Symbiotic DNA in eukaryotic genomes

Clifford Zeyl and Graham Bell

The recent explosive growth of molecular genetic databases has yielded increasingly detailed insights into the evolutionary dynamics of eukaryotic genomes. DNA sequences with the self-encoded ability to transpose and replicate are unexpectedly abundant and widespread in eukaryotic genomes. They seem to be sexual parasites. By dispersing themselves among the chromosomes, they increase their transmission rates and can invade outcrossing populations despite reducing host fitness. Once established, molecular parasites may themselves be parasitized by other elements, and through selection for reduced virulence may become beneficial genes. Elements have been isolated at various stages in this progression, from transposons that regulate their own transposition rates, to fundamental components of eukaryotic cytology, such as telomeres.

Clifford Zeyl and Graham Bell are at the Dept of Biology, McGill University, 1205 ave Dr. Penfield, Montréal, Québec, Canada H3A 1B1.

Eukaryotic cells carry a fascinating diversity of DNA molecules. The most familiar of these are the nuclear, mitochondrial and chloroplast chromosomes on which conventional genes are carried. The cellular environment created by these genes, with its efficient mechanisms for replicating and expressing DNA, has been repeatedly colonized by plasmids, supernumerary or B chromosomes, and retroviruses. Because the transmission rates of these molecules are not governed by an equitable process such as meiosis, they can spread despite reducing host fitness.

In addition, the chromosomes themselves carry a variety of DNA sequences that replicate and transpose within genomes. Based on their structures, transposition mechanisms and inferred phylogenies[1], mobile elements may be classified as transposons (Box 1), retrotransposons (Box 2), LINEs (Box 3) or mobile introns (Box 4). The evolutionary relationship between these mobile elements and the genomes they inhabit is one of the most intriguing aspects of molecular genetics.

The relationship between mobile and stationary DNA

Mobile elements have been thought to spread by increasing individual fitness, as do conventional genes, either by regulating gene expression or by increasing the long-term adaptability of a lineage.

Because some mobile elements are inserted preferentially into regulatory regions upstream of genes, rather than into open reading frames, they may be more likely than other mutagenic agents to produce adaptive novelty[9]. Mobile elements are responsible for many mutations that alter the timing or tissue specificity of gene expression, but few animal, fungal or protist examples of adaptive patterns of gene regulation attributable to mobile element insertions have been found. However, recent applications of the polymerase chain reaction (PCR) and database searching by Bureau and Wessler[10] revealed that plants are strikingly different in this respect. *Tourist* and *Stowaway* are small (<350 bp) mobile elements associated with the introns or flanking sequences of dozens of plant genes. *Tourist* and *Stowaway* apparently supply many wild-type regulatory and processing signals. We shall explain why the production of adaptive mutations does not demonstrate that an element has spread because it increases organismal fitness, but *Tourist* and *Stowaway* have clearly had important effects on the evolution of angiosperm genes.

Mobile elements can also alter gene expression by causing chromosomal mutations. Translocations, inversions, deletions and duplications are induced by the presence of the same sequence at several locations, and by the recombinogenic nature of most transposition mechanisms. Chromosomal instability resulting from ectopic recombination is so consistently deleterious or lethal that it may limit the abundance of some mobile elements, if the frequency of chromosomal mutation is determined by their copy number[11].

Mobile elements also mutate single loci, by transposing into coding regions. In well-studied eukaryotes, mobile elements are responsible for a large proportion of mutations. Seventy-nine per cent of spontaneous mutations at 10 loci in *Drosophila* have been attributed to insertions of mobile elements[12], although transposition rates vary greatly among elements, target loci, genetic backgrounds and environmental conditions. While the vast majority of mutant alleles produced in this way are null or defective, it has been suggested that this additional source of mutation gives to populations carrying a mobile element the long-term advantage of greater evolutionary flexibility. Transposition in experimental populations can increase the variability of metabolic and quantitative traits[13,14] and consequently the rate of response to artificial selection[13]. However, this can happen only when the element is present in a large number of individuals. Because the ratio of beneficial to harmful mutations is very low, mobile elements are unlikely to produce beneficial mutations when rare[15]. The production of beneficial mutations does not appear to be a significant factor in the population dynamics of even well-established elements in *D. melanogaster*. For three different families of mobile element, many different sites of insertion are found, but the population frequency of each insertion is very low, which is inconsistent with any of these insertions causing an increase in fitness[15].

This does not mean that established elements cannot have major evolutionary effects. They may well affect

Box 1. Transposons: cut-and-paste elements

Transposons move by excision from a chromosome, followed by reinsertion into a new locus (see Fig. 1). They are delimited by short (usually 10–40 bp) inverted repeats and contain one or more open reading frames, which encode the protein(s) responsible for transposition. The insertion of a transposon produces a duplication of a specific length, usually three to nine base pairs, at the target site, and subsequent excision of the transposon usually leaves several base pairs behind as a 'footprint'.

