance cue A. The form of Dscam shown on the right contains different sequences encoded by exons 4, 6 and 9 and thus interacts with guidance cue B, rather than guidance cue A. Neurons expressing the form of Dscam shown on the right will be attracted in a different direction than neurons expressing the form shown on the left.

The proteins encoded by some of these alternatively spliced mRNAs have altered specificities for their ligands[19,20], and it has been proposed that the diversity of neurexin proteins might be a component of a code that specifies neural connectivity[21]. For example, it is now known that the interaction of b-neurexins present in pre-synaptic cells with neuroligins on the surface of post-synaptic cells is sufficient to trigger synapse formation[17]. Importantly, this interaction only occurs if the b-neurexin is encoded by an mRNA lacking an alternative exon (exon 20) at splice site 4 – proteins synthesized from exon 20-containing transcripts do not interact with neuroligins[19]. This is consistent with the observation that, in tissue culture, excess recombinant b-neurexin I produced from a cDNA lacking exon 20 can inhibit synaptogenesis, whereas b-neurexin I synthesized from a cDNA containing exon 20 does not[17]. Although the precise function of b-neurexins containing exon 20-encoded sequences is not known, it is tempting to speculate that the relative ratio of the two protein isoforms could determine whether functional synapses are formed or broken (Fig. 2b). Thus, alternative splicing of the neurexin transcripts might have a direct role in controlling the formation and maintenance of synapses. Elucidating how the splicing of neurexin pre-mRNAs is regulated will undoubtedly provide significant insight into the function of the vertebrate nervous system.

Axon guidance

Perhaps the most complex event that takes place during development is the migration and connection of neurons. Even in a 'simple' organism such as *Drosophila melanogaster*, which contains only ~250 000 neurons, accurately wiring neurons together would appear to be a daunting task. This process must involve a remarkable system to accurately and reproducibly direct growing axons to their destination. In flies, the gene encoding the Down syndrome cell adhesion molecule (*Dscam*) appears to fulfill at least part of this role[22]. *Dscam* encodes an axon guidance receptor with an extracellular domain that contains ten immunoglobulin (Ig) repeats. The most striking feature of the *Dscam* gene is that it's pre-mRNA can be alternatively spliced into over 38 000 different mRNA isoforms (Fig. 3a). This is 2–3 times the number of predicted genes in the entire organism. Each mRNA encodes a distinct receptor with the potential ability to interact with different molecular guidance cues, directing the growing axon to its proper location (Fig. 3b).

This gene has a very unusual exon–intron organization, the like of which has not been seen before[22]. As shown in Fig. 3, the majority of the alternatively spliced exons encode variants of Ig repeats 2, 3 and 7, which are encoded by the alternative cassette exons 4, 6 and 9, respectively. There are 12 variants of exon 4, 48 variants of exon 6 and 33 variants of exon 9. Each mature *Dscam* mRNA contains only one variant of exon 4, 6 and 9. In addition, there are two variants of exon 17, of which only one is included in *Dscam* mRNAs. The synthesis of *Dscam* transcripts has many parallels with today's popular lottery drawings. Imagine there are four containers each containing multiple numbered balls – 12, 48, 33 and 2 in this case. To arrive at the winning number (transcript), one ball from each container is

selected. Is this how *Dscam* transcripts are spliced together, or is the splicing somehow regulated? Although nothing is known about this, there must be some degree of control because *Dscam* transcripts always contain one variant exon at each position, yet never more than one variant[22]. Because each of the variant exons contains both a 5¢ and 3¢ splice site, it is possible that each of the alternative exons could be joined together in a long string. Thus, we must understand not only the mechanism by which one variant is selected at each position, but also why the other variants are not used. Understanding how the complex splicing of the transcripts encoded by this gene is regulated will help to elucidate how the specificity of axon guidance is established.

