THE OUTS AND INS OF TRANSPOSITION: FROM MU TO KANGAROO

M. Joan Curcio[*§] *and Keith M. Derbyshire*[‡§]

Transposons are ubiquitous in prokaryotic and eukaryotic organisms and are major determinants of genome structure. Transposition — the movement of discrete segments of DNA without a requirement for homology — occurs by a handful of mechanisms that are used over and over again in different combinations. Understanding these mechanisms provides an important key to unlocking the secrets of genome organization and evolution.

TRANSPOSABLE ELEMENT
A transposable element or a transposon is a defined segment of DNA that has the ability to move, or copy itself, into a second location without a requirement for DNA homology.

**Laboratory of Developmental Genetics and* [‡]*Division of Infectious Disease, Wadsworth Center, New York State Department of Health, PO Box 22002, Albany, New York 12201-2002, USA.*
[§]*Department of Biomedical Sciences, School of Public Health, State University of New York at Albany, Albany, New York 12201–2002, USA.*
Correspondence to K.M.D.
e-mail: keith.derbyshire@wadsworth.org
doi:10.1038/nrm1241

Fifty years ago, when Watson and Crick published the structure of DNA in their landmark paper[1], who would have thought that TRANSPOSABLE ELEMENTS would make up a significant fraction of the DNA of many genomes? However, recent DNA sequence analyses show that 40% or more of the human[2], mouse[3] and rice[4,5] genomes are composed of transposon-derived sequences. In lower eukaryotes and bacteria this fraction is smaller (1–5%), but it is still significant. Although many of these elements, particularly in higher eukaryotes, are 'fossils' of active elements, the mobility of transposons has had an important role in the structure and evolution of genes and genomes from bacteria to humans[6,7]. The insertion of transposons can both activate and inactivate genes depending on the location of their target (upstream or within a gene, respectively). Transposons also promote inversions and deletions of chromosomal DNA, either as a direct result of intramolecular transposition, or by providing dispersed regions of homology that can be recognized by the DNA recombination machinery of the host. Transposition of some elements can result in the transduction of flanking DNA, and so provide yet another means of rearranging host genes[8,9]. As many transposons, especially bacterial elements, encode additional functions such as antibiotic resistance and virulence factors, their dissemination among species (carried by plasmids and viruses) has contributed to the shared bacterial gene pool. Similarly, in eukaryotes, transposon-associated functions have been co-opted by the host to alter the genome structure and gene expression, as evidenced by V(*D*)J (variable (diversity) joining) rearrangements[10,11], nuclear intron splicing[12] and telomere maintenance[13,14].

How do all these different elements move? What are the different mechanisms of transposition? How similar are the pathways for related elements found in different organisms? Many of the answers have been provided by the elegant genetic, molecular and structural studies carried out on transposons in model organisms such as bacteria and yeast. Although the mechanistic details of different transposition systems have been determined using a relatively small number of transposons, these studies have shed light on the major pathways that are used by all transposons to 'break-and-join' or 'copy-and-join' DNA. Here, we will compare and contrast the known transposition mechanisms to provide a suitable framework for predicting the movement and rearrangements caused by elements still to be characterized.

Classifying transposons by their transposases

In the past, the classification of transposons has been based on a variety of factors that are not necessarily related to their mechanism of transposition. Although this approach has been useful, it is limited because many of the newly identified transposons do not contain the signature structural elements that are found in the earlier classes of transposon. For example, not all elements are flanked by inverted or direct repeats, or generate target-site duplications on insertion. In this era

TRANSPOSASE
An enzyme that is responsible for the catalysis of transposition.

DDE
A triad of highly conserved amino acids (aspartate (D), aspartate, glutamate (E)) that is found in one class of transposases, which are required for the coordination of metal ions that are necessary for catalysis.

REVERSE TRANSCRIPTASE
A DNA polymerase that can use RNA as a template for the synthesis of a cDNA.

ENDONUCLEASE
An enzyme that catalyzes the cleavage of a phosphodiester bond within a DNA molecule.

TYROSINE RECOMBINASE
An enzyme that is capable of rearranging DNA using a conserved tyrosine residue to cleave and reseal its substrates by a 3′-phosphotyrosine linkage (also known as integrase).

of large-scale genome sequencing, in which new elements are being described from diverse organisms at an unprecedented rate, a more informative way to categorize transposons is by the way they move, which is dictated by their TRANSPOSASE proteins.

So far, five protein families have been described that mediate transposition: DDE-transposases, rolling-circle (RC)- or Y2-transposases, tyrosine (Y)-transposases, serine (S)-transposases, and a fifth family, which encodes a combination of REVERSE TRANSCRIPTASE and ENDONUCLEASE activities (RT/En) (FIG. 1). These proteins have different catalytic mechanisms to mediate the various DNA breaks and joins that are necessary for transposition into a new target. Accordingly, some transposons cut, whereas others copy, themselves out from their starting site. Similarly, some elements 'paste', whereas others 'copy', themselves into the target. The variations and combinations of mechanisms add layers of complexity that hint at how these functions were acquired and spread between organisms, and they indicate that yet more are still to be discovered.

The best-characterized family of transposons includes those originally described by Barbara McClintock[15]. Most elements of this family move by excised DNA intermediates, employ a 'cut-and-paste' mechanism and are called DNA transposons, type II transposons and, now probably more correctly, DDE-transposons (for example, IS, P, Ac, Tc and Mariner elements)[7]. These elements all encode a transposase that has a related amino-acid motif, the DDE motif[16,17], which is responsible for excising the transposon from the donor and integrating it into the target (FIG. 1a).

Two other families of DNA transposons also use a 'cut-and-paste' mechanism. These elements encode proteins related to either site-specific TYROSINE[18,19] or SERINE[20] RECOMBINASES. These protein families are unrelated and mediate DNA-strand exchange by different mechanisms, but generate similar intermediates and products of transposition (FIG. 1e,f). The transposon is excised from its original site to form a circular DNA intermediate, and the original 'empty' site is restored. Following capture of a target, a reversal of the excision steps results in transposon insertion. These transposons have been found almost exclusively in bacteria (where they are called CONJUGATIVE TRANSPOSONS or integrating conjugative elements; ICEs)[21,22], but there is one family of eukaryotic elements[23,24] that encodes related proteins (for example, Kangaroo and DIRS1 elements) and are likely to move by a similar mechanism (FIG. 1d).

The 'copy' pathways require the incorporation of a replicative step in the transposition mechanism. The RC-transposons of bacteria[25] (and probably the related eukaryotic helitrons[26]) copy their DNA directly into the target by DNA replication such that each copy (new and old) of the transposon contains one newly synthesized strand (FIG. 1g). This process requires the DNA replication machinery of the host, but is initiated by a transposon-encoded protein that makes

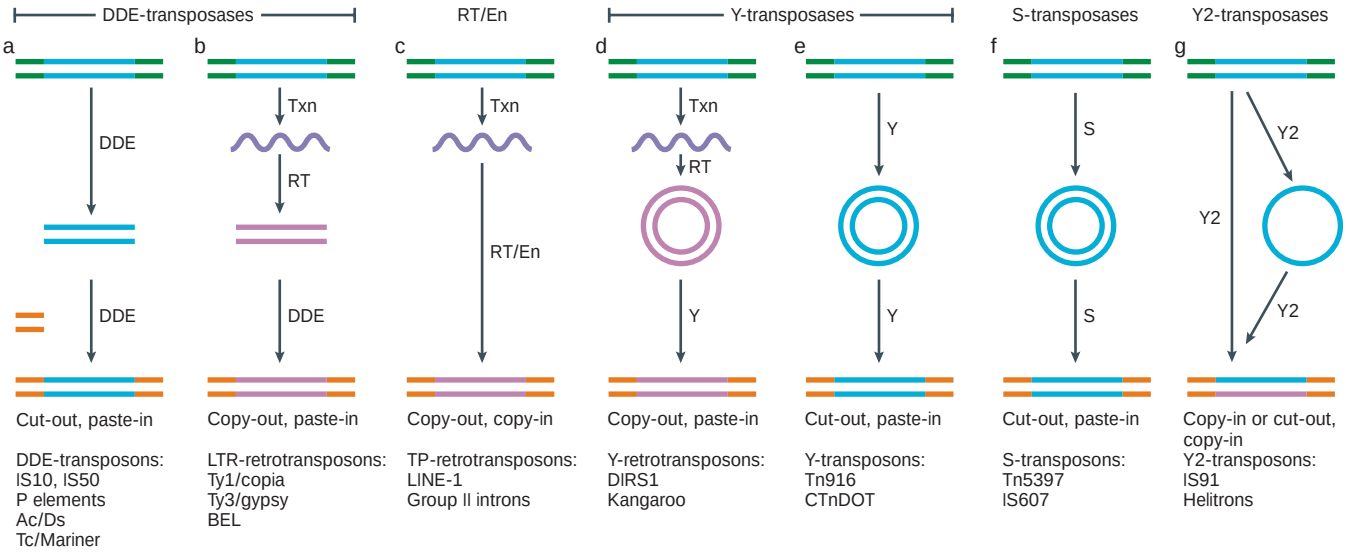


Figure 1 | **Transposons move in different ways.** Five protein families dictate different transposition pathways: DDE-transposases, reverse transcriptase/endonucleases (RT/En), tyrosine (Y)-transposases, serine (S)-transposases and rolling-circle (RC)- or Y2-transposases. Transposons (blue) can be either 'cut-out' or 'copied-out' of the flanking donor DNA (green). **a** | Most DDE-transposons excise from the flanking DNA to generate an excised linear transposon, which is the substrate for integration into a target (orange). **b** | Retrotransposons copy-out by reverse-transcribing (RT) a full-length copy of their RNA (purple) that is generated by transcription (Txn). Long-terminal repeat (LTR)-retrotransposons make a full-length cDNA copy (pink represents newly replicated DNA) from their RNA and integrate this into a target using a DDE-transposase. **c** | TP-retrotransposons use reverse transcriptase (RT) to copy their RNA directly into a target that has been nicked by a transposon-encoded nuclease (En). **d** | Y-retrotransposons are thought to generate a circular cDNA intermediate by reverse transcription. A Y-transposase integrates the element into the target. **e** | and **f** | Y- and S-transposons encode either a tyrosine or serine transposase, which mediates excision of the transposon to form a circular intermediate. A reversal of the catalytic steps results in transposon insertion. **g** | Y2-transposons 'paste' one strand of the transposon into a target and use it as a template for DNA replication. Two models have been proposed for Y2-transposition. Representatives of each type of transposon are listed below each pathway.

nicks at the ends of the transposon and in the target DNA. The resulting free 3′ OH ends are thought to prime DNA synthesis using the transposon strand as a template.