This cut-and-paste mechanism seems inherently conservative, or incapable of increasing copy numbers. However, in current models of the transposition mechanism, excision preferentially occurs during DNA replication, from a replicated locus into an unreplicated region[2]. In addition, the double-stranded break at the donor site may be repaired using its sister chromatid as a template, restoring the source copy[3]. These models are supported by major increases in the intragenomic copy numbers of transposons such as the *Drosophila* P-element[4] and maize *Activator* element following their introduction into new genomes by transformation or crosses.

Box 2. Retrotransposons and their similarities to retroviruses

Retrotransposons are transcribed and encode enzymes that insert DNA copies of these transcripts at new loci. Their transposition is thus replicative. Retrotransposons are bounded by repeated sequences, which in contrast to those of transposons are hundreds of base pairs in length and are in the same orientation. In addition to full-length retrotransposons, host genomes may contain dozens or hundreds of isolated copies of these long terminal repeats (LTRs), produced when recombination between the LTRs deletes the retrotransposon and leaves behind a single LTR.

The process of retrotransposition is fairly well-characterized, and is very similar to the retroviral life cycle. Transcripts are exported from the nucleus to the cytoplasm packaged inside capsid proteins encoded by *gag* genes, and are translated by host enzymes into a polypeptide that cleaves itself into separate proteins. The transcripts are copied into DNA by reverse transcriptase, and integrases then insert these copies into the chromosomes (Fig. 1). Retroviruses differ from retrotransposons in their capacity for extracellular transmission. In addition to the gene products required for retrotransposition, they encode envelope proteins, which mediate the interactions with host cell membranes that are required to leave one cell and infect another.

Retrotransposons fall into two classes. One group, typified by the *17.6*, *297* and *gypsy* or *mdg4* elements of *Drosophila*, resembles some retroviruses in the arrangement of its protein-coding sequences. Some also encode an envelope protein. In fact, Kim *et al.*[5] consider *gypsy* to be the first known retrovirus of an invertebrate, having found that it can be transmitted via food sources. The other group of retrotransposons, which includes the *Drosophila copia* and *S. cerevisiae* Ty1 elements, has the protein-coding sequences in a different order and lacks envelope proteins, and so cannot produce infective particles capable of extracellular transmission.

Box 3. LINEs

Most of what is known about LINEs (**l**ong **in**terspersed **e**lements) is based on the L1 family, the predominant LINE in mammalian genomes, although the F- and I-elements of *Drosophila* and *Cin4* of maize are also well-studied LINEs[6]. The copy numbers of LINEs vary widely, from a single active I-element to tens of thousands of L1 elements. LINEs are believed to transpose as RNA, and encode reverse transcriptases, but differ from retrotransposons in lacking structural similarities to retroviruses, such as LTRs. LINE transposition frequently generates deletions of varying lengths at the 5′ ends, so that the 3′ ends of some LINEs are much more abundant than their 5′ ends.

Box 4. Mobile introns

Introns that interrupt the coding sequences, and that are spliced from pre-mRNA transcripts by a complex of RNA and proteins called the spliceosome, are typical of eukaryotic genes. Mobile introns differ from such conventional introns in having a specific secondary structure, and in the relative autonomy of their splicing. Mobile introns fall into two distinct groups. Group I introns differ from group II introns in their secondary structure and self-splicing mechanism[7], which is sufficiently similar to the splicing of conventional introns by spliceosomes that group II introns have been hypothesized to be ancestral to nuclear mRNA introns.

The most independent of mobile introns are self-splicing molecules, which excise themselves from transcripts. Others depend on enzymes encoded by host genes. In addition to self-splicing, at least some mobile introns encode an ability to insert copies of themselves into homologous sites in alleles lacking the intron. The mitochondrial ω element of *Saccharomyces cerevisiae* contains an open reading frame encoding an endonuclease that makes double-stranded breaks within a specific 18-bp sequence. This activity results in gene conversion of such efficiency that in crosses between ω^+ and ω^- strains, nearly all of the progeny inherit the intron[8]. Finally, there is recent evidence that some mobile introns can transpose by reverse-splicing into RNA, followed by reverse transcription and integration into a chromosome. Although most mobile introns discovered thus far are located in organelle genomes, nuclear mobile introns are also known[7].

speciation[16], macroevolution[17], or chromosome architecture and evolution[18], and their legacy in eukaryotic cytology may be more pervasive than is generally thought[19]. However, this is not why they evolved the capacity for mobility, since it does not explain their initial spread.

