

Chapter 14

DDE Transposon as Public Goods



Louis Tsakou, Célia Vicari, Laura Helou, Vivek Keshri, Sabyasachi Das, Yves Bigot, and Pierre Pontarotti

Abstract DDE Transposons have been recruited at least four times as site-specific recombination activating gene allowing programmed DNA elimination in eukaryotes. The described cases are RAG in jawed vertebrates, Kat 1 and Alpha 3 in the *Kluyveromyces lactis* yeast and Piggymac/TPB1 TPB2 and TPB6 in ciliates. The domesticated RAG is the most known case. It constitutes the enzymatic core of the Jawed vertebrates V(D)J recombination machinery. It directs random assembly and joining of gene segments during the development of B and T cells helping in the generation of the enormous gene diversity encoding antibodies or T cell receptors. It was shown in the case of RAG that the shift from DDE transposon to site-specific recombination activating gene is an evolutionary phenomenon that did not require dramatic changes. This explains why the co-option of DDE transposon as site-specific recombination activating gene can occur in a convergent manner. As numerous genes coding for DDE transposases are widespread through numerous members of the life tree, it is expected that several of them might correspond to domesticated transposons involved in programmed DNA elimination and maybe in the generation of receptor diversity. The domestication of DDE transposon could have been and still be of an extreme importance for organisms' evolution.

L. Tsakou · C. Vicari · V. Keshri · P. Pontarotti (✉)
Aix Marseille Univ IRD, APHM, MEPHI, IHU Méditerranée Infection, Marseille, France
e-mail: Pierre.pontarotti@univ-amu.fr

L. Helou · Y. Bigot
UMR INRAE 0085, CNRS 7247, PRC, Centre INRAE Val de Loire, 37380 Nouzilly, France

S. Das
Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University,
1462 Clifton Road North-East, Atlanta, GA 30322, USA

P. Pontarotti
SNC5039 CNRS, Paris, France

© Springer Nature Switzerland AG 2020
P. Pontarotti (ed.), *Evolutionary Biology — A Transdisciplinary Approach*,
https://doi.org/10.1007/978-3-030-57246-4_14

1

14.1 Preamble

Eukaryotes and prokaryotes do not have the same DNA content in their genome over stages of their cell differentiation. Indeed genomes are modified during development or cell differentiation. This phenomenon is called developmentally regulated genome rearrangement (DRGR) (Zufall et al. 2005). DRGRs start by a DNA breaks either due to a specific endonuclease activity or specific chromatin structure that is related to a mechanism inducing a DNA recombination event.

14.2 Developmentally Regulated Genome Rearrangement via a Specific Endonuclease

All the reported cases so far correspond to the domestication of selfish DNA that displayed at the minimum an endonuclease activity derived from a DDE transposase, a homing endonuclease, or a prophage recombinase.

The homothallic switching endonuclease (HO) is a domesticated homing endonuclease (Fig. 14.1). The HO endonucleases are mobile genetic elements that correspond to DNA fragments encoding these proteins over all their length, from their first 5' to the last 3' nucleotide. These DNA fragments use the homing endonuclease they encode to mediate their mobility (Keeling and Roger 1995; Koufopanou and Burt 2005). Homing endonucleases specifically recognize and cleave a specific oligonucleotide motif that is generally located within a very conserved motif of some house-keeping protein-coding sequence. After DNA cleavage, a DNA fragment coding an HO endonucleases specifically inserts in frame with the parasitized gene in the middle of its own recognition sequences (Fig. 14.1). In a diploid cell that is heterozygous for a homing endonuclease, the gene lacking the parasitic element becomes cleaved at the recognition site, and the broken chromosome is invaded by the parasitic DNA fragment by homologous recombination using the homing endonuclease-containing gene as a template. The mating type in budding yeasts is derived from the domestication of one of these parasitic DNA fragments. The mating types in budding yeasts are encoded by the mating-type (MAT) loci *MAT*^a and *MAT*^α. Some yeasts have the ability to change their mating type (mating-type switching) without going through mating or meiosis." In *Saccharomyces cerevisiae*, switching is initiated when the homothallic switching (HO) endonuclease induces a DNA double-strand break (DSB) in the *MAT* locus. Next, the replacement is completed through a gene conversion, in which transcriptionally silent copies of *MAT* genes, known as "hidden MAT left, *HML*^α" and "hidden MAT right, *HMR*^a," are copied into the expressed *MAT* locus.