RETROTRANSPOSONS generate a copy of their DNA (cDNA) by reverse transcription of their RNA genome. For the purpose of this review, retrotransposons will be separated into three classes depending on which enzymatic activities they encode in addition to reverse transcriptase[23,27], as the mechanism with which they insert into a target is dictated by these other activities. For example, long-terminal repeat (LTR)- or DDE-retrotransposons (such as Ty1, gypsy and BEL)[27] and retroviruses (such as HIV-1)[17,28] have co-opted DDE-transposases to integrate their cDNA into a target — a step that is mechanistically equivalent to DDE-transposon insertion (FIG. 1b).

The DIRS1 family forms a second class of retrotransposons — Y-retrotransposons — that encodes Y-transposases rather than DDE-transposases and lack LTRs[23,24]. These elements are thought to generate excised circular cDNA copies of their transposon by reverse transcription of their RNA[29]. The Y-transposase is proposed to insert the circular DNA intermediate into the genome[23,24] (FIG. 1d).

The third class of retrotransposons is often referred to as non-LTR-retrotransposons[7]. However, this name has lost its use because of the recent characterization of DIRS-like elements as another class of retrotransposons that lack canonical LTRs. Therefore, we will use a name based on their mechanism of mobility: target-primed. TP-retrotransposons represent the fifth family of transposon, which encodes an endonuclease in addition to reverse transcriptase (RT/En). The endonuclease nicks the target site, and the nick then serves as a primer for reverse transcription of an RNA copy of their genome[30,31] (FIG. 1c). Therefore, unlike LTR-retrotransposons and Y-retrotransposons, the RNA intermediate of TP-retrotransposons is an integral component of the integration process.

DDE-transposons

The invariant DDE residues that are encoded by DDE-transposases are found in three non-contiguous patches of more weakly conserved residues[16,32,33]. Their importance was established by mutagenesis studies, which showed that the DDE motif was essential for transposition *in vitro* and *in vivo*[7,17]. Crystallographic studies have since shown that, despite the lack of sequence similarity, the structures of the catalytic domains, including the location of the invariant DDE residues, are largely superimposable for both transposases (Mu, REF. 34; and IS50, REF. 35) and retroviral INTEGRASES (HIV-1 and ASLV[36–38]). This, together with the structural data that locate divalent metal ions and both transposon ends at the active site, supports the role of the DDE motif in coordinating the metal ions to facilitate catalysis[35,39,40]. DDE-enzymes catalyse two chemical reactions. First, HYDROLYSIS of the phosphodiester backbone at each end of the transposon generates free 3′ OH ends. Second, the exposed 3′ OH ends are joined to the target DNA in a *TRANS*-ESTERIFICATION reaction[41]. These steps take place within a nucleoprotein complex, called the TRANSPOSOSOME, which ensures the concerted insertion of each transposon end into the target[42,43]. The NUCLEOPHILIC ATTACK by the two 3′ OH groups on each strand of the target is staggered and, generally, separated by 2–9 nucleotides. Repair of this segment results in a target duplication — a hallmark of DDE-transposition — the length of which is characteristic for each transposon.

Mu and Tn3 cleave only one transposon strand. Although the DDE-transposases perform the same chemical reactions, the mechanisms that are used to generate a DNA substrate for strand transfer differ significantly between the transposons[44] (FIG. 2). Several bacterial transposons (epitomized by the Mu and the Tn3 families) simply nick and join the 3′ ends of the transposon to the target[45,46]. This fuses donor and target DNAs to form a strand-transfer, or Shapiro, intermediate[47]. Replication from the newly exposed 3′ OHs in the flanking target DNA results in a co-integrate molecule, in which the donor and target DNAs are joined by two copies of the transposon (FIG. 2a).

Hairpin formation on the transposon ends. DNA transposons that excise themselves from the donor DNA have developed a variety of strategies to cleave the non-transferred strand. Excision of the bacterial elements IS10 and IS50 is achieved through a hairpin intermediate[48,49] (FIG. 2b). The 3′ OHs at the transposon ends attack the phosphate backbone at the 5′ ends of the transposon to form a hairpin, a reaction that is chemically equivalent to strand transfer, except that the target is the complementary strand. The hairpin is resolved by a reiteration of the first chemical step — hydrolysis at the 3′ ends to form an excised linear transposon ready for insertion. The fate of the double-stranded break at the site of excision is unclear. In bacteria, the chromosome could be either degraded, or repaired, if a second copy of the chromosome exists (for example, after replication and before cell division). In eukaryotes, the gap is repaired by NON-HOMOLOGOUS END JOINING, or by gene conversion using the sister chromatid or homologue as a template[50,51].

Hairpin formation on the flanking DNA. Even before the development of *in vitro* systems to study transposition, a model that invokes hairpin formation on the flanking DNA was proposed based on the 'footprints' that are left behind after excision of certain elements[52]. These 'footprints' consisted of small deletions or insertions of DNA at the empty site and were often palindromic in nature and were therefore called P NUCLEOTIDES. These have been described for Tam3 Ac/Ds, Ascot and Hobo elements[53]. It was proposed that nicks were introduced at the 5′ ends of the transposon DNA, thereby generating 3′ OH nucleophiles in the flanking DNA (FIG. 2c). Direct *trans*-esterification by the 3′ OH on the opposing strand would result in a hairpin and the release of a linear transposon. P nucleotides would be formed if opening of the

SERINE RECOMBINASE
Recombination protein (also known as resolvase or invertase) that uses an active-site serine nucleophile to perform strand-cleavage and religation steps through a 5′-phosphoserine intermediate.

CONJUGATIVE TRANSPOSON
A transposon that encodes functions allowing transfer of the transposon DNA between donor and recipient bacterial cells.

RETROTRANSPOSON
A transposon, the movement of which occurs through an RNA intermediate, which is copied into a cDNA molecule.

LTR
(Long terminal repeat).
A directly repeated sequence at each end of a retrovirus or retrotransposon, which is necessary for reverse transcription, integration and transcription.

INTEGRASE
A term used to describe two different protein families that integrate one DNA molecule into another, but by two different mechanisms.
1: a DDE-containing protein of LTR-retrotransposons and retroviruses that catalyses the integration of a cDNA molecule into a target DNA. 2: a member of the tyrosine site-specific recombinases that is responsible for catalysing the integration and excision of DNA, most notably bacteriophage λ.

DNA HYDROLYSIS
Cleavage of a DNA backbone by an activated water molecule, resulting in the addition of an OH group to the 3′ or 5′ position of the ribose.

TRANS-ESTERIFICATION
The direct exchange of an alcohol moiety of an ester for another alcohol. In this case, the OH group on the 3′ end of the transposon or an amino acid (Y or S) is exchanged for the phospho-ester group on a DNA target.

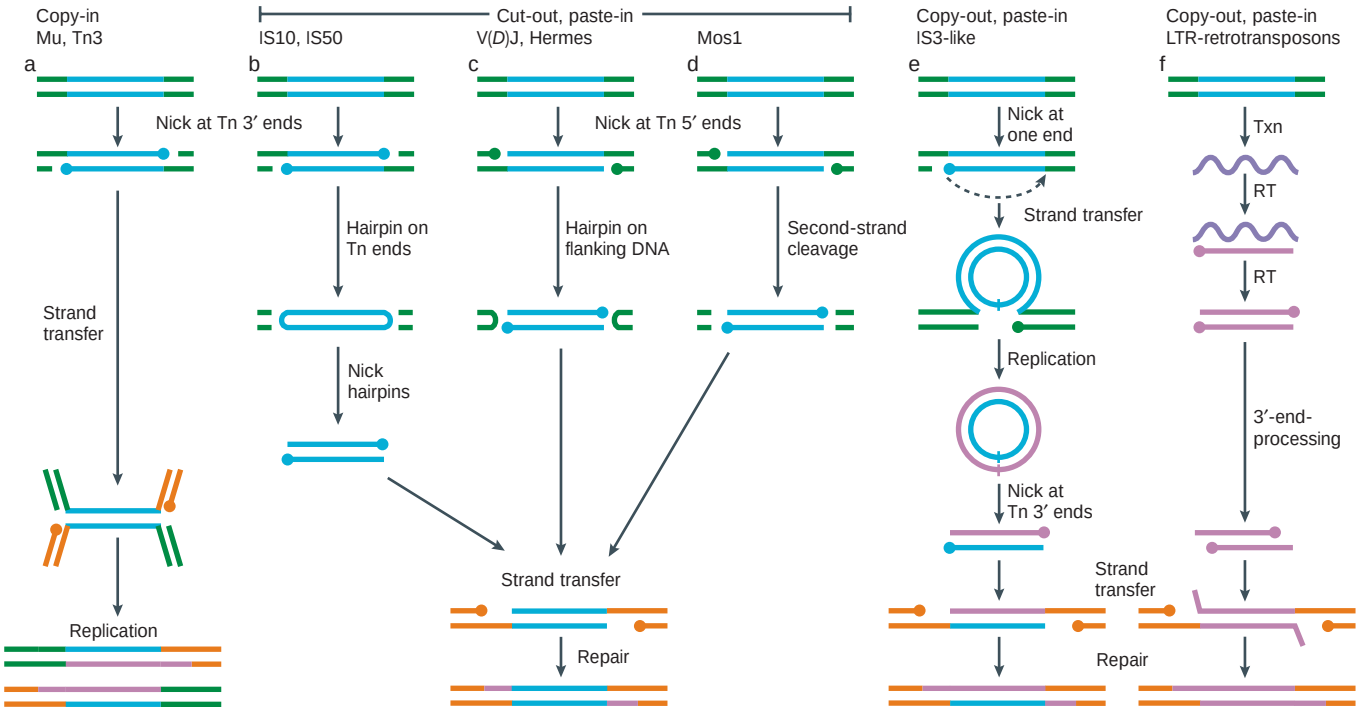


Figure 2 | **DDE-transposases excise their transposon DNA by different mechanisms.** Pathways are drawn in parallel to emphasize the different DDE-transposition pathways. **a** | Mu and Tn3-like transposases nick each transposon (Tn) at the 3′ end and join these to the target. The transposon intermediate is replicated using the 3′ OHs in the target to prime DNA replication. **b–d** | The 'cut-out, paste-in' pathways sever connections with the donor flank and therefore result in a simple insertion. **e** | IS3-like elements only nick at one 3′ end. The resulting 3′ OH attacks the same strand immediately outside the transposon (dotted line with arrow). Replication is thought to resolve this intermediate releasing a transposon circle with the transposon ends abutted and regenerating the donor DNA. A second round of 3′-end nicking generates a linear excised transposon. **f** | LTR-retrotransposons generate a copy of their genome by transcription (Txn), followed by reverse transcription (RT). The 3′ ends of this cDNA either contain a terminal CA dinucleotide, or are processed by the DDE-transposase to expose a CA terminal dinucleotide that is joined to the target. All DDE-transposases insert the transposon between two staggered nicks in the target. Repair of this gap by host-replication enzymes results in a target-site duplication at both transposon ends (pink/orange duplex). Blue lines represent transposon DNA; green lines represent donor DNA; orange lines represent target DNA; pink lines represent newly replicated DNA; purple lines represent RNA; filled circles at the ends of DNA indicate exposed 3′ OH groups.

hairpins was not precise and required repair of the break before religation (BOX 1). Recent biochemical analysis of the Hermes transposon, which is found in insects, has confirmed that this pathway involves hairpin formation and transposon excision using 3′ OH ends that are generated in the flanking DNA (N. L. Craig, personal communication). Quite remarkably, V(*D*)J immunoglobulin-gene rearrangements occur by a similar mechanism involving hairpin formation, P nucleotides and a DDE-related protein. This has led to the proposal that V(*D*)J recombination evolved from a DDE-transposon[54,55] (BOX 1).