Even more diversity

These three genes are by no means the only examples of complex alternative splicing events, but they illustrate how this process can significantly expand the complexity of the proteome. This issue becomes even more complex when other RNA processing events, such as RNA EDITING, are considered. For example, the *paralytic* (*para*) gene, which encodes the major voltage-gated action potential sodium channel in *Drosophila*, contains 13 alternative exons[23,24] (R. Reenan, pers. commun.) and can potentially synthesize 1536 different mRNAs. However, *para* transcripts also undergo post-transcriptional RNA editing at 11 known positions[25]. Considering both RNA editing and alternative splicing, 1 032 192 different *para* transcripts can theoretically be synthesized from this single gene. Importantly, the majority of both the alternative splicing and RNA editing sites are likely to be functionally relevant as they are conserved between distantly related *Drosophila* species[23,26]. Of course, it is not known whether each of the possible variants is actually expressed: answering this question will require heroic efforts. The tremendous diversity of transcripts expressed from the *slo*, neurexin, *Dscam* and *para* genes illustrates the magnitude of the task faced in determining an organism's proteome.

Is all of this for real?

The previous examples illustrate how alternative splicing can generate an amazingly diverse collection of proteins from a single gene. But how much alternative splicing is real, and how much of it is actually 'noise' in the system? Here I define noise as either a mistake made by the splicing machinery or an alternative exon that has no impact whatsoever on the function of the encoded protein. Below I will address both of these definitions of noise and their relative contributions to the large amount of alternative splicing that occurs.

Don't blame the spliceosome

Most of the documented splicing errors are caused by mutations in the genomic DNA that either destroy a normal splicing signal or create a new one[27]. Other instances of 'mis-splicing' are caused by mutations in splicing regulatory proteins or their binding sites[28–30]. Although it has not been documented, it is also possible that splicing 'errors' could result from mutations introduced in the pre-mRNA during transcription that create or destroy splice sites. In all of these instances, the SPLICEOSOME itself is not at fault – it is simply following the instructions present in the pre-mRNA. In fact, there is not a single report of a splicing error in which the 'blame' can be placed clearly upon the spliceosome.

Like other macromolecular machines, the spliceosome is probably not perfect and will occasionally make mistakes. The RIBOSOME, for instance, makes an error once every 10 000 codons[31]. If we assume that the fidelity of splicing is similar to that of translation and that an average human gene contains ten introns, then one in every 1000 transcripts would be inappropriately spliced. Of course, these numbers are quite speculative and rigorous systems for the quantitative analysis of splicing accuracy need to be developed. Regardless, the take home message is that true splicing errors probably occur at a very low frequency and thus constitute a very small fraction of entries in the EST database.

There are two possible fates for inappropriately spliced transcripts – translation or degradation. On the one hand, these transcripts can be exported from the nucleus and translated, resulting in the production of aberrant proteins. In many cases, such aberrant proteins would not be functional, and would not affect the cell negatively. However, it is possible that some of the aberrant proteins synthesized from mis-spliced mRNAs would have dominant functions – a situation that could rapidly turn catastrophic. Thus, the translation of aberrantly spliced transcripts must occur at an extremely low frequency.

On the other hand, aberrantly spliced transcripts might not be translated at all, but rather degraded to prevent the synthesis of potentially harmful proteins. Eukaryotes possess an mRNA surveillance system that scans newly synthesized mRNAs for the presence of premature stop codons and, if detected, degrades the defective mRNAs to prevent their translation[32]. A number of different splicing errors could lead to the insertion of premature stop codons in a transcript. For example, improperly including exons or retaining introns that contain stop codons can direct a transcript to the mRNA surveillance pathway. Additionally, some splicing errors could cause a shift in the reading frame of a normal downstream exon, thus generating an in-frame stop codon. It has been proposed that the primary function of the mRNA surveillance system is to eliminate improperly processed transcripts[33]. However, recent data in *Caenorhabditis elegans* suggests that the major targets of the mRNA surveillance system are not aberrantly spliced RNAs, but rather transcripts that are deliberately spliced to contain premature stop codons as part of an intricate