Mobile genetic elements as genomic parasites

As the abundance and ubiquity in eukaryotic genomes of mobile DNA became apparent, Orgel and Crick[20] and Doolittle and Sapienza[21] hypothesized that mobile elements are molecular parasites, multiplying within and spreading among genomes despite reducing their fitness. The evolutionary success of conventional genes is based on the reproductive success of the genomes in which they reside. By contrast, mobile elements can overcome the equal segregation imposed by meiosis, by replicating within a genome, dispersing themselves among the chromosomes, and increasing their likelihood of infecting new sexual lineages. In a cross between individuals with and without a mobile element, the element is inherited by 50–100% of the offspring, depending on the efficiency of its intragenomic replication. A mobile element can therefore spread through a population despite reducing the fitness of individuals by up to half[22].

This parasitic invasion requires sex, since in an asexual population mobility within a genome permits no spread to uninfected genomes and, if it reduces fitness, will only hasten the extinction of a lineage bearing the mobile element. Thus, a direct and testable prediction of the hypothesis that eukaryotic mobile elements are parasitic is that they can invade only sexual populations. No experimental test of eukaryotic elements has yet been reported, but we may consider whether their molecular biology is consistent with that expected of molecular parasites.

The distribution among genomes of parasitic DNA should be restricted by the occurrence of sex. Some genomes are transmitted asexually, while fusion and recombination occur in others. Parasitic DNA should be restricted to the latter[22], with the provision that evolutionary losses of sexuality and rare horizontal transfer may strand parasitic DNA in currently asexual genomes. A rigorous test of this prediction must await a broader knowledge of the distributions of sex and of mobile DNA, but observations to date broadly support it. The mitochondrial and chloroplast genomes of most plants and animals are transmitted uniparentally, and are also largely free of mobile elements, while their generally sexual nuclear genomes carry numerous elements of all types. Both mobile introns and biparental inheritance are widespread among the mitochondrial and plastid genomes of algae and fungi.

A striking exception to this trend is the chloroplast genome of *Euglena gracilis*, which is burdened by at least 155 mobile introns, comprising 39% of the genome and forming complex nested structures of introns within introns[23]. As the euglenoid lineage appears to be totally asexual, we view this extreme proliferation of mobile introns as an evolutionary paradox.

The activity of parasitic DNA should be deleterious to its host. The spread of mobile elements is thought to require biased transmission to sexual progeny because they reduce host fitness, primarily through the genetic damage caused by transposition. The Ty1 retrotransposon of budding yeast is probably typical in reducing mean fitness while increasing the variance of fitness among individuals[9].

Mutations of the mitochondrial chromosome caused by the mobile intron α of the fungus *Podospora anserina* are now

Box 5. Hybrid dysgenesis

Hybrid dysgenesis is a syndrome of lethality and sterility in the progeny of crosses among some *Drosophila* strains. When male flies carrying particular mobile elements are crossed with females lacking the elements, transposition rates increase by orders of magnitude, causing greatly elevated rates of mutation and chromosomal rearrangement. P-elements, which are transposons, and I-elements, which are LINEs, are well-characterized elements that cause hybrid dysgenesis. Both transpose only in gonadal tissues. The germline specificity of P-elements results from tissue-specific splicing: an intron between the second and third open reading frames is spliced out only in gonadal tissue, so that in somatic tissues the translation product is truncated by a stop codon, and, instead of acting as a transposase, is a possible repressor of P-element transposition. P-elements that have been abundant in a *Drosophila* strain for several generations cease to transpose even in the germline, possibly repressed by the same mechanism that inhibits transposition in somatic tissues[26].

P-elements, and less clearly I-elements, appear to have invaded wild populations of *D. melanogaster* within the past 60 to 70 years, as rapid range expansions and dispersal brought localized populations harbouring the elements into contact with those lacking them. Despite the fitness costs imposed by hybrid dysgenesis[27], P-elements spread rapidly from initial frequencies as low as 1% to near fixation within 20 generations in experimental populations[4].

known to be responsible for two fatally degenerative processes: premature-death syndrome, which results from a specific deletion of mitochondrial DNA (mtDNA) involving the 5′ end of α, and senescence itself, which is associated with extrachromosomal amplification of the intron. The deletion causing premature-death syndrome results from recombination between α at its usual location and a second copy inserted by transposition into a nearby transfer RNA (tRNA) gene[24]. The persistence of this intron would be inexplicable if *P. anserina* were not a sexual fungus in which mitochondrial recombination permits the parasitic spread of α.