Prophage excision involved in the specific rearrangement. (Feiner et al. 2015, Fig. 14.2)

During the differentiation process in procaryote, prophages and derived prophages are excised from the genomic region using a prophage recombinase in order to

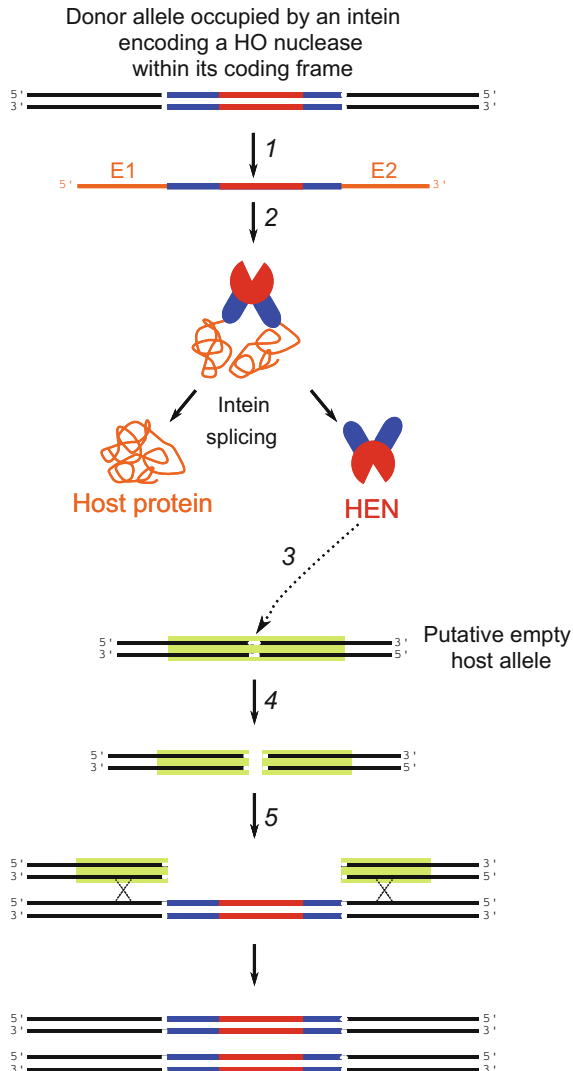
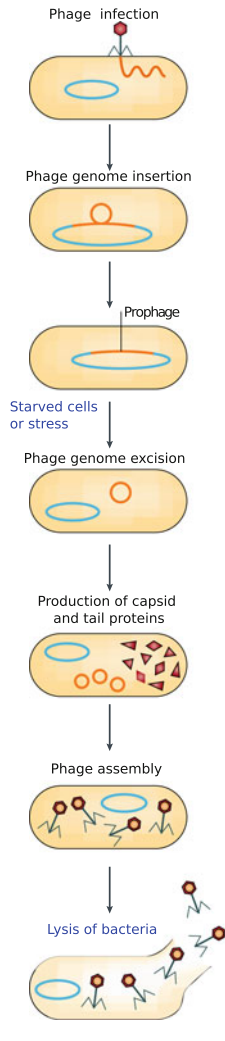