Box 2. Alternative splicing and functional genomics

An extremely important area of the functional genomics field will be to determine the functional differences of proteins encoded by alternatively spliced mRNAs. This could be done on a gene-by-gene basis in transgenic animals that are unable to use a specific alternative exon either globally or in a tissue-specific manner. However, given the large number of genes that encode alternatively spliced transcripts, it will be necessary to develop high-throughput methods to achieve this goal. The recent development of methods for performing RNA-MEDIATED INTERFERENCE (RNAi) on a large-scale in *C. elegans*[a,b] could be applied to this problem. Libraries encoding double-stranded RNA directed against alternative exons could be used to inactivate transcripts containing those exons systematically. Future technological advances might make this, or a similar, approach feasible in other model organisms such as *D. melanogaster* and mice.

References

a Fraser, A.G. *et al.* (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408, 325–330

b Gönczy, P. *et al.* (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408, 331–336

autoregulatory system[34]. These results point out that putative erroneous splicing events deserve close scrutiny before their dismissal.

Tuning out the noise

How much of the alternative splicing that takes place is noise? Of all the alternative splicing known, only a handful of examples have clear functions and biological relevance ascribed. The most famous example is the *Drosophila* sex-determination system in which an alternative splicing cascade involving the transcripts of at least five genes (*sxl*, *tra*, *msl2*, *dsx*, and *fru*) culminates in the expression of male or female physical and behavioral traits[35]. However, in most cases, the relevance of alternative splicing is not known. For some alternative splicing events, functional consequences have been inferred based on the changes in the protein sequence introduced by an alternative exon (for example, *Dscam*[22]). For other genes, however, the functional consequences of alternative splicing are not obvious. Some groups have failed to detect differences in the activity of protein isoforms. However, negative results by no means indicate that there are no functional differences. It is equally, if not more, likely that the sequences encoded by an alternative exon confer very subtle, yet important biological effects upon the protein, undetectable by the experiments performed. Unfortunately, in most instances, experiments that address functional differences between protein isoforms encoded by alternatively spliced transcripts have not even been performed. Thus, the fact remains that the functional relevance of the vast majority of alternative splicing events is unknown and will probably remain that way for some time to come. One of the greatest challenges of the functional genomic and proteomic research in the future will be to fill this void (see Box 2).

One way to determine whether an alternative splicing event is real or represents noise is to examine whether it also occurs in a related species. The nucleotide sequence of functionally neutral alternative exons will not be under selective pressure and should not be conserved in distantly related species, whereas functionally relevant alternative exons should be. In fact, a recent comparison of the *C. elegans* genome with a portion of the genome of the related worm *Caenorhabditis briggsae* revealed that the intron sequences flanking alternatively spliced exons are frequently conserved between the two species[36]. Upon the completion of the mouse genome, this type of analysis will be relatively straightforward for human genes encoding alternatively spliced transcripts. There is, however, one caveat – alternative exons that are not evolutionarily conserved are not necessarily unimportant but might simply have evolved recently. This is not that far-fetched because single point mutations can create new splice sites[27] and potentially introduce a new alternative exon. In fact, there are documented cases in which the splicing patterns of paralogous genes differ across species[23].

Conclusion

Alternative splicing has a significant and under-appreciated role in expanding proteome diversity. Hopefully, many will take from this review the fact that we are generally ignorant of the scope of functionally relevant alternative splicing events. This simply underscores the tremendous amount of work ahead of us. One confounding fact emerging from the deluge of genomic sequence is that the number of genes does not seem to correlate with the complexity of an organism. *C. elegans*[37], *D. melanogaster*[38] and *Homo sapiens*[3,4] are predicted to contain ~19 000, 14 000 and 35 000 genes, respectively. It does not seem possible that the complexity of an organism can be explained by the one gene, one protein hypothesis. Thus, what some consider noise might actually be crucial in facilitating the development of complex organisms from a limited number of genes.