Mobile elements should be more active in germline than somatic cells. If transposition occurs in order to increase the transmission rates of mobile elements, and reduces host fitness, then selection should favour elements with limited mobility in somatic cells, since somatic transposition can cause deleterious mutations but provides no increase in transmission rates[25]. The elements responsible for hybrid dysgenesis in *Drosophila* (Box 5) epitomize this behaviour. Although ciliates are unicellular, they furnish analogous examples. Ciliates carry two nuclei: a macronucleus, responsible for gene expression, and a micronucleus, responsible for sexual transmission of the genes. The macronucleus and micronucleus are thus analogous to metazoan somatic and germline tissues, respectively. During macronuclear formation, the ciliate transposons Tec and TBE1 are excised, permitting the transmission of thousands of phenotypically silent micronuclear copies, which are hypothesized to encode their own excision as an extension of their transposase activity[28].

Molecular parasites of unicellular organisms may alternatively restrict their activity to sexual phases of the life cycle. The Ty3 retrotransposon of budding yeast is stimulated to transpose in haploid cells by pheromones produced by the complementary mating type, and is repressed in diploids, which cannot mate[29]. Ty3 therefore transposes only when imminent mating offers the potential for spread to new lineages.

Parasitic mobile elements should increase their rates of transposition when their host cells encounter stressful conditions, by the same reasoning that predicts increased sexuality and recombination in response to stress: if a particular genotype is faring poorly in a particular environment, either the environment has deteriorated or that genotype is inferior. In either case, its constituent genes maximize their chances of persistence by dispersing themselves among a variety of genotypes, some of which may be fitter in that environment[19]. Mutagens and thermal stress increase the mobility of at least some transposons[27]: Ty1 transposition rates can increase by two to 13 times in response to DNA damage caused by the mutagen NQO, and by 100 times when yeast is grown at 15° or 20°C instead of the optimal 30°C. Mobile elements, such as the *Tnt1* retrotransposon of tobacco, increase their activity during the stressful process by which protoplasts are prepared for cell culture. Some mobile elements transpose more frequently when either the activity of unrelated elements or ionizing radiation increases rates of DNA damage and mutation[27].

Parasites of parasitic DNA

Because many proteins encoded by mobile elements are *trans*-acting, they can be used by elements other than those that encode them. This frees the latter from the need to encode their own proteins, and deletion mutants may gain increased efficiency of intragenomic replication without sacrificing mobility. Because transposition mechanisms are recombinogenic, such mutants arise frequently and often come to outnumber full-length copies[30]. Such defective transposons are analogous to the defective interfering viruses that are dependent on their ancestors for gene products that they themselves no longer encode[31]. These truncated versions of viral genomes appear late in the course of some infections, outcompeting their ancestors. Most transposon families are dominated by defective copies that are mobile only in the presence of a copy encoding a functional transposase. For most transposons, the minimal requirements for transposition are the terminal inverted repeats, because the specific sequences with which the transposase interacts are usually within or near these terminal repeats. Defective copies in which these inverted repeats are lost or so mutated that they cannot be mobilized are no longer under any selective constraint, and drift in nucleotide sequence, eventually becoming unrecognizable. Typical eukaryotic genomes are littered with elements in varying stages of decay.

Parasitic dependence also arises between mobile introns. The yeast mitochondrial group I intron *al4* encodes an endonuclease permitting its insertion into unoccupied alleles, but is dependent for self-splicing on a maturase encoded by the *bl4* mobile intron[7]. Hyperparasitism could explain the spread of SINEs, a type of highly repeated sequence. SINEs are **s**hort (<500 bp) **in**terspersed **e**lements that are so highly repeated that particular families comprise substantial fractions of vertebrate genomes: 500 000 or so copies of the *Alu* element constitute about 5% of the human genome[32]. Although most SINEs currently appear to be immobile, some continue to transpose and spread[33]. Because SINEs transpose via RNA intermediates, they require reverse transcriptase for mobility, but unlike retrotransposons or LINEs they do not encode their own reverse transcriptase. They may have spread as parasites on other elements, such as retroviruses, retrotransposons, or LINEs, which could explain their relatively recent origin (about 65 million years ago[34]).

Coevolution between mobile elements and host genomes

Other than abstaining from sex, most eukaryotic genomes seem to have no method of preventing their invasion by mobile elements. However, mechanisms specifically evolved to combat molecular parasites may be difficult to detect using the usual technique of screening model organisms for mutants, because apart from suppressing

transposition or deleting mobile elements, such mechanisms are unlikely to have easily scored phenotypic effects. The repeat-induced point mutation (RIP) phenomenon may result from such a mechanism in the genome of *Neurospora crassa*. In both copies of any duplicated sequence, 10% to 50% of G:C pairs are mutated to A:T per sexual generation, rapidly inactivating any function that they encoded initially[35]. This process only occurs between fertilization and nuclear fusion, precisely when invasion by a mobile element would begin. The RIP system is thus ideally qualified to mutate and inactivate any sequence undergoing duplicative transposition, and is presumably responsible in part for the relative dearth of repetitive DNA in *N. crassa*.