Similar to Hermes (and V(*D*)J recombination), cleavage of the Mariner-like element, Mos1, first occurs at the 5′ ends of the transposon, but second-strand cleavage does not seem to occur by a hairpin intermediate[56] (FIG. 2d). Hairpins have not been detected *in vitro*, and efficient and accurate second-strand cleavage is still observed, even in the absence of a 3′ OH group in the flanking DNA (that is, when using a pre-nicked substrate without 3′ OH ends). It remains to be determined how Mos1 performs this second hydrolysis cleavage step, and whether this is true for all related Mariner-like elements. We note that one difference between the prokaryotic and eukaryotic DDE-transposases described so far seems to be the site of hairpin formation (on the transposon or on flanking DNA, respectively). Whether this is a consequence of small sample size or selective evolution will require analysis of other DDE-transposons.

Two proteins release Tn7 from flanking DNA. The bacterial transposon Tn7 requires four proteins for transposition, of which two, TnsA and TnsB, carry out cleavage and strand transfer[57,58]. A second pair of proteins — either TnsC and TnsD, or TnsC and TnsE — is responsible for coordinating transposition and target selection[59,60]. TnsA and TnsB function as a heterodimer and each cleaves a different flanking strand. TnsB is a DDE-transposase, which binds the transposon termini and nicks precisely at the transposon 3′ ends similar to IS10 and IS50 (FIG. 2b). However, no hairpin is involved in second-strand cleavage; instead, TnsA cleaves the transposon at its 5′ ends[58,61], and structural studies show that it most closely resembles a TYPE IIS RESTRICTION ENDONUCLEASE and not a DDE-transposase[62].

TRANSPOSOSOME
A protein–DNA complex that mediates all the steps of transposition, ensuring the fidelity of the reaction. The complex contains both *cis*-acting sites (for example, transposon ends), target and transposase protein (and sometimes other host factors).

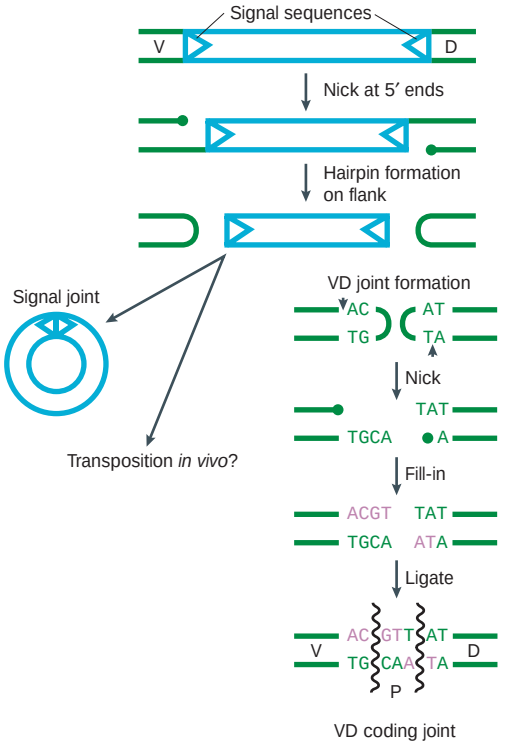
NUCLEOPHILIC ATTACK
A reaction that involves the transfer of electrons from a nucleophile; for example, a hydroxyl group from H_2O or a serine or tyrosine residue.

NON-HOMOLOGOUS END-JOINING
The joining of DNA ends that share no, or only a few, nucleotides of DNA homology.

Box 1 | V(*D*)J recombination and transposition

There are remarkable similarities between V(*D*)J (variable (diversity) joining) recombination and transposition at both the protein and the mechanistic level, which indicates that V(*D*)J recombination evolved from a DDE-transposon[54,55]. The recombination-activating gene (RAG)1 protein contains a DDE motif, which, when mutated, abolishes nicking and hairpin formation during V(*D*)J recombination[119–121]. Together with RAG2, RAG1 recognizes specific binding sites, known as signal sequences, which are formally equivalent to the inverted repeats that are found at the ends of DDE-transposons and flank the immunoglobulin coding regions. The excision of blunt-ended signal sequences (equivalent to an excised transposon) and the formation of a hairpin occurs as in FIG. 2c (REFS 122,123). The hairpins at the coding ends are opened and rejoined by host-encoded proteins[124,125]. The opening is not always precise and, if nicked off-centre, results in the formation of single-strand extensions, which can be filled in by host-repair enzymes and then religated. The nucleotides at the new junction are palindromic (P) with respect to each flank.

To strengthen the transposon connection even further, *in vitro* reactions have shown that the RAG PROTEINS will integrate the released signal sequences into a DNA target in a step that is equivalent to strand transfer. Furthermore, if insertions are coupled, a five-base-pair staggered cleavage occurs in the target similar to DDE-transposition[10,11]. The chemistry of the hydrolysis and *trans*-esterification reactions is identical to that of DDE-transposases. It had been thought that signal-sequence ends were always precisely joined to produce circles *in vivo*, and so prevent deleterious insertions into essential genes. But recently, products that are consistent with intermolecular transposition of signal sequences have been described, which indicates that they can insert like a transposon *in vivo*, although the frequency of these events is unknown[126].



P NUCLEOTIDE
(Palindromic nucleotide). A small, palindromic DNA sequence that is introduced at the site of hairpin resolution during excision of Tam3-related elements and V(*D*)J signal sequences.

TYPE IIS RESTRICTION ENDONUCLEASE
A DNA endonuclease that recognizes an asymmetric sequence and cleaves both DNA strands at fixed positions outside the recognition site.

RAG PROTEIN
Protein product of the recombination-activating genes that are required for V(*D*)J recombination.

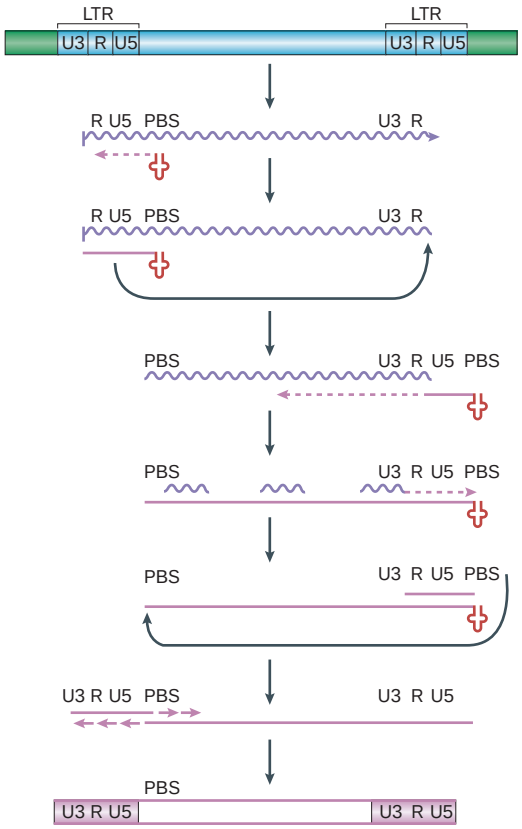
SYNAPTIC COMPLEX
A protein–DNA complex, or transpososome, that contains DNA sites in the correct alignment for effective recombination.

Excision of the IS3 family requires replication. The excision of IS911, IS2, IS3 and probably many other related elements requires host-replicative proteins to facilitate excision, and results in the formation of a circular intermediate[63–65] (FIG. 2e). Unlike other DDE-transposon reactions, the first step is asymmetric and involves the generation of a nick at just one 3′ end of the transposon. This exposed 3′ OH then acts as a nucleophile and directs strand transfer to the phosphodiester backbone of the same strand immediately outside the other transposon end. This targeted strand-transfer reaction is presumably facilitated by the juxtaposition of the two ends that are held together by the transposase in a SYNAPTIC COMPLEX[66]. It is thought that DNA replication that is primed from the nick at the transposon junction resolves this intermediate into an excised transposon circle (and also regenerates the parent molecule). A linear excised transposon that is suitable for insertion is generated from this circle by nicking at each transposon end[67]. Therefore, in contrast to other DDE-transposons in which the excised transposon DNA is generated by multiple cleavages, DNA replication is required to generate an excised linear transposon intermediate that is suitable for insertion into a target.

LTR-retrotransposons 'copy-out and paste-in'. LTR-retrotransposons are structurally and mechanistically related to retroviruses, except that they lack an *env* gene, which renders their RNA-containing virus-like particles (VLPs) non-infectious. Work on the mechanism of LTR-retrotransposition has relied heavily on comparing steps in retrotransposition to known steps in retroviral replication. LTR-retrotransposons encode Gag, a structural protein that binds RNA and forms VLPs, and Pol, which includes three distinct enzymatic activities: an aspartic protease that is required for protein processing, a reverse transcriptase and a DDE-transposase, more commonly known as integrase. Like retroviruses, LTR-retrotransposons generate an extrachromosomal cDNA by reverse transcription of their RNA transcript within a cytoplasmic VLP[68,69] (BOX 2). The cDNA is bound by integrase, forming a pre-integration complex. In retroviruses and some LTR-retrotransposons, the integrase introduces nicks at each 3′ end of the proviral cDNA, releasing the two terminal nucleotides and exposing a 3′-CA dinucleotide, which is conserved among retrotransposons and retroviruses[28,70,71] (FIG. 2f). Other LTR-retrotransposons have the conserved CA dinucleotide at the 3′ end of the cDNA, and integrase catalyses the integration reaction using blunt cDNA ends[72]. The exposed 3′ OH ends are joined to the target, where it is thought that host-repair enzymes, which fill in the single-stranded gaps at the host junctions, also remove the two unpaired 5′ nucleotides at the transposon ends (when present). So, these elements move by a 'copy-out and paste-in' process.