Fig. 14.1 Organization, expression, splicing, and mobility from one allele occupied by an intein to an unoccupied allele. The main steps of the intein mobility and homing are numbered. Step 1 is the RNA transcription. Black lines indicate DNA strands of the occupied and unoccupied alleles by an intein-coding DNA fragment, and cyan blue and light green bars indicate the coding-frame of the DNA region transcribed for each allele kind. Blue and red lines indicated regions coding for the intein (blue) and the homing endonuclease (HEN) moieties (red) in the DNA and the corresponding RNA transcripts. Step 2 is translation into protein. “pacmans” HEN moieties are shown in red and blue ellipses are N- and C-terminal intein moieties. Step 3 is site-specific recognition of an unoccupied allele by HEN. Blank spaces indicated the insertion site specifically cleaved by the HEN in both alleles. Step 4 is specific cleavage by the HEN. Finally, steps 5 and 6 depicted the invasion of a DNA fragment encoding intein from an occupied allele toward an unoccupied one by DNA strand invasion at DNA replication. This figure was constructed from derived information and graphic elements from Piègu et al. (2015)

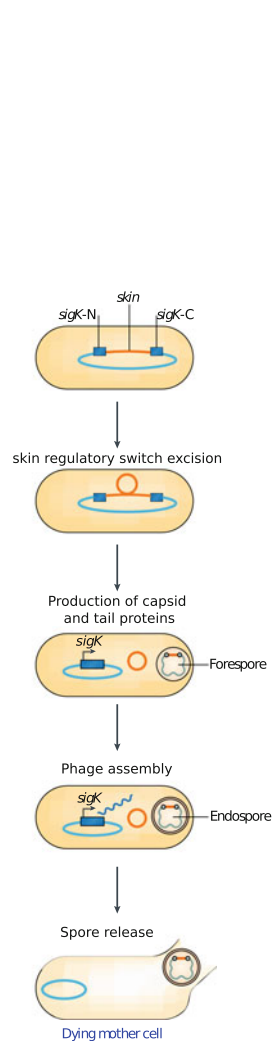
57 combine ORFs in frame. This process is involved in at least three phenomena:
 58 sporulation (*Bacillus* and *Clostridium*), heterocyst differentiation (Cyanobacteria),
 59 and monocytogenes phagosomal escape (*Listeria*).

60 In the case of the “well analyzed” *Bacillus subtilis*, gene rearrangement occurs
 61 in the sigK gene that is disrupted into two segments by the insertion of SKIN (sigK
 62 intervening element) which is a remnant of the ancestral prophage (Abe et al. 2017a).
 63 During sporulation, SKIN is excised from the chromosome to combine the ORFs in

a. Lysogenic phage cycle



b. Spore development in *Bacillus subtilis*



◀**Fig. 14.2** Parallel between the lysogenic cycle of a phage (a) and the bacterial process involving a phage-derived mechanism activating sporulation (b). **a** Temperate phages enter a cycle in which their phage genome is first going to be integrated into the bacterial chromosome to become a prophage that is going to persist in what is considered a phage latent or dormant state in which bacterial cells will be viable and will not produce of phage particles. Prophages are replicated together with the bacterial host chromosome during host-cell replication and switch into lytic production upon exposure to stress. Chronologically, the entry in the lytic phase starts with the excision of the phage genome, its gene transcription then the production of phage capsid and tail proteins that are thereafter assembled in phage particles. The last step is the lysis of the bacteria that releases phages and kills the bacteria. **b** Regulation of mother cell-specific genes during sporulation in *Bacillus subtilis*. A phage regulatory switch (phageRS), named *skin*, is inserted within the *sigK* gene. *SigK* encodes a protein, σ K, that regulates the expression of late-stage sporulation genes in the mother cell. *skin* excises itself at the initiation of the sporulation process, leaving an intact *sigK* gene able to encode a functional σ K protein. Post excision, σ K expression activates the mother cell's late-stage sporulation genes. Following excision, the excised *skin* element is eventually lost in the mother cell, which dies late during sporulation. By contrast, the forespore, which did not undergo element excision, gives rise to an endospore that still encodes the *skin* element within its *sigK* gene. This figure was constructed from derived information and graphic elements from