Prud'homme *et al.*[36] recently discovered a *D. melanogaster* gene that controls the retrotransposon *gypsy*. In flies homozygous for *flamenco* mutations, the copy numbers and transposition rates of *gypsy* are elevated, but no other phenotypic effects are detected, even when *flamenco* is entirely deleted. Genome sequencing projects are likely to yield additional examples of host genes controlling molecular parasites.

Selection among vertically transmitted organismal parasites favours those that maximize transmission rates, while minimizing damage to their hosts. The same principle applies to genomic parasites, whose transmission requires host reproduction: the most successful molecules will be those with high rates of transposition and thus transmission, but with minimal mutagenic effects. The parasitic DNA hypothesis therefore leads to contrasting predictions regarding the activities of mobile elements in naive versus colonized hosts, and implies that through long-term coevolution with colonized hosts, established mobile elements will be domesticated through selection for minimized virulence.

Parasitic DNA should show greater activity in naive hosts than in colonized hosts. Transposition yields the greatest increases in transmission rates in newly invaded genomes, in which nearly all chromosomes are unoccupied, whereas elements gain nothing by further transposition if they would already be transmitted to all sexual progeny. Transposition rates are thus highest in crosses between genomes in which an element is established and uninfected genomes, as in hybrid dysgenesis (Box 5).

Once established, some elements regulate their own transposition rates. Modelling suggests that, in outcrossing diploids, self-regulation can evolve only for elements that cause dominant lethal or sterile mutations[25]. Transposons, such as the *Activator* (*Ac*) family of maize and the P-elements of *Drosophila*, which cause chromosome breakage and fusion, are the type of element that is most likely to meet these requirements. It is these transposons that encode self-regulatory functions (see Box 5). The evolution of P-element self-regulation has been observed in the laboratory following the invasion of an experimental population by initially-active P-elements[4].

Established parasitic DNA should be phenotypically silent. Long-term selection for reduced impact on host fitness can favour adaptations to host genomes that minimize the deleterious effects of mobile elements. The germline specificity discussed above is one such attribute. Others include insertion-site specificity and the splicing of elements from transcripts of target genes.

Although local hotspots for insertion exist, most transposons, retrotransposons, LINEs and SINEs of large nuclear genomes show no sequence specificity beyond general preferences for AT-rich or GC-rich regions. However, the retrotransposons of *Saccharomyces cerevisiae*, which has an unusually concise nuclear genome, have unusual insertion-site specificity. Ty3 elements insert immediately upstream of tRNA genes[29]. While the specificity of Ty1 is less absolute, it too avoids open reading frames and favours tRNA regions[37]. Because multiple copies of tRNA genes exist, an insertion at any one of them is unlikely to have very harmful effects[29]. Mobile introns are the most specific of all mobile elements, often being restricted to one locus, where self-splicing also renders them phenotypically silent. Mobile introns are also the only elements commonly found in mitochondrial and chloroplast genomes, which in general contain very little non-coding DNA. Thus the insertion specificities of mobile elements reflect the sizes of their host genomes and the stringency of selection against their mutagenic activity.

Finally, while transposons and retrotransposons are unable to splice themselves from transcripts as do mobile introns, some carry splicing signals, which are recognized by the spliceosomes responsible for excising conventional introns. Some insertions are thus removed from transcripts, partially or completely restoring wild-type function to these alleles despite interrupting an open reading frame[38].