Acknowledgements

I thank R. Reenan for many stimulating discussions on this topic, and D. Black, E. Fleming, K. Hertel, K. Lynch, T. Maniatis and members of my laboratory for their comments on the manuscript. I apologize to colleagues whose work was not cited due to space limitations.

References

1 Smith, C.W. and Valcárcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388

2 Lopez, P.J. and Seraphin, B. (2000) YIDB: the Yeast Intron DataBase. *Nucleic Acids Res.* 28, 85–86

3 Roest Crollius, H. *et al.* (2000) Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat. Genet.* 25, 235–238

4 Ewing, B. and Green, P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* 25, 232–234

5 Davis, C.A. *et al.* (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* 28, 1700–1706

6 Spingola, M. and Ares, M., Jr (2000) A yeast intronic splicing enhancer and Nam8p are required for Mer1p- activated splicing. *Mol. Cell* 6, 329–338

7 Mironov, A.A. *et al.* (1999) Frequent alternative splicing of human genes. *Genome Res.* 9, 1288–1293

8 Hanke, J. *et al.* (1999) Alternative splicing of human genes: more the rule than the exception? *Trends Genet.* 15, 389–390

9 de Souza, S.J. *et al.* (2000) Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12690–12693

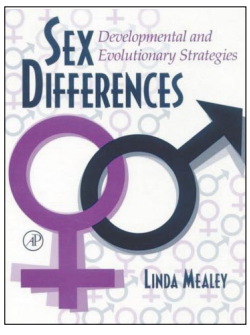
10 Dias Neto, E. *et al.* (2000) Shotgun sequencing of

the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3491–3496
11 Croft, L. *et al.* (2000) ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat. Genet.* 24, 340–341
12 Black, D.L. (1998) Splicing in the inner ear: a familiar tune, but what are the instruments? *Neuron* 20, 165–168
13 Navaratnam, D.S. *et al.* (1997) Differential distribution of Ca^{2+}-activated K^+ channel splice variants among hair cells along the tonotopic axis of the chick cochlea. *Neuron* 19, 1077–1085
14 Rosenblatt, K.P. *et al.* (1997) Distribution of Ca^{2+}-activated K^+ channel isoforms along the tonotopic gradient of the chicken's cochlea. *Neuron* 19, 1061–1075
15 Ramanathan, K. *et al.* (1999) A molecular mechanism for electrical tuning of cochlear hair cells. *Science* 283, 215–217
16 Missler, M. and Südhof, T.C. (1998) Neurexins: three genes and 1001 products. *Trends Genet.* 14, 20–26
17 Scheiffele, P. *et al.* (2000) Neuroligin expressed in non-neuronal cells triggers presynaptic development in contacting axons. *Cell* 101, 657–669
18 Ullrich, B. *et al.* (1995) Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* 14, 497–507
19 Ichtchenko, K. *et al.* (1995) Neuroligin 1: A splice site-specific ligand for b-neurexins. *Cell* 81, 435–443
20 Sugita, S. *et al.* (1999) Neurexins are functional a-latrotoxin receptors. *Neuron* 22, 489–496
21 Missler, M. *et al.* (1998) The making of neurexins. *J. Neurochem.* 71, 1339–1347
22 Schmucker, D. *et al.* (2000) *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671–684
23 Thackeray, J.R. and Ganetzky, B. (1995) Conserved alternative splicing patterns and splicing signals in the *Drosophila* sodium channel gene *para*. *Genetics* 141, 203–214
24 Loughney, K. *et al.* (1989) Molecular analysis of the *para* locus, a sodium channel gene in *Drosophila*. *Cell* 58, 1143–1154
25 Palladino, M.J. *et al.* (2000) A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. *Cell* 102, 437–449
26 Hanrahan, C.J. *et al.* (2000) RNA editing of the *Drosophila para* Na^+ channel transcript. Evolutionary conservation and developmental regulation. *Genetics* 155, 1149–1160
27 Cooper, T.A. and Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.* 61, 259–266
28 Jensen, K.B. *et al.* (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* 25, 359–371
29 Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25, 106–110
30 Reenan, R.A. *et al.* (2000) The *mle*$^{(napts)}$ RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the *para* Na^+ channel transcript in a region of RNA editing. *Neuron* 25, 139–149
31 Ibba, M. and Soll, D. (1999) Quality control mechanisms during translation. *Science* 286, 1893–1897
32 Hilleren, P. and Parker, R. (1999) Mechanisms of mRNA surveillance in eukaryotes. *Annu. Rev. Genet.* 33, 229–260
33 Pulak, R. and Anderson, P. (1993) mRNA surveillance by the *Caenorhabditis elegans smg* genes. *Genes Dev.* 7, 1885–1897
34 Mitrovich, Q.M. and Anderson, P. (2000) Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*. *Genes Dev.* 14, 2173–2184
35 Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* 32, 279–305
36 Kent, W.J. and Zahler, A.M. (2000) Conservation, regulation, synteny, and introns in a large-scale *C. briggsae–C. elegans* genomic alignment. *Genome Res.* 10, 1115–1125
37 The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018
38 Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195