Box 2 | Reverse transcription of LTR-retrotransposon RNA

A terminally redundant RNA, and two critical strand-transfer events during retrotransposition, allow a full-length long terminal repeat (LTR)-retrotransposon cDNA to be synthesized from a terminally truncated transcript. The transcript of LTR-retrotransposons (wavy line), and retroviruses, begins and ends within the LTRs, using transposon-encoded promoter and termination sequences. The RNA contains a region that is repeated at either end (R), a segment of the upstream LTR that is unique to the 5′ end (U5), a segment of the downstream LTR that is unique to the 3′ end (U3), and the primer-binding site (PBS) — which typically hybridizes to the 3′ end of a cellular tRNA (red cloverleaf structure) that primes reverse transcription. The first product of reverse transcription is minus-strand cDNA (pink strand with attached cloverleaf). The RNASE H activity of reverse transcriptase (RT) degrades the complementary RNA, and the elongating cDNA is transferred to the 3′ end of the retrotransposon transcript, hybridizing to the R region. Reverse transcription extends minus-strand cDNA to the end of the RNA template. The RNA strand of the RNA–cDNA hybrid is partially degraded, leaving short RNA fragments, one of which primes synthesis of plus-strand cDNA. Plus-strand synthesis continues into the tRNA primer, copying the PBS. Following RNaseH degradation, the plus-strand cDNA is transferred to the 5′ end of the minus-strand cDNA, hybridizing in the PBS region. Extension from both 3′ ends of the cDNA by RT completes synthesis of the full-length double-stranded cDNA.



Y2-transposons: a 'rolling-circle' mechanism

Rolling-circle or Y2-transposons are typified by the IS91 family of bacterial insertion sequences[25]. The RC-transposases are related to a family of bacterial proteins that are required for plasmid and bacteriophage replication, which occurs by an RC mechanism[73,74] (BOX 3). Indeed, insights from analyses of these related proteins have provided the foundation for understanding the mode of transposition[9]. This family is a relative newcomer to the transposon collection, in part because they are difficult to identify; they lack terminal inverted repeats, the two ends of the transposon are different and they do not generate target duplications. However, it is these properties that add further clues to support models for their mechanism of transposition. One transposon end contains conserved structural elements that are related to those found in the origin region of RC-plasmids and phage. Moreover, a conserved tetranucleotide (5′ CTTG, or 5′ GTTC for IS91) is the site of plasmid- and phage-origin nicking, as well as a preferred transposon insertion site (that is, the site of transposase cleavage). The 3′ ends of IS91 are defined by a related 5′ CTCG sequence at the terminus and a region of twofold symmetry, which resembles terminator sequences of RC-plasmids. The different functions that are performed by each end are highlighted by the fact that the 3′ ends are not essential for transposition; in the absence of a normal 3′ end, longer segments of DNA transpose and terminate at a CTTG (or GTTC) sequence in the flanking DNA[9]. So, it is probably more correct to consider the ends of RC-elements in their functional role as a replication origin (*ori*) and terminus (*ter*).

Although RC-transposases and phage-encoded replication proteins share several conserved amino-acid motifs, the most striking is the absolute conservation of a pair of tyrosines that are separated by three residues[25]. No structural information is available for these proteins — however, some insight is provided from crystallographic studies of the adeno-associated virus (AAV) replicase protein[75]. This protein contains the same conserved tyrosine motif and is proposed to nick DNA, thereby forming 5′ phosphotyrosine intermediates and 3′ OH ends that are suitable for priming viral DNA synthesis[76,77]. The replicase structure shows that both tyrosines are located in the active site in close proximity to a pair of highly conserved histidines, which are responsible for metal-ion coordination. Substitution of the tyrosine residues in the IS91 transposase abolishes transposition[78]. In keeping with using the catalytic motif to define a family, we will call IS91 and related elements Y2-transposons.

Models of Y2-transposition. Two models have been proposed to describe Y2-transposition, which differ primarily in the order of cleavage events observed (that is, whether donor and target cleavages are sequential or concerted; see FIG. 3). Each model involves multiple rounds of tyrosine-mediated cleavage (*trans*-esterification) and

RNASE H
An enzyme that degrades the RNA strand of an RNA–DNA duplex.

Box 3 | Rolling-circle replication

Rolling-circle (RC) replication is a form of DNA replication that is often used by bacterial plasmids and bacteriophage[73,74]. Replication is primed from a 3′ OH that is generated by a single-stranded nick in the plasmid or phage DNA and uses the uncut strand as a template (compare with replacement-strand synthesis in the sequential model of FIG. 3). The site-specific nick is introduced by the replication protein through an active-site tyrosine residue, which becomes covalently attached to the 5′ end of the nick. As the newly synthesized strand is extended, the pre-existing strand is displaced. As the plasmid/phage DNAs are circular, continuous replication around the molecule can result in the generation of multiple tandem copies of a single-stranded DNA if replication is not terminated — hence the name 'rolling-circle' replication. For protein A of the bacteriophage ϕx174, strand cleavage in the origin is catalysed by one of two conserved tyrosine residues and results in a 5′ phosphotyrosine linkage[127,128]. The second tyrosine residue is required for terminating DNA replication and resealing the newly synthesized single strand. Conserved tyrosines in related motifs of RC-transposases, replication proteins and CONJUGAL RELAXASES are thought to perform a comparable role[25,73,129].

DNA replication, primed from a 3′ OH exposed at the cleavage site. The sequential model (FIG. 3a) is more similar to that proposed for replication of bacteriophage ϕX174 and involves excision of a single strand of the transposon to form a circular intermediate[79]. So, cleavage at each end of the transposon occurs before target cleavage. This 'cut-out and copy-in' model is supported by the detection of *in vivo* single-stranded circular DNA intermediates for IS91, the formation of which is dependent on the presence of both conserved tyrosine residues[78].

In the concerted cleavage, or 'copy-in', model (FIG. 3b), a transposase dimer binds to the *ori* end of the transposon and a target CTTG, and nicks to form two 5′ phosphotyrosyl linkages[9,25]. The *ori*-bound transpose then catalyses the first strand-transfer event by joining the 5′ end of the transposon to the 3′ end of the target. Replacement-strand synthesis of the donor transposon, followed by a second tyrosine-mediated cleavage and strand-transfer event at *ter*, results in a single pre-existing strand of the transposon being transferred to the target DNA.

Both models account for the unusual features of this family of transposons. The ends are different, as they perform distinct roles in transposition, initiation and termination. The target specificity is directly linked to cleavage specificity at both ends of the transposon, and there are no target duplications as only one strand of transposon DNA is inserted into a target. The ability to transduce DNA that flanks the 3′ end of the transposon to a new target DNA is due to inefficient *ter* recognition, and, instead, transposase recognises an arbitrary CTTG sequence in the flanking DNA and so extends the segment of transposed DNA.

Eukaryotic Y2 elements. Elements related to Y2-transposons, known as helitrons, have recently been identified by computer-assisted searches in the genomes of *Arabidopsis*, *Caenorhabditis*, rice and *Drosophila*[26,80]. These elements encode Y2-transposases that contain conserved tyrosine and histidine motifs, have a preferred target site (5′ CTRR (where R is a purine)), and do not generate direct repeats on insertion. In contrast to prokaryotic Y2-transposases, the helitron Y2-transposases also encode a carboxy-terminal HELICASE domain, which is presumed to unwind and displace the transposon DNA during replacement-strand synthesis[9] (FIG. 3). Many of these elements have acquired additional host genes, which are found at the 3′ ends of the transposon. This is consistent with the 3′ end acting as a *ter* sequence, which, if not recognized by the transposase, can be substituted by pseudo-*ter* sites in the flanking DNA and thereby allowing 'capture' of flanking genes (as proposed originally for *ter*-defective prokaryotic Y2 elements[9]). A sequence that is related to helitrons has recently been described in maize, except that it does not encode a DNA helicase; instead, it carries a pseudo-gene for a DEAD-BOX RNA helicase[81]. Intriguingly, the genome of *Saccharomyces cerevisiae* harbours a sub-telomeric repeat, coincidentally named Y′, which also encodes an RNA helicase[82]. Y′ elements can be excised as circular DNA molecules and reintegrated, a process which has been implicated in promoting telomeric rearrangements[83]. So, might Y′ sub-telomeric repeats be specialized helitrons?

Y- and S-transposons

Two families of proteins that mediate DNA recombination in prokaryotes are the Y- and S-recombinases, exemplified by λ-integrase and γδ-resolvase, respectively[19,20]. These proteins generally mediate site-specific recombination between two homologous sites. This results in integration if the sites are on different DNAs (for example, *attB* and *attP* of bacteriophage λ) and excision (the reverse of integration) when the sites are on the same molecule[84]. However, several transposons have been described that use these recombinases to mediate transposition to many different, rather than specific, target sites. As we will discuss, the mechanistic steps of transposition (equivalent to integration and excision) are likely to be identical to those of their site-specific cousins, with the most important difference being the ability of Y- and S-transposons to recognize unrelated targets for insertion.

The largest and best-studied group of these elements are the conjugative transposons[21], which use a Y-recombinase (hereafter referred to as Y-transposase to distinguish between site-specific recombination and transposition), such as Tn916, although a few encode S-transposases, such as Tn5397 (REF. 85). The conjugative transposons are so-called because they encode transfer functions that allow the excised transposon to be transferred from a donor to a recipient bacterium. A word of caution here: the literature and databases are full of alleged conjugative transposons (also known as CONSTINs or ICEs[22,86]). However, most of these integrate into specific sites and so do not fit the strict definition of a transposon. Indeed, even Tn916 and Tn5397 exhibit site-specific insertion in some bacterial hosts and not others, thereby adding to the confusion. Moreover, there are examples of these transposons that are not conjugative (for example, IS607; REF. 87) and it is assumed that there will be more. We will focus on the mechanism of transposition of those elements that use multiple target sites and will designate each family depending on its active-site nucleophile — Y or S.

CONJUGAL RELAXASE
A protein that is responsible for initiating DNA transfer in bacteria. It contains a conserved tyrosine motif that is required to nick the transferred DNA strand at *oriT*, which it forms a 5′-phosphotyrosine bond with.

HELICASE
An enzyme that promotes ATP-dependent disassociation of the complementary DNA or RNA strands of a duplex molecule.

DEAD BOX
An evolutionarily conserved array of amino acids, also known as the Walker B motif, that confers ATP-binding activity.