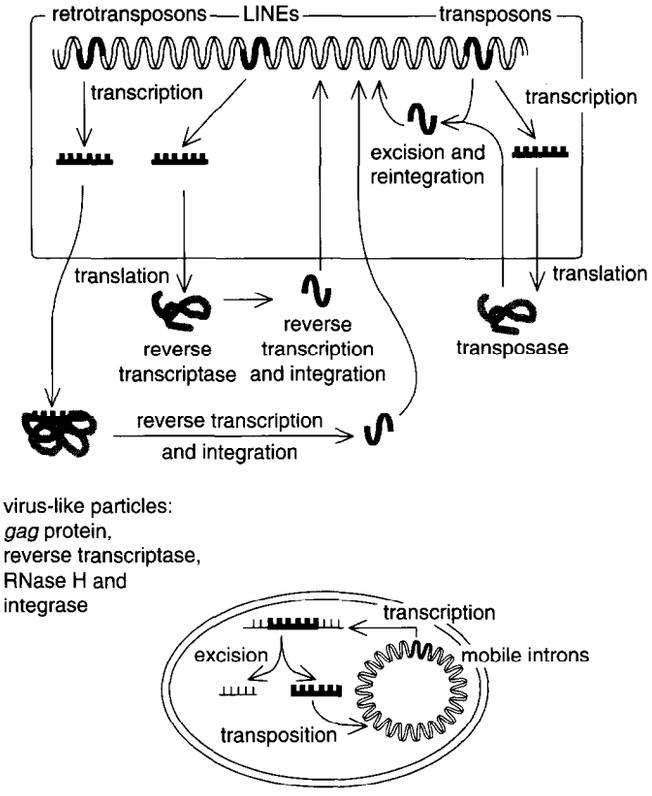

Fig. 1. Schematic diagram of mobile elements in the nuclear and mitochondrial genomes of a hypothetical eukaryotic cell. Transposons, retrotransposons and LINEs are transcribed into nuclear pre-mRNA, processed and exported to the cytoplasm, and translated. Transposases enter the nucleus and are responsible for excising and reinserting the transposons that encode them (Box 1). The reverse transcriptases that are encoded by LINEs copy LINE transcripts into DNA that is then integrated into chromosomes. Retrotransposons encode several gene products. Following translation, these form complexes of protein and retrotransposon RNA resembling the structures formed by active retroviruses. Enzymatic components of these virus-like particles produce DNA copies of the retrotransposon RNA, enter the nucleus, and insert the new retrotransposons into chromosomes. Mobile introns are often found in mitochondrial or chloroplast genomes. In addition to self-excision from transcripts of the genes into which they are inserted, some mobile introns are capable of transposition into unoccupied loci. Because it is unclear how many mobile introns can do so, and the mechanisms vary among those that do (Box 4), a transposition mechanism is not indicated here for mobile introns.

The domestication of initially parasitic elements may be responsible for some of the most characteristic features of the eukaryotic genome, such as mating-type genes and telomeres. The potential for parasitic DNA to spread via host sexuality led to the hypothesis that the genes encoding sexuality are themselves descendants of transposons[22], a conjecture supported by the non-homologous, and in some cases mobile, nature of mating-type genes in many fungi[19]. Since telomeres encode a form of reverse transcriptase and contain simple repeated sequences, they may have arisen as mobile elements which, after selection for beneficial association with a complement of genes, now stabilize linear chromosomes[19]. The telomeres of the ciliate *Oxytrichia* are the terminal repeats of transposons that are still mobile[28]. One telomere of *N. crassa* is unrelated to other telomeres and is a dispersed repeated sequence 'hypothesized to be presently, or at one time, transposable'[39]. The end of the right arm of chromosome 3 in *D. melanogaster* is made up of tandem repeats of two different LINEs[40]. These telomeric arrays are continually eroded from the chromosome ends, and are replenished by occasional transposition from interstitial source copies. As the self-sacrificial behaviour of telomeric copies of these LINEs stabilizes the transmission of the interstitial copies, this can be seen as a case of molecular kin selection, which at the organismal level can favour altruism among closely related individuals. Having invaded a genome as parasitic DNA, these LINEs have become mutualistic symbionts whose continued self-replication stabilizes their transmission as well as that of linked stationary genes.

Mobile elements play major roles in the evolution of eukaryotic genomes. Their molecular biology shows that they spread as sexual parasites, but after they become established in a population, the very features that made them successful as parasites equip their domesticated descendants to drive some of the basic processes of the eukaryotic cell. As eukaryotic genomes are characterized in ever-increasing detail, we expect the pervasiveness and evolutionary significance of genomic parasites, their interactions with host genomes, and their descendants, to become increasingly apparent. Forty years ago, it became widely recognized that the eukaryotic cell has evolved as a community of symbiotic microbes of various kinds. Today, we may be on the brink of a similar realization, the interpretation of major components of the eukaryotic genome as a community of domesticated elements with independent ancestries as parasites.

Acknowledgements

Supported by grants from the Natural Sciences and Engineering Research Council and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche. We thank C. Waddell and three anonymous reviewers for comments on an earlier version of the manuscript, and S. Rosenberg for bringing the *Neurospora* RIP phenomenon to our attention.