Book Review

Cocksure Jones*

Sex Differences: Developmental and Evolutionary Strategies
by Linda Mealey
Academic Press, 2000. $79.95 hbk (xix + 1480 pages) ISBN 0 12 487460 6



Type the term 'masculinity' into Amazon.com (not, perhaps, the most obvious source) and 879 titles emerge. A quarter are by women. They dissect the hapless Y-bearer in volumes that range from the baffling – *The Penis Book, an Owner's Manual* by the (no doubt well informed) Margaret Gore – and bilious (*If Men could Talk, What would they Say?*) to the beastly (*Men and Other Reptiles*) and botanical (Risa Glickman's *Why Cucumbers are Better than Men*).

Why do they hate us so much? And why do so few males stand up for their own? Faced with this torrent of oestrogen-fuelled invective, *The Phallic Quest; Priapus and Masculine Inflation* or *Naked and Erect: Male Sexuality and Feeling* can but wilt (although *Awakening from the Deep Sleep: a Powerful Guide for Courageous Men* might go some way to help).

Such works reflect a potent sense of difference, fuelled by biology. Men, it seems, are slaves to the tedious little chromosome that switches their sexual machinery onto the rocky track to manhood. The pathetic fallacy, that human nature is driven by genes, is alive, well and available on-line.

Males are, of course, different from females. They have smaller gametes, for a start. The reason behind, and the consequences of, that protoplasmic imbalance lie behind much of biology. Linda Mealey's new book is, in part, a clear statement of where we stand in the battle for sexual understanding. From the origin of sex, to the ratios of males to females and why some women cannot draw the line at which the surface of a half-full glass of beer settles when tipped, much of reproductive life is there. Some is familiar, but much is not and, with 2000 citations, this is, it seems, a worthy addition to the scientific study of sex differences.

That is until it gets to humans; then it begins to fall apart. Of course men and women differ in behaviour – peek into any urinal to see how. Perhaps, indeed, man's penchant for murder or masturbation traces itself to the hypothalamus: as might his moderately ridged (and perhaps sexually selected) genital organ. But, damn it, let's have some *proof*. After all, *Drosophila* behaviour asks for no less; but when it comes to ourselves, scientific standards fly out of the window.

Some of the research reported here is fine, some is batty ('killer sperm' come to mind), but huge tranches are simply soft-headed. Thus, we learn that mate-guarding is an evolved strategy whereby

*Commissioned by *TREE* and published in the January 2001 issue of *TREE*.