Y-transposons cleave DNA strands sequentially. Consistent with a mechanism that is similar to site-specific recombination[18,41], Y- or conjugative transposons excise precisely from a target and regenerate the original target site. They generate a circular intermediate and, following insertion, do not generate a target duplication[21]. In addition to the Y-transposase, many elements also encode a small basic protein, Xis, which is required for efficient excision. Xis is thought to have a structural role in facilitating synapsis of the transposon ends, based on the role of λXis in λ excision[19,88]. Following synapsis, excision is initiated by cleavages on the same DNA strand at each transposon end by the active-site tyrosine nucleophile (FIG. 4a). This results in the generation of a 3′ phosphotyrosine linkage and an exposed 5′ OH. The exposed 5′ OH then attacks the phosphotyrosine linkage at the opposite end to promote strand exchange. This process is repeated on the opposite strand (that is, strand cleavage is sequential), but the cleavages are staggered, to complete excision and regeneration of the target site[88]. Integration is thought to occur by a reverse process after target capture.

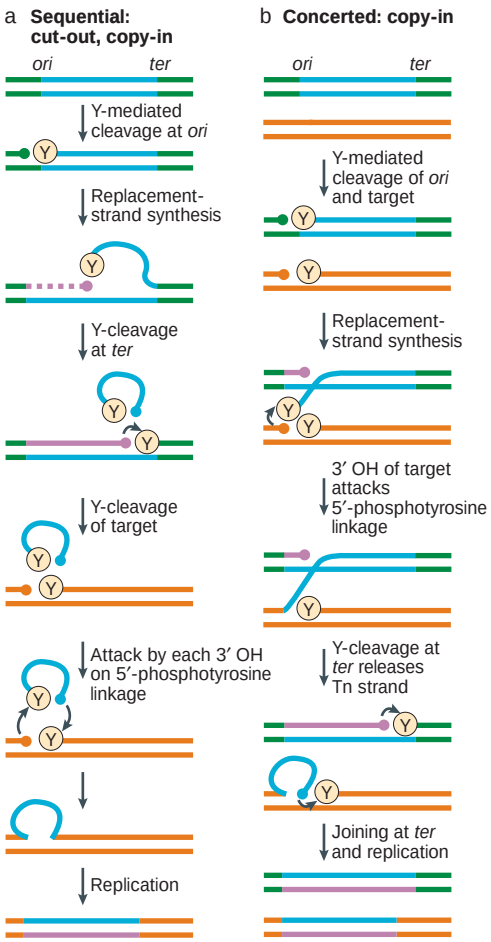


Figure 3 | **Y2-transposases form phosphotyrosine linkages during transposition.** Two models for Y2-transposition are outlined. **a** | In the sequential pathway, the transposase cleaves the *ori* end of the transposon forming a 5′-phosphotyrosine linkage (Y in circle). The 3′ OH of the flanking DNA is used to prime replacement-strand synthesis (pink dotted line). The Y-mediated cleavage at *ter* results in transposon excision. A nucleophilic attack by the 3′-OH group of the nascent DNA strand on the phosphotyrosine linkage at *ter* is thought to reseal the donor DNA (indicated by arrow). A second tyrosine cleavage nicks the target DNA adjacent to the preferred target sequence, and the attack by each exposed 3′ OH on the 5′-phosphotyrosine linkage on the opposite strand of DNA results in insertion. Passive DNA replication by the host completes the process, and therefore each transposon contains one newly synthesized strand. **b** | In the concerted pathway, the 5′ ends at the *ori* and the target are cleaved together. The transposon strand is joined to the 3′ OH at the target and is displaced by DNA synthesis from the flanking donor DNA. Cleavage at *ter* releases the transposon strand. Each junction is resealed by an attack on the phosphotyrosine linkage by the exposed 3′ OH. The transposon strand is copied in the target by host replication. Blue lines represent transposon DNA; green lines represent donor DNA; orange lines represent target DNA; pink lines represent newly replicated DNA; filled circles represent exposed 3′ OH groups.

An important difference between the reaction that is mediated by the site-specific Y-recombinases and the Y-transposases is that the junctions formed in the circular intermediates by Y-transposases can accommodate mismatches. For the site-specific Y-recombinases, the staggered cleavages occur in a region of homology (the core) and therefore the two strands are complementary. Base-pairing within this region is crucial for strand exchange, as it has been shown that mismatches introduced within the cross-over sites are sufficient to block strand joining[89,90]. By contrast, the equivalent core regions of Tn916 (known as coupling sequences), found at the transposon flanks, are not necessarily complementary and therefore might contain mismatches, but still allow strand exchange for both excision and integration[91,92]. The lack of a requirement for homology of the coupling sequences allows these elements to insert into and excise out of nonhomologous sites, and so contributes to their relaxed target specificity and their classification as transposons. How these Y-transposases accommodate these mismatches within their active sites will clearly require more detailed mechanistic and structural studies. Such studies will also reveal which components of the process contribute to the relaxed specificity of target capture.

Finally, we note that putative open-reading frames that encode Y-recombinases have recently been identified in Tec elements of the ciliate *Euplotes crassus*[93,94]. These elements are precisely removed from the micronucleus during differentiation into a macronucleus. This process involves the formation of excised, circular transposons that contain mismatched DNA at their junctions, the sequence of which is derived from the original transposon flanks[93,95]. The mechanistic similarity to Y-transposons indicates that the Tec-encoded Y-recombinase mediates this process.

Transposons that encode an S-transposase. A few bacterial transposons encode a protein that is closely related to the site-specific S-recombinases[87,96,97]. The S-recombinases (for example, γδ-resolvase and Hin invertase) carry out recombination between a pair of identical sites in which a serine nucleophile is used to generate strand cleavages[20,98]. Each crossover site is recognized by a dimer of proteins. Recombination occurs

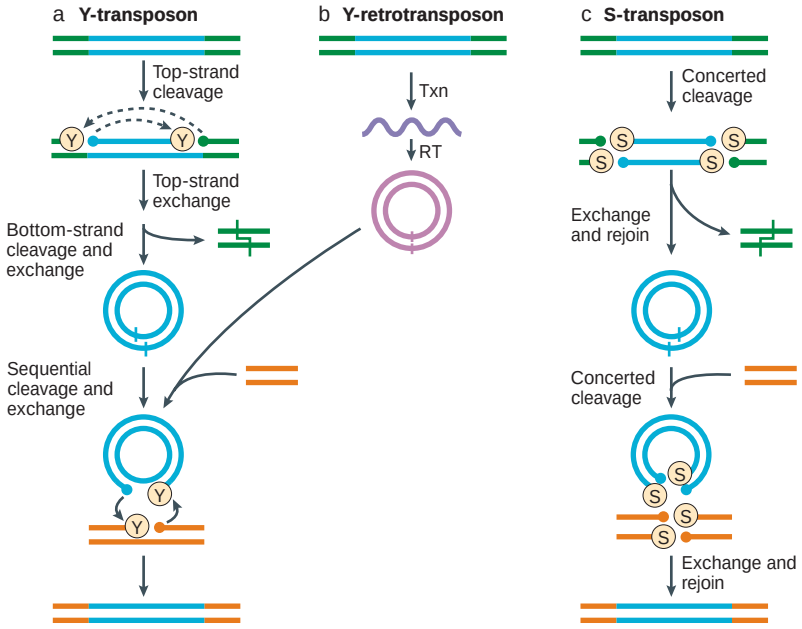


Figure 4 | **Tyrosine and serine transposition involves a series of protein–DNA covalent intermediates. a** | Tyrosine (Y)-transposases nick each end of one transposon strand forming 3′ phosphotyrosine intermediates. The exposed 5′ OHs attack the phosphotyrosine linkage at the opposite end (dotted lines with arrows) to circularize one transposon strand and join the two flanking DNAs. This is repeated on the bottom strand to regenerate the host DNA and a circular intermediate. Nicking occurs with a stagger such that the circle junctions might contain mismatches (which are thought to be resolved by DNA replication). Following target capture, the sequential strand cleavage and joining process is repeated and results in an insertion. **b** | The Y-retrotransposons have been proposed to generate a circular cDNA of the element by reverse transcription (RT), and then integrate this circular intermediate by a process that is identical to Y-transposons. **c** | Serine (S)-transposition occurs through concerted double-stranded breaks at each transposon end, which results in 5′ phosphoserine linkages. A concerted nucleophilic attack by the 3′ OHs on the phosphoserine linkages at the other junction results in a circular intermediate and joins the original flanking DNA. The circular intermediate is inserted into a target by a repeat of the above process. The cleavages result in a two-base-pair 3′ overhang. Blue lines represent transposon DNA; green lines represent donor DNA; orange lines represent target DNA; pink lines represent newly replicated DNA; purple lines represent RNA; S in circle, 5′-phosphoserine; Txn, transcription; Y in circle, 3′-phosphotyrosine.

through a concerted four-strand cleavage (in contrast to the Y-recombinases), strand switching and rejoining event (FIG. 4c). Staggered cleavages at each crossover site result in a two-base-pair 3′overhang with the catalytic serines linked to the 5′ phosphate ends. After strand switching, the exposed 3′ OH residues attack each phosphoserine linkage to rejoin the four DNA strands.

Little is known regarding the precise mechanism of the S-transposons, although the similarity of their transposases to S-recombinases and the nature of the transposition products indicate that these elements transpose by an almost identical mechanism. A circular intermediate has been detected for three S-transposons — Tn5397, Tn4451 and IS607 (REFS 85,99; and N.D.F. Grindley, personal communication). These excised circles have the predicted transposon joints with a two-base-pair coupling sequence and regenerate the original target. There is a strong preference for complementarity within the coupling sequence, which is also consistent with mismatches in the crossover site inhibiting site-specific recombination[96,100]. Although these transposases must recognize specific sequences at their transposon ends to allow synapsis and excision, the variable nature of their target sites indicates that they recognize a target (flanking a conserved dinucleotide) with relaxed specificity.