64 frame. In addition to *sigK*, many other examples of sporulation-specific gene rear-
 65 rangement occur, suggesting that this phenomenon is widespread and common in
 66 spore-forming bacteria (the intervening sequence can correspond to only a recombi-
 67 nase of prophage origin. The recombinase has two functions: it catalyzes the DNA
 68 cleavage at the recombination site and join the DNA molecule ends of the restored
 69 protein-coding genes (Abe et al. 2017b) (Fig. 14.2)

70 Heterocysts in cyanobacteria are specialized cells with a role in nitrogen fixation.
 71 They provide nitrogen to vegetative cells. Differentiation in a heterocyst is due the
 72 excision of fragments that interrupt three different loci: *nitD*, *fdxN*, and *hupL* that
 73 encodes a dinitrogenase alpha subunit, a ferredoxin and an uptake hydrogenase large
 74 subunit, respectively. The excised fragments have prophage origins and include at
 75 least the recombinase gene (Hilton et al. 2016).

76 Finally, to promote phagosomal escape, *Listeria monocytogenes* need to excise a
 77 temperate prophage integrated into the *comK* locus during bacterial phagocytosis in
 78 order to activate the *comK* expression (Feiner et al. 2015)

79 14.3 Programmed Break Specified by Chromatin Signals

80 14.3.1 Mating-Type Switching in the Fission Yeast *S. pombe*

81 As in the case of *S. cerevisiae*, the yeast *Schizosaccharomyces pombe* is able of
 82 mating-type switching. This mechanism occurred in an independent manner, the two
 83 species that belong to two highly divergent clades of yeast. The *S. pombe* genome
 84 contains one active (*mat1*) and two silenced (*mat2* and *mat3*) mating-type loci that
 85 share similarities with the *S. cerevisiae* MAT proteins. However, the mechanism

86 of switching is different from that involved in *S. cerevisiae*. Instead of cleavage
87 mediated by an HO endonuclease (Fig. 14.2), a fragile chromosomal site consisting
88 of an epigenetic mark at the *mat1* locus in switching-competent cells leads to a
89 dsDNA break during replication (Klar et al. 2014).

90 **14.3.2 Immunoglobulin Switching**

91 Immunoglobulin isotype class switching is a mechanism that changes a B cell's
92 production of immunoglobulin from one type to another. It involves an intrachromo-
93 somal deletional rearrangement that focuses on the region of 1–10 kb of repetitive
94 switch (S) located upstream of each heavy chain isotype gene. Class switch recom-
95 bination is induced by a break done by cytidine deaminase AID (AID-APOBEC
96 family) on accessible chromatin which is expressed during the B cell development
97 (Yu and Lieber 2019).

98 **The variable lymphocyte receptors (VLRs) system** (Fig. 14.3). In agnathes
99 (cyclostomes), the VLR system involved in adaptive immunity includes a germline
100 VLR that do not code for functional protein but instead encodes the only portion
101 of the amino and carboxyl termini of the mature VLRLs. The sequences encoding
102 those portions are separated by non-coding intervening regions. In lymphocytes, the
103 germline VLRLs are assembled by somatic DNA rearrangement into a mature VLR
104 that encodes the functional receptor via the insertion of LRR cassette that flank the
105 germline VLR. The germline VLR is broken by the AID-APOBEC enzyme at the
106 intervening sequences between the C-terminal and N-terminal portions of the VLR
107 genes where the chromatin is accessible. Then gene conversion starts thanks to
108 sequence identity between the intervening sequence and sequences surrounding the
109 LRR cassettes (Boehm et al. 2018).