References

1 Xiong, Y. and Eickbush, T.H. (1990) **Origin and evolution of retroelements based upon their reverse transcriptase sequences,** *EMBO J.* 9, 3353–3362

2 Coen, E.S., Robbins, T.P., Almeida, J., Hudson, A. and Carpenter, R. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 413–436, American Society for Microbiology

3 Daniels, S.B. and Chovnick, A. (1993) **P element transposition in *Drosophila melanogaster*: an analysis of sister-chromatid pairs and the formation of intragenic secondary insertions during meiosis,** *Genetics* 133, 623–636

4 Good, A.G., Meister, G.A., Brock, H.W., Grigliatti, T.A. and Hickey, D.A. (1989) **Rapid spread of transposable P elements in experimental populations of *Drosophila melanogaster*,** *Genetics* 122, 387–396

5 Kim, A. *et al.* (1994) **Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*,** *Proc. Natl Acad. Sci. USA* 91, 1285–1289

6 Hutchison, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 593–617, American Society for Microbiology

7 Cech, T.R. (1990) **Self-splicing of group I introns,** *Annu. Rev. Biochem.* 59, 543–568

8 Perlman, P.S. and Butow, R.A. (1989) **Mobile introns and intron-encoded proteins,** *Science* 246, 1106–1109

9 Wilke, C.M. and Adams, J. (1992) **Fitness effects of Ty transposition in *Saccharomyces cerevisiae*,** *Genetics* 131, 31–42

10 Bureau, T.E. and Wessler, S.R. (1994) **Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants,** *Plant Cell* 6, 907–916

11 Charlesworth, B. and Langley, C.H. (1989) **The population genetics of *Drosophila* transposable elements,** *Annu. Rev. Genet.* 23, 251–287

12 Green, M.M. (1988) in *Eukaryotic transposable elements as mutagenic agents* (Lambert, M.E., McDonald, J.F. and Weinstein, I.B., eds), pp. 41–50, Cold Spring Harbor

13 Torkamanzehi, A., Moran, C. and Nicholas, F. (1992) **P element transposition contributes substantial new variation for a quantitative trait in *Drosophila melanogaster*,** *Genetics* 131, 73–78

14 Clark, A.G., Wang, L. and Hulleberg, T. (1995) **P-element-induced variation in metabolic regulation in *Drosophila*,** *Genetics* 139, 337–348

15 Charlesworth, B. (1987) **The population biology of transposable elements,** *Trends Ecol. Evol.* 2, 21–23

16 Fontdevila, A. (1992) **Genetic instability and rapid speciation: are they coupled?** *Genetica* 86, 247–258

17 McDonald, J.F. (1990) **Macroevolution and retroviral elements,** *BioScience* 40, 183–191

18 Korenberg, J.R. and Rykowski, M.C. (1988) **Human genome organization: Alu, lines and the molecular structure of metaphase chromosome bands,** *Cell* 53, 391–400

19 Bell, G. (1993) **The sexual nature of the eukaryotic genome,** *J. Hered.* 84, 351–359

20 Orgel, L.E. and Crick, F.H.C. (1980) **Selfish DNA: the ultimate parasite,** *Nature* 284, 604–607

21 Doolittle, W.F. and Sapienza, C. (1980) **Selfish genes, the phenotype paradigm and genome evolution,** *Nature* 284, 601–603

22 Hickey, D.A. (1982) **Selfish DNA: a sexually-transmitted nuclear parasite,** *Genetics* 101, 519–531

23 Hong, L. and Hallick, R.B. (1994) **A group III intron is formed from domains of two individual group II introns,** *Genes Dev.* 8, 1589–1599

24 Sellem, C.H., Lecellier, G. and Belcour, L. (1993) **Transposition of a group II intron,** *Nature* 366, 176–178

25 Charlesworth, B. and Langley, C.H. (1986) **The evolution of self-regulated transposition of transposable elements,** *Genetics* 112, 359–383

26 Rio, D.C. (1991) **Regulation of *Drosophila* P element transposition,** *Trends Genet.* 7, 282–287

27 Ajioka, J.W. and Hartl, D.L. (1989) in *Mobile DNA* (Berg, D.E. and Howe, M.M., eds), pp. 939–958, American Society for Microbiology

28 Williams, K., Doak, T.G. and Herick, G. (1993) **Developmental precise excision of *Oxytrichia trifallax* telomere-bearing elements and formation of circles closed by a copy of the flanking target duplication,** *EMBO J.* 12, 4593–4601

29 Kinsey, P.T. and Sandmeyer, S.B. (1995) **Ty3 transposes in mating populations of yeast: a novel transposition assay for Ty3,** *Genetics* 139, 81–94

30 Brookfield, J.F.Y. (1991) **Models of repression of transcription in P–M hybrid dysgenesis by P cytotype and by zygotically encoded repressor proteins,** *Genetics* 128, 471–486

31 Nee, S. and Maynard Smith, J. (1990) **The evolutionary biology of molecular parasites,** *Parasitology* 100, S5–S18

32 Okada, N. (1991) **SINEs: short interspersed repeated elements of the eukaryotic genome,** *Trends Ecol. Evol.* 6, 358–361