DIRS1 retrotransposons encode a Y-transposase. Given the abundance of Y- and S-recombinases in prokaryotes (both site-specific- and transposon-recombinases) and the myriad of rearrangements they mediate, it is surprising that so few mechanistically similar elements from eukaryotes have been described (which is why we anticipate many more). Intriguingly, a family of retroelements in eukaryotes, the DIRS1-like family, encodes both a reverse transcriptase and a Y-transposase and integrates without generating target-site duplications[23,24] — therefore, we propose that they be named Y-retrotransposons. Although originally thought to be a variant type of LTR-retrotransposon because of homology between the reverse transcriptase domains in the two families, Y-retrotransposons lack the characteristic aspartic proteases and DDE-integrases of LTR-retrotransposons. Furthermore, they lack directly repeated LTR sequences, and instead have either inverted terminal repeats (DIRS1) or 'split' direct repeats (PAT, Kangaroo). These repeats are likely to have dual functions — namely, to produce full-length cDNA from terminally truncated RNAs and serve as recombination sites for the Y-transposase. Although the mechanism of retrotransposition of this family has not been determined experimentally, molecules that are consistent with a closed circular DNA of the highly mobile element, Kangaroo, have been detected in *Volvox carteri*[24]. This fact, combined with the lack of target-site duplications and the presence of a Y-recombinase, indicates that these elements transpose by a mechanism that resembles that used by Y-transposons — namely, that a closed circular DNA intermediate is generated by reverse transcription and integrated by Y-mediated recombination (FIG. 4b). Related elements have been found in slime mold, green algae, nematodes and fish, which indicates that this new family of retrotransposons is only beginning to be appreciated.

TP-retrotransposons: combining RT/En activities

The fifth class of transposon — TP-retrotransposons — are more ancient and structurally more simple than LTR-retrotransposons[27]. They are typified by the mammalian LINE-1, or L1, elements, the most abundant and active transposable elements in the human genome (500,000 copies)[101]. They lack terminal repeats altogether, but often have a polyA or A-rich sequence at their 3′ end. At least one open-reading frame is present, which encodes a protein with at least two distinct activities: reverse transcriptase and endonuclease. In some TP-retrotransposons, a second open-reading frame with nucleic-acid-binding activity is present. In mouse L1 elements, the nucleic-acid-binding domain also has chaperone activity, which is thought to facilitate formation of the DNA-primer–RNA-template complex for the initiation of reverse transcription[102].

LINE
(Long interspersed nuclear element). A retrotransposon, the mobility of which is dependent on target-primed reverse transcription.

GROUP II INTRON
An autocatalytic intron from one of two families that catalyses its own splicing from an RNA transcript and encodes a protein that mediates its mobility as a DNA element. Group II introns are the likely progenitors of spliceosomal introns.

Three endonuclease-domain families have been described so far, including apurinic-apyrimidinic endonucleases (APE; found in L1, I and R1 elements[103,104]), site-specific endonucleases that resemble type II restriction endonucleases (found in R2Bm[105]) and a third, recently identified class of GIY-YIG endonucleases (found in Penelope-like elements, a phylogenetically separate class of presumed TP-retrotransposons[106]). A complex of full-length RNA and RT/En protein initiates transposition when the endonuclease activity cleaves the bottom strand of the target, generating a 3′ OH that serves as the primer for reverse transcription of the retrotransposon RNA[30,103] (FIG. 5a). Importantly, studies of human L1 target-primed reverse transcription both *in vitro* and *in vivo* have demonstrated that pre-existing DNA nicks and double-stranded breaks can serve as targets, indicating that reverse transcription is not necessarily coupled to endonuclease-mediated nicking of the target[31,107]. Reverse transcription of TP-retrotransposons rarely proceeds all the way to the 5′ end of the RNA, which accounts for the fact that most members of LINE-like retrotransposon families are truncated at the 5′ end.

The endonuclease of TP-retrotransposons[30], or possibly a host-encoded activity[31], cleaves the top strand of the target, which generates a double-stranded break. Often, top-strand cleavage occurs downstream of the lower-strand nick, which creates a staggered cut that is repaired to create a variable target-site duplication[27]. Many questions remain about how the cDNA is attached to the upstream end of the target and how second-strand synthesis occurs. One currently favoured model suggests that the top strand of the target anneals to the first-strand cDNA through microhomology, and the 3′ OH of the top strand primes second-strand cDNA synthesis using the first-strand cDNA as a template[102,108]. In a second model, the reverse transcriptase switches templates, or 'jumps', from the RNA to the top strand of the target DNA, where it continues synthesis of the first-strand cDNA[109]. Second-strand synthesis would then occur by reverse transcriptase or host-repair enzymes.

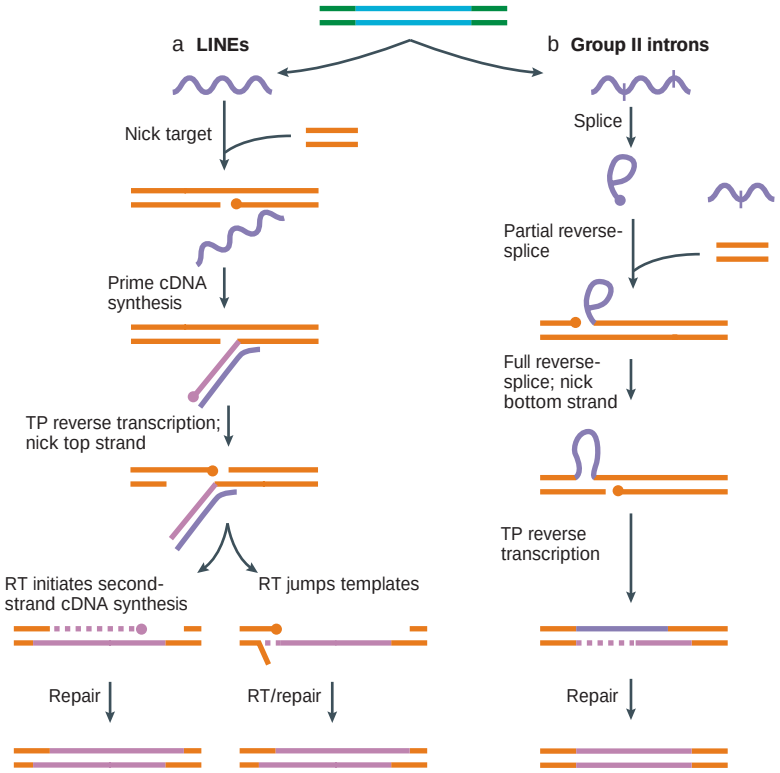


Figure 5 | **TP-retrotransposons insert by reverse-transcribing RNA into a target.** Target-primed (TP) retrotransposons include LINE-like elements and group II introns. **a** | LINE elements encode an endonuclease, which makes a nick in the target DNA. The resulting 3′ OH primes cDNA synthesis, which uses the retrotransposon RNA as a template. The top strand is cleaved by the endonuclease or by a host enzyme. The nascent cDNA anneals to the top strand of the target by microhomologies, allowing the top strand to act as a primer for second-strand cDNA synthesis. The gaps are filled by host enzymes. Alternatively, reverse transcriptase (RT) might switch templates from the RNA to the top strand of the target, continuing first-strand cDNA synthesis. Either RT or host-repair enzymes synthesize the second strand and seal the gap. **b** | Group II introns splice themselves out of a pre-mRNA. The 3′ OH of the resulting lariat RNA attacks one strand of the target DNA (*trans*-esterification), leaving a 3′ OH in the target DNA. The 3′ OH of the target attacks the lariat branch point, which results in complete reverse-splicing of the intron RNA into the top strand of the target. A nick that is made downstream of the intron insertion site by the endonuclease primes cDNA synthesis, using the intron RNA as a template. Second-strand synthesis by RT or host-repair functions displaces the RNA and reseals the gap. Blue lines represent transposon DNA; green lines represent donor DNA; orange lines represent target DNA; pink lines represent newly replicated DNA; purple lines represent RNA; filled circles represent exposed 3′ OH groups.

Group II introns reverse-splice into DNA. A fascinating class of mobile elements are the GROUP II INTRONS, which are found primarily in bacteria and eukaryotic organelles. Group II introns are mobile, autocatalytic introns capable of both retrohoming (movement to a cognate DNA allele that lacks the intron) and TP-retrotransposition[110,111]. Like the LINE-like elements, group II introns encode reverse transcriptase and endonuclease activities. But their target site is chosen by a different and remarkable mechanism: intron RNA, which has excised itself out of pre-mRNA, reverse-splices itself into DNA, which results in the covalent attachment of each end of the intron RNA to one strand in the DNA target[112,113] (FIG. 5b). Target specificity for group II introns is primarily determined by RNA–DNA base-pairing interactions, although the endonuclease does show some specificity for nucleotides around the cleavage site. Once the intron is inserted, the intron-encoded endonuclease cleaves the bottom strand of the DNA slightly downstream of the intron insertion site. The resulting 3′ OH is used as a primer for reverse transcriptase to copy the top-strand RNA into a bottom-stand cDNA[114,115]. As with LINE-like elements, it is not clear whether synthesis of the second-strand cDNA is carried out by the reverse transcriptase or by host-repair enzymes. Interestingly, the endonuclease is essential for retrohoming but dispensable for retrotransposition[116], which indicates that the 3′ OH that serves as a primer for reverse transcriptase might be generated by a different mechanism during retrotransposition, possibly by random nicks or by DNA replication.

Telomeres are synthesized by TP reverse transcription. Target-primed reverse transcription is essential for the maintenance of most eukaryotic genomes, as it is the mechanism by which telomeric DNA, which consists of

short tandem DNA repeats at the ends of linear chromosomes, are synthesized. Telomerase, a reverse transcriptase that is related to TP-retrotransposon reverse transcriptases, synthesizes telomeric DNA using the 3′ OH at the end of the chromosome as a primer. A remarkable exception to this standard eukaryotic mechanism is found in *Drosophila*, which lacks telomerase and simple telomeric repeats. Instead, *Drosophila* chromosomal ends are maintained by repeated transposition of the specialized TP-retrotransposons, HeT-A and TART[117]. Although the mechanistic parallels between TP-retrotransposons and telomerase are clear — and in *Drosophila* this even extends to a functional parallel — it is not known if they both evolved from an ancient progenitor or whether one is the progenitor of the other[13,14].

Concluding remarks

Transposon biology has come a long way since the elucidation of the DNA double helix. Transposable elements are far more diverse and abundant than previously imagined, and new forms of transposons will inevitably be discovered as more genomes are analysed. It is especially striking how different 'blends' of transposon have evolved as a consequence of shuffling transposon-associated catalytic activities (even between kingdoms). We note in particular how retrotransposons can be divided into three mechanistic subgroups depending on which other transposon-associated activity they have acquired (DDE, Y or En).

Who would have predicted Y-retrotransposons? With only a little hindsight, we might predict that S-retrotransposons will be described (see FIG. 4). What other combinations of activity might we anticipate to expand the mechanistic repertoire of how to move 'out and in' to a new target? Recently, an LAGLIDADG endonuclease and a GIY-YIG endonuclease, which are normally associated with mobile group I introns, were identified in group II introns[118] and Penelope retrotransposons[106], respectively. These discoveries add further fuel to the transposon shuffle and hint that novel endonucleases will be found in future TP-retrotransposons. The presence of type IIS endonucleases in both Tn7 and TP-retrotransposons[62,105] further strengthens the argument that transposons are capable of acquiring, and presumably disseminating, functions that are necessary for strand-breakage and joining. Consequently, we anticipate that novel host processes involving DNA rearrangements will find their roots buried in the sequence of our favourite transposons.