110 **14.3.3 Immunoglobulin Diversity Driven by Gene Conversion** 111 **in Birds and Some Mammals**

112 In birds and some mammals, only one pair of functional V and J segments is found
113 for both the Ig light and heavy chain loci. Therefore, the diversity generated by
114 V(D)J recombination is limited. However, several pseudo-V coding segments are
115 found upstream the functional V segment in genes coding light and heavy chains.
116 These pseudo-V segments are used as a template for gene conversion to diversify the
117 single functional V segment. To launch the diversification process, a break done by
118 the Activation-induced deaminase (AID) within the single functional V segment is
119 required. (see for example Arakawa et al. 2002).

120 Other DRGRs have been described in the literature (Wang and Davis 2014), and
121 they likely correspond to an infinitesimal number of cases that occur in nature. At the

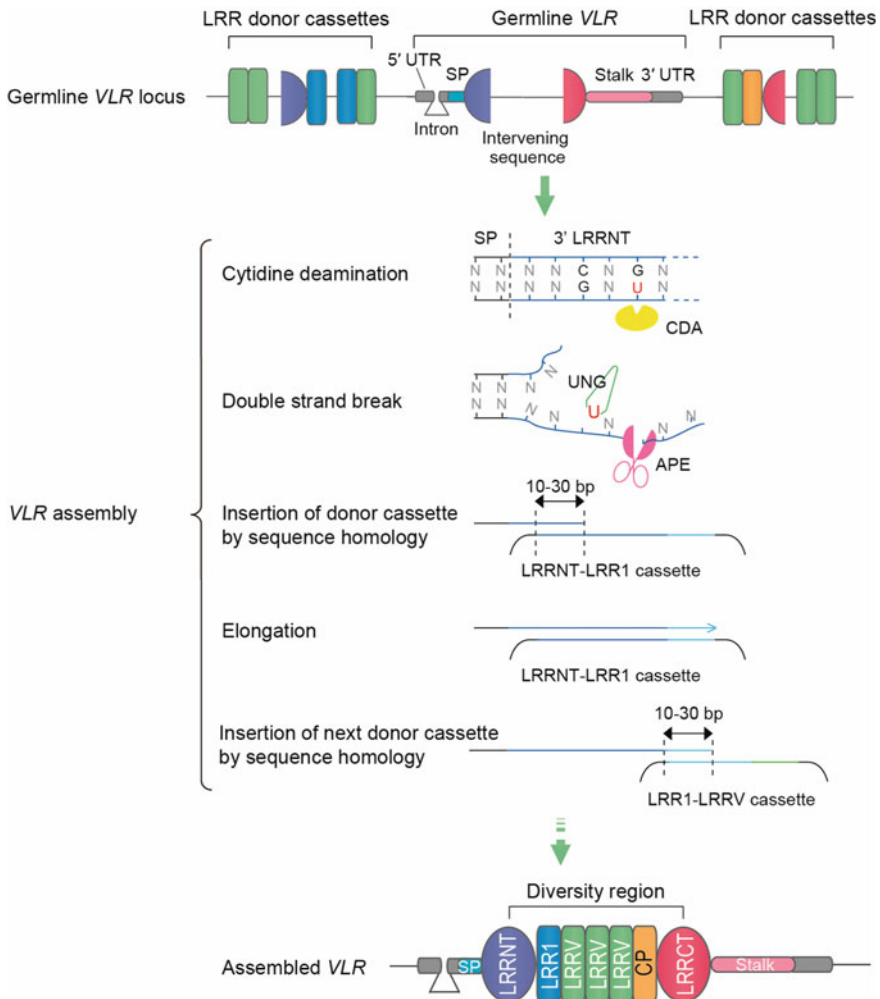


Fig. 14.3 Variable lymphocyte receptors (VLRs) system. Schematic diagram of a germline VLR gene and the postulated gene conversion-like (copy choice) mechanism for VLR assembly. The germline VLR gene is incomplete and contains invariant regions encoding for 5'-end of the N-terminal LRR (5'LRRNT) and 3'-end of the C-terminal LRR (3'LRRCT) and stalk. Hundreds of different LRR cassettes are located upstream and downstream of a pre-assembled germline VLR gene. The non-coding intervening sequence between 5'LRRNT and 3'LRRCT is replaced by the donor LRR cassettes that are sequentially copied either from 5' to 3' or 3' to 5' direction. At the beginning of the gene assembly process, a cytosine deaminase (CDA) converts cytosine (C) to uracil (U) in the germline VLR gene. The uracil is then removed by uracil-DNA glycosylase (UNG), leaving an apurinic (AP) site. The AP site activates nicking activity of AP endonuclease (APE) which leads to a DNA double-strand break. To repair this break, homologous recombination starts based on the sequence homology of 10-30 bp between the donor and acceptor LRR cassettes. This process is repeated along with deletion of the intervening sequence until the completion of a mature VLR gene. The diversity region of a mature VLR is composed of a 3'LRRNT, LRR1, multiple LRRVs, connecting peptide (CP), and a 5'LRRCT

122 molecular level, as will be seen in the next section beside the prophage and among
123 other domesticated selfish elements, domesticated DDE transposons seem a solution
124 often retained by evolution to achieve programmed DNA elimination.

125 **14.4 DDE Transposon and Domesticated DDE Transposon**