33 Takasaki, N. *et al.* (1994) **Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by**

retroposition: a process of parasitization of entire genomes during the evolution of salmonids, *Proc. Natl Acad. Sci. USA* 91, 10153–10157
34 Deininger, P.L. and Batzer, M.A. (1993) in *Evolutionary Biology* (Vol. 27) (Hecht, K. *et al.*, eds), pp. 157–197, Plenum
35 Cambareri, E.B., Singer, M.J. and Selker, E.U. (1991) **Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*,** *Genetics* 127, 699–710
36 Prud'homme, N., Gans, M., Masson, M., Terzian, C. and Bucheton, A. (1995) ***Flamenco*, a gene controlling the *gypsy* retrovirus of *Drosophila melanogaster*,** *Genetics* 139, 697–711
37 Ji, H. *et al.* (1993) **Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences,** *Cell* 73, 1007–1018
38 Purugganan, M.D. (1993) **Transposable elements as introns: evolutionary connections,** *Trends Ecol. Evol.* 8, 239–243
39 Schlechtmann, M.G. (1990) **Characterization of telomere DNA from *Neurospora crassa*,** *Gene* 88, 159–165
40 Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A. and Sheen, F-M. (1993) **Transposons in place of telomeric repeats at a *Drosophila* telomere,** *Cell* 75, 1083–1093

Reconstructing shifts in diversification rates on phylogenetic trees

Michael J. Sanderson and Michael J. Donoghue

Few issues in evolutionary biology have received as much attention over the years or have generated as much controversy as those involving evolutionary rates. One unresolved issue is whether or not shifts in speciation and/or extinction rates are closely tied to the origin of 'key' innovations in evolution. This discussion has long been dominated by 'time-based' methods using data from the fossil record. Recently, however, attention has shifted to 'tree-based' methods, in which time, if it plays any role at all, is incorporated secondarily, usually based on molecular data. Tests of hypotheses about key innovations do require information about phylogenetic relationships, and some of these tests can be implemented without any information about time. However, every effort should be made to obtain information about time, which greatly increases the power of such tests.

Michael Sanderson is at the Section of Evolution and Ecology, University of California, Davis, CA 95616, USA; Michael Donoghue is at the Dept of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

Many questions in evolutionary biology revolve around the observation that some groups of organisms have more species than others. We wonder, for example, why beetles are so species-rich. Is it the result of some feature that evolved in the ancestor of beetles (an intrinsic cause, such as the evolution of elytra), or is this species richness related to an environmental shift that favored the diversification of organisms that had already evolved a particular set of attributes (an extrinsic cause, such as a climatic change or changes occurring in coevolving groups of organisms) or some more complicated combination of such causes? Still other questions pertain to replicated episodes of diversification. For example, did the evolution of latex in several clades of vascular plants promote their diversification by warding off insect pests[1]? Answers to such questions hinge on an accurate reconstruction of where and when shifts in diversification rate occured in relation to the appearance of possible intrinsic or extrinsic causes.

Techniques for reconstructing the evolution of organismal features in phylogenies have been well studied[2], but reconstructing the where and when of shifts in diversification rate in phylogenies has only recently received critical attention. Our aim is to compare and contrast these methods. A key issue is how phylogenetic information can provide new insights into species diversification. But we also hope to clarify the role that other information can play, especially the absolute timing of speciation and extinction events in evolutionary history.

Trees, times and durations

Ideally, diversification would be studied in clades in which phylogenetic relationships are clear and a rich fossil record has placed tight constraints on the timing of speciation and extinction events (Fig. 1a). Although a few such cases are known (e.g. North American Neogene horses[3]), they are not common. A complete set of information about diversification includes two items: the phylogeny, and times of events (including speciation, extinction, and perhaps the origin of novelties). Durations ('lengths') of branches, which are often used in studies of diversification, can be derived from information about the timing of events. Various methods for studying diversification use these items singly or in combination (Table 1).

Much of the ambiguity in the study of diversification rates stems from the use of a subset of the necessary information. For example, inferences based on durations alone are confounded by uncertainty about what marks the beginning and end of a lineage. Is it a branching event, a true extinction event, or a pseudo-extinction or pseudo-speciation event caused by phenotypic transformation into a 'new' species[4]? Were a phylogeny available, the length of a lineage could be clearly defined. In the absence of complete information, however, compromise is inevitable. Below we consider what can and cannot be inferred when almost everything is known about a clade (as in Fig. 1a), and what can be inferred when less is known (as illustrated by Figs 1b–d). One may know relationships but know nothing about time; know time but know nothing about relationships; and so on.

A variety of models have been used to study species diversification (Box 1). Because these inevitably entail some probability of error, inferences based upon them are statistical. Such errors may involve systematic biases in estimates of rates, large average errors (deviations from the true parameter regardless of direction) in those estimates,