1. Watson, J. D. & Crick, F. H. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Waterston, R. H. *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
4. Goff, S. A. *et al.* A draft sequence of the rice genome (*Oryza sativa L*. ssp. *japonica*). *Science* **296**, 92–100 (2002).
5. Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa L*. ssp. *indica*). *Science* **296**, 79–92 (2002).
6. Berg, D. E. & Howe, M. M. *Mobile DNA* (American Society for Microbiology, Washington DC, 1989).
7. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. *Mobile DNA II*, (American Society for Microbiology, Washington DC, 2002).
   This compilation of reviews and its parent (reference 6) are the 'bibles' of transposition. We refer interested readers to these books for more detailed descriptions of the many different types of mobile element.
8. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H. Jr. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530–1534 (1999).
9. Mendiola, M. V., Bernales, I. & de la Cruz, F. Differential roles of the transposon termini in IS91 transposition. *Proc. Natl Acad. Sci. USA* **91**, 1922–1926 (1994).
   A model for Y2-transposition is presented based on the similarity of these elements and their transposases to rolling-circle replication systems that are found in certain bacterial plasmids and phage (see reference 79 for an alternative model).
10. Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
11. Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
    References 10 and 11 describe *in vitro* transposition of signal sequences that are mediated by the V(*D*)J recombination proteins and provide evidence for the evolution of the immune system from a transposon.
12. Sharp, P. A. "Five easy pieces". *Science* **254**, 663 (1991).
13. Eickbush, T. H. Telomerase and retrotransposons: which came first? *Science* **277**, 911–912 (1997).
14. Nakamura, T. M. & Cech, T. R. Reversing time: origin of telomerase. *Cell* **92**, 587–590 (1998).
15. McClintock, B. Mutable loci in maize. *Carnegie Institute Washington Year Book* **47**, 155–169 (1948).
16. Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. & Skalka, A. M. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* **12**, 2331–2338 (1992).
17. Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53**, 245–281 (1999).
18. Van Duyne, G. D. in *Mobile DNA II* Vol. II (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 93–117 (American Society for Microbiology, Washington DC, 2002).
19. Azaro, M. A. & Landy, A. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 118–148 (American Society for Microbiology Press, Washington DC, 2002).
20. Grindley, N. D. F in *Mobile DNA II* Vol. II (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 272–302 (American Society for Microbiology, Washington DC, 2002).
21. Churchward, G. in *Mobile DNA II* Vol. II (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 177–191 (American Society for Microbiology, Washington DC, 2002).
22. Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.* **46**, 601–610 (2002).
23. Goodwin, T. J. & Poulter, R. T. The DIRS1 group of retrotransposons. *Mol. Biol. Evol.* **18**, 2067–2082 (2001).
24. Duncan, L., Bouckaert, K., Yeh, F. & Kirk, D. L. Kangaroo, a mobile element from *Volvox carteri*, is a member of a newly recognized third class of retrotransposons. *Genetics* **162**, 1617–1630 (2002).
25. Garcillan-Barcia, M. P., Bernales, I., Mendiola, M. V. & de la Cruz, F. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 891–904 (American Society for Microbiology, Washington DC, 2002).
26. Kapitonov, V. V. & Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **98**, 8714–8719 (2001).
    The discovery of helitrons by a computational approach is reported, and a model for their mechanism of mobility is presented based on reference 9.
27. Eickbush, T. H. & Malik, H. S. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 1111–1144 (American Society for Microbiology, Washington DC, 2002).
28. Craigie, R. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 613–630 (American Society for Microbiology, Washington DC, 2002).
29. Cappello, J., Handelsman, K. & Lodish, H. F. Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell* **43**, 105–115 (1985).
    The initial description of the DIRS1 transposon of *Dictyostelium discoideum* as a retrotransposon and a model for its mechanism of reverse transcription and transposition through a circular cDNA intermediate are presented.
30. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
    The mechanism of TP reverse transcription was determined by *in vitro* analysis of the endonuclease and reverse transcriptase activities of the RT/En protein encoded by the R2Bm element of *Bombyx mori*.
31. Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.* **21**, 5899–5910 (2002).
32. Chandler, M. & Mahillon, J. in *Mobile DNA II* Vol. II (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 305–366 (American Society for Microbiology, Washington DC, 2002).
33. Rezsohazy, R., Hallet, B., Delcour, J. & Mahillon, J. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* **9**, 1283–1295 (1993).
34. Rice, P. & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209–220 (1995).
    References 34–40 describe and compare the structures of several DDE-transposases and integrases, highlighting their structural similarity despite their sequence diversity.
35. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. & Rayment, I. The three-dimensional structure of the Tn5 synaptic complex intermediate. Science **289,** 77–85 (2000).
    This study presents the first structure of a synaptic complex between the transposase and excised transposon ends.
36. Bujacz, G. *et al.* High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **253**, 333–346 (1995).
37. Bujacz, G., Alexandratos, J., Qing, Z. L., Clement-Mella, C. & Wlodawer, A. The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Lett.* **398**, 175–178 (1996).

38. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
39. Bujacz, G. *et al.* Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.* **272**, 18161–18168 (1997).
40. Lovell, S., Goryshin, I. Y., Reznikoff, W. R. & Rayment, I. Two-metal active-site binding of a Tn5 transposase synaptic complex. *Nature Struct. Biol.* **9**, 278–281 (2002).
41. Mizuuchi, K. & Baker, T. A. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 12–23 (American Society for Microbiology, Washington DC, 2002).
42. Mizuuchi, M., Baker, T. A. & Mizuuchi, K. Assembly of phage Mu transpososomes: cooperative transitions assisted by protein and DNA scaffolds. *Cell* **83**, 375–385 (1995).
43. Surette, M. G., Buch, S. J. & Chaconas, G. Transpososomes: stable protein–DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell* **49**, 253–262 (1987).
44. Turlan, C. & Chandler, M. Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.* **8**, 268–274 (2000).
45. Craigie, R. & Mizuuchi, K. Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell* **41**, 867–876 (1985).
46. Mizuuchi, K. Mechanism of transposition of bacteriophage Mu: polarity of the strand transfer reaction at the initiation of transposition. *Cell* **39**, 395–404 (1984).
This classic paper and other papers by Mizuuchi and colleagues describe the establishment of the first *in vitro* transposition system. Importantly, they assign the polarity of strand transfer, identify and characterize the strand-transfer intermediate and define the chemistry of the reaction.
47. Shapiro, J. A. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc. Natl Acad. Sci. USA* **76**, 1933–1937 (1979).
48. Bhasin, A., Goryshin, I. Y. & Reznikoff, W. S. Hairpin formation in Tn5 transposition. *J. Biol. Chem.* **274**, 37021–37029 (1999).
49. Kennedy, A. K., Guhathakurta, A., Kleckner, N. & Haniford, D. B. Tn10 transposition via a DNA hairpin intermediate. *Cell* **95**, 125–134 (1998).
The first demonstration of the involvement of hairpins in the excision of transposons from their flanking donor DNA (see also reference 48). The mechanism explains how a single-protein active site can cleave multiple DNA strands by using reiterative steps of hydrolysis and *trans*-esterification.
50. Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. & Sved, J. High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* **62**, 515–525 (1990).
51. Plasterk, R. H. & Groenen, J. T. Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. *EMBO J.* **11**, 287–290 (1992).
52. Coen, E. S., Carpenter, R. & Martin, C. Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47**, 285–296 (1986).
53. Kunze, R. & Weil, C. F. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 565–610 (American Society for Microbiology, Washington DC, 2002).
54. Roth, D. B. & Craig, N. L. V(D)J recombination: a transposase goes to work. *Cell* **94**, 411–414 (1998).
55. Gellert, M. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 705–729 (American Society for Microbiology, Washington DC, 2002).
56. Dawson, A. & Finnegan, D. J. Excision of the *Drosophila* mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. *Mol. Cell.* **11**, 225–235 (2003).
57. Biery, M. C., Stewart, F. J., Stellwagen, A. E., Raleigh, E. A. & Craig, N. L. A simple *in vitro* Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucl. Acids Res.* **28**, 1067–1077 (2000).
58. Sarnovsky, R. J., May, E. W. & Craig, N. L. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996).
59. Waddell, C. S. & Craig, N. L. Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. *Genes Dev.* **2**, 137–149 (1988).
60. Craig, N. L. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 423–456 (American Society for Microbiology, Washington DC, 2002).
61. May, E. W. & Craig, N. L. Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996).
62. Hickman, A. B. *et al.* Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol. Cell* **5**, 1025–1034 (2000).
63. Rousseau, P. *et al.* in *Mobile DNA II* Vol. II (eds. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 367–383 (American Society for Microbiology, Washington DC, 2002).
64. Sekine, Y., Aihara, K. & Ohtsubo, E. Linearization and transposition of circular molecules of insertion sequence IS3. *J. Mol. Biol.* **294**, 21–34 (1999).
65. Lewis, L. A. & Grindley, N. D. Two abundant intramolecular transposition products, resulting from reactions initiated at a single end, suggest that IS2 transposes by an unconventional pathway. *Mol. Microbiol.* **25**, 517–529 (1997).
66. Polard, P. & Chandler, M. An *in vivo* transposase-catalyzed single-stranded DNA circularization reaction. *Genes Dev.* **9**, 2846–2858 (1995).
67. Ton-Hoang, B., Polard, P., Haren, L., Turlan, C. & Chandler, M. IS911 transposon circles give rise to linear forms that can undergo integration *in vitro*. *Mol. Microbiol.* **32**, 617–627 (1999).
68. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. Ty elements transpose through an RNA intermediate. *Cell* **40**, 491–500 (1985).
69. Garfinkel, D. J., Boeke, J. D. & Fink, G. R. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42**, 507–517 (1985).
References 68 and 69 represent the first demonstration that a retrotransposon moves through an RNA intermediate. Moreover, reverse transcription of Ty1 RNA was shown to occur in cytoplasmic virus-like particles.
70. Fujiwara, T. & Mizuuchi, K. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**, 497–504 (1988).
71. Kirchner, J. & Sandmeyer, S. B. Ty3 integrase mutants defective in reverse transcription or 3′-end processing of extrachromosomal Ty3 DNA. *J. Virol.* **70**, 4737–4747 (1996).
72. Eichinger, D. J. & Boeke, J. D. A specific terminal structure is required for Ty1 transposition. *Genes Dev.* **4**, 324–330 (1990).
73. Kahn, S. A. Plasmid rolling-circle replication: recent developments. *Mol. Microbiol.* **37**, 477–484 (2000).
74. Novick, R. P. Contrasting lifestyles of rolling-circle phages and plasmids. *Trends Biochem. Sci.* **23**, 434–438 (1998).
75. Hickman, A. B., Ronning, D. R., Kotin, R. M. & Dyda, F. Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell* **10**, 327–337 (2002).
76. Smith, R. H. & Kotin, R. M. An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. *J. Virol.* **74**, 3122–3129 (2000).
77. Smith, R. H. & Kotin, R. M. in *Mobile DNA II*, Vol. II (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 905–926 (American Society for Microbiology, Washington DC, 2002).
78. Garcillan-Barcia, M. P., Bernales, I., Mendiola, M. V. & de la Cruz, F. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol. Microbiol.* **39**, 494–501 (2001).
79. Tavakoli, N. *et al.* IS1294, a DNA element that transposes by RC transposition. *Plasmid* **44**, 66–84 (2000).
80. Kapitonov, V. V. & Jurka, J. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl Acad. Sci. USA* **100**, 6569–6574 (2003).
81. Lal, S. K., Giroux, M. J., Brendel, V., Vallejos, C. E. & Hannah, L. C. The maize genome contains a helitron insertion. *Plant Cell* **15**, 381–391 (2003).
82. Louis, E. J. & Haber, J. E. The structure and evolution of subtelomeric Y′ repeats in *Saccharomyces cerevisiae*. *Genetics* **131**, 559–574 (1992).
83. Horowitz, H. & Haber, J. E. Identification of autonomously replicating circular subtelomeric Y′ elements in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **5**, 2369–2380 (1985).
84. Nash, H. A. in Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology* 2nd edn (eds Neidhardt, F. C. *et al.*) 2363–2376 (American Society for Microbiology, Washington DC, 1996).
85. Wang, H. *et al.* Characterization of the ends and target sites of the novel conjugative transposon Tn5397 from *Clostridium difficile*: excision and circularization is mediated by the large resolvase, TndX. *J. Bacteriol.* **182**, 3775–3783 (2000).
86. Hochhut, B. & Waldor, M. K. Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol. Microbiol.* **32**, 99–110 (1999).
87. Kersulyte, D., Mukhopadhyay, A. K., Shirai, M., Nakazawa, T. & Berg, D. E. Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. *J. Bacteriol.* **182**, 5300–5308 (2000).
88. Rudy, C., Taylor, K. L., Hinerfeld, D., Scott, J. R. & Churchward, G. Excision of a conjugative transposon *in vitro* by the Int and Xis proteins of Tn916. *Nucl. Acids Res.* **25**, 4061–4066 (1997).
89. Kitts, P. A. & Nash, H. A. Homology-dependent interactions in phage λ site-specific recombination. *Nature* **329**, 346–348 (1987).
90. Nunes-Duby, S. E., Yu, D. & Landy, A. Sensing homology at the strand-swapping step in λ excisive recombination. *J. Mol. Biol.* **272**, 493–508 (1997).
91. Caparon, M. G. & Scott, J. R. Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. *Cell* **59**, 1027–1034 (1989).
92. Trieu-Cuot, P., Poyart-Salmeron, C., Carlier, C. & Courvalin, P. Sequence requirements for target activity in site-specific recombination mediated by the Int protein of transposon Tn1545. *Mol. Microbiol.* **8**, 179–185 (1993).
93. Jacobs, M. E., Sanchez-Blanco, A., Katz, L. A. & Klobutcher, L. A. Tec3, a new developmentally eliminated dna element in *Euplotes crassus*. *Eukaryot. Cell* **2**, 103–114 (2003).
94. Doak, T. G., Witherspoon, D. J., Jahn, C. L. & Herrick, G. Selection on the genes of *Euplotes crassus* Tec1 and Tec2 transposons: evolutionary appearance of a programmed frameshift in a Tec2 gene encoding a tyrosine family site-specific recombinase. *Eukaryot. Cell* **2**, 95–102 (2003).
95. Klobutcher, L. A., Turner, L. R. & LaPlante, J. Circular forms of developmentally excised DNA in *Euplotes crassus* have a heteroduplex junction. *Genes Dev.* **7**, 84–94 (1993).
96. Crellin, P. K. & Rood, J. I. The resolvase/invertase domain of the site-specific recombinase TnpX is functional and recognizes a target sequence that resembles the junction of the circular form of the *Clostridium perfringens* transposon Tn4451. *J. Bacteriol.* **179**, 5148–5156 (1997).
97. Wang, H. & Mullany, P. The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn5397. *J. Bacteriol.* **182**, 6577–6583 (2000).
98. Johnson, R. C. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 230–271 (American Society for Microbiology, Washington DC, 2002).
99. Bannam, T. L., Crellin, P. K. & Rood, J. I. Molecular genetics of the chloramphenicol-resistance transposon Tn4451 from *Clostridium perfringens*: the TnpX site-specific recombinase excises a circular transposon molecule. *Mol. Microbiol.* **16**, 535–551 (1995).
100. Stark, W. M., Grindley, N. D. F., Hatfull, G. F. & Boocock, M. R. Resolvase-catalysed reactions between *res* sites differing in the central dinucleotide of subsite I. *EMBO J.* **10**, 3541–3548 (1991).
101. Prak, E. T. & Kazazian, H. H. Jr. Mobile elements and the human genome. *Nature Rev. Genet.* **1**, 134–144 (2000).
102. Martin, S. L. & Bushman, F. D. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell Biol.* **21**, 467–475 (2001).
103. Feng, Q., Moran, J. V., Kazazian, H. H. Jr. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
104. Feng, Q., Schumann, G. & Boeke, J. D. Retrotransposon R1Bm endonuclease cleaves the target sequence. *Proc. Natl Acad. Sci. USA* **95**, 2083–2088 (1998).
105. Yang, J., Malik, H. S. & Eickbush, T. H. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl Acad. Sci. USA* **96**, 7847–7852 (1999).
106. Arkhipova, I. R., Pyatkov, K. I., Meselson, M. & Evgen'ev, M. B. Retroelements containing introns in diverse invertebrate taxa. *Nature Genet.* **33**, 123–124 (2003).
107. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31**, 159–165 (2002).
Demonstration that the human TP-retrotransposon, LINE-1, can integrate into DNA lesions by an endonuclease-independent mechanism.
108. Symer, D. E. *et al.* Human l1 retrotransposition is associated with genetic instability *in vivo*. *Cell* **110**, 327–338 (2002).
109. Eickbush, T. H. in *Mobile DNA II* (eds. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 813–835 (American Society for Microbiology, Washington DC, 2002).
110. Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B. & Lambowitz, A. M. Mobile introns: pathway and proteins. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 761–783 (American Society for Microbiology, Washington DC, 2002).

111. Cousineau, B., Lawrence, S., Smith, D. & Belfort, M. Retrotransposition of a bacterial group II intron. *Nature* **404**, 1018–1021 (2000).
112. Cousineau, B. *et al.* Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**, 451–462 (1998).
113. Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381**, 332–335 (1996).
114. Zimmerly, S. *et al.* A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83**, 529–538 (1995).
115. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**, 545–554 (1995).
References 114 and 115 represent the first demonstration of the mechanism of group II intron mobility by TP reverse transcription of an intron RNA that reverse-spliced into a DNA target.
116. Ichiyanagi, K. *et al.* Retrotransposition of the LI. LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.* **46**, 1259–1272 (2002).
117. Pardue, M.-L. & DeBaryshe, P. G. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 870–887 (American Society for Microbiology, Washington DC, 2002).
118. Toor, N. & Zimmerly, S. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* **8**, 1373–1377 (2002).
119. Fugmann, S. D., Villey, I. J., Ptaszek, L. M. & Schatz, D. G. Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol. Cell* **5**, 97–107 (2000).
120. Kim, D. R., Dai, Y., Mundy, C. L., Yang, W. & Oettinger, M. A. Mutations of acidic residues in RAG1 define the active site of the V(*D*)J recombinase. *Genes Dev.* **13**, 3070–3080 (1999).
121. Landree, M. A., Wibbenmeyer, J. A. & Roth, D. B. Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(*D*)J recombination. *Genes Dev.* **13**, 3059–3069 (1999).
122. van Gent, D. C. *et al.* Initiation of V(*D*)J recombination in a cell-free system. *Cell* **81**, 925–934 (1995).
123. McBlane, J. F. *et al.* Cleavage at a V(*D*)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83**, 387–395 (1995).
124. Gellert, M. V(*D*)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101–132 (2002).
125. Dai, Y. *et al.* Nonhomologous end joining and V(*D*)J recombination require an additional factor. *Proc. Natl Acad. Sci. USA* **100**, 2462–2467 (2003).
126. Messier, T. L., O'Neill, J. P., Hou, S. M., Nicklas, J. A. & Finette, B. A. *In vivo* transposition mediated by V(*D*)J recombinase in human T lymphocytes. *EMBO J.* **22**, 1381–1388 (2003).
127. Hanai, R. & Wang, J. C. The mechanism of sequence-specific DNA cleavage and strand transfer by phi X174 gene A* protein. *J. Biol. Chem.* **268**, 23830–23836 (1993).
128. van Mansfeld, A. D., van Teeffelen, H. A., Baas, P. D. & Jansz, H. S. Two juxtaposed tyrosyl-OH groups participate in phi X174 gene A protein catalysed cleavage and ligation of DNA. *Nucl. Acids Res.* **14**, 4229–4238 (1986).
129. Zechner, E. L. *et al.* in *The Horizontal Gene Pool, Bacterial Plasmids and Gene Spread* (ed. Thomas, C. M.) 87–175 (Harwood Academic Publishers, Amsterdam, 2000).

Acknowledgements

We thank M. Belfort for comments on this work and, in particular, N. Grindley for his enthusiastic discussions on transposon families and mechanisms. This work was supported by the National Institutes of Health.

Online links

DATABASES

The following terms in this article are linked online to:
Swiss-Prot: http://us.expasy.org/sprot/
Hin invertase | λ-integrase | RAG1 | RAG2 | TnsA | TnsB | TnsC | TnsD | TnsE | γδ-resolvase |

FURTHER INFORMATION

ACLAME Database: http://aclame.ulb.ac.be/
Flybase, Drosophila Transposons, Transgene Constructs, and Insertions:
http://flybase.bio.indiana.edu/transposons/
IS Database:
http://www-IS.biotoul.fr/IS.html
Swiss-Prot keyword, Transposition:
http://us.expasy.org/cgi-bin/get-entries?KW=Transposition
Access to this interactive links box is free online.