126 DNA transposition is the process by which a discrete segment of DNA is either moved
127 or copied into a new genomic location. Several distinct types of enzymes catalyze
128 DNA transposition, one of the most abundant kinds are the DDE transposases thus
129 named for conserved essential acidic residues located at the active catalytic site. The
130 DDE transposase coding gene in the transposon is flanked by two Terminal Inverted
131 Repeats (TIRs). To achieve transposition, the transposase recognizes these TIRs to
132 perform the excision of the transposon which is after this or in a concerted manner
133 with excision is re-inserted into a new genomic location. Upon insertion, the target
134 site DNA is duplicated, resulting in Target Site Duplications (TSDs).

135 Several DDE transposons have been domesticated by their hosts retaining their
136 DNA binding capacity (see for review (Sinzelle et al. 2009; Jangam et al. 2017)) and
137 in some cases, the transposase and their related TIR as site-specific recombination
138 activating gene. The few well-studied examples will be presented in the next sections.

139 **14.4.1 Shift from a DDE Transposon to a Mechanism** 140 **of Programmed DNA Elimination**

141 DDE transposons have been recruited as site-specific recombination activating gene
142 at least four times as programmed DNA elimination system, mediating biological
143 differentiation processes: RAG (RAG1-RAG2) in jawed vertebrates, *Kat 1* and *Alpha*
144 *3* in *K. lactis* yeast and *PiggyBac* (TPB1 TPB2 and TPB6) in the ciliate *Tetrahymena*.
145 The first described and best known example is the V(D)J RAG.

146 **14.4.2 The Recombination Activating Gene (RAG) Paradigm**

147 Jawed vertebrates has a specific adaptive immune system based on lymphocytes
148 that express highly diverse, clonally distributed antigen receptors encoded by genes
149 that are non-functional in the germline and assembled by recombination during
150 lymphocyte development (Teng and Schatz 2015). This assembly reaction, known as
151 V(D)J recombination, operates on arrays of V, D, J polypeptide-coding segments of
152 immunoglobulin and T-cell receptor loci. V(D)J recombination is initiated early in
153 lymphocyte development by a site-specific endonuclease: RAG1 and 2. The RAGs

154 cleave at a conserved recombination signal (RSS) that flanks each V/D and J segments.
 155 The finding that the RAGs have transposase activity, supports a model in which co-
 156 option of the components of a transposon played a critical role in the evolution of the
 157 jawed vertebrate adaptive immune system (see for review Flajnik 2016) (Fig. 14.4).

158 The hypothesis was that the RAG proteins derive from transposase genes of RAG
 159 transposons while the split antigen receptor genes derive from the insertion of the
 160 terminal inverted repeats (TIRs) of this transposon into a Ig-like receptor gene exons
 161 with the inserted TIRs becoming the RSSs. The presence of a RAG transposase core
 162 and TIRs in non-vertebrates (Kapitonov and Jurka 2005) was consistent with this
 163 model. The discovery of a complete RAG transposon in amphioxus (*Branchiostoma*
 164 *belcheri*) (Huang et al. 2016) strengthens this hypothesis. The discovery of an active

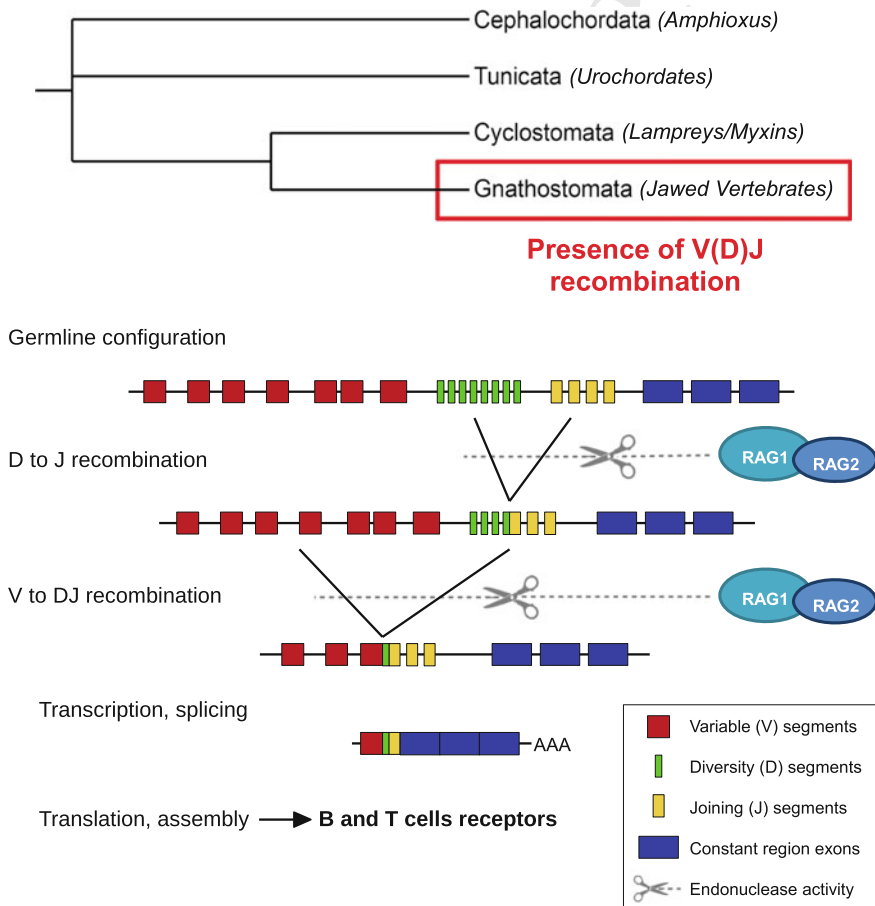


Fig. 14.4 RAG implication in the V(D)J recombination mechanism in jawed vertebrates. The chordates consensus tree shows the position of jawed vertebrates among chordates. The V(D)J mechanism was adapted from Janeway et al., 2001

165 RAG transposon in the hemichordata *Ptychodera flava* and several fossilized trans-
166 posons in several Deuterostomia (Morales Poole et al. 2017) indicate that the RAG
167 transposon was present in the deuterostomia ancestor and remained active to date in
168 several lineages of this clade, while it was co-opted as part of V(D)J recombinase
169 in jawed vertebrates about five hundred millions years ago. The RAG transposon
170 includes TIRs that are related in sequence to the RSS heptamer and the transcribed
171 open reading frame encodes RAG1-like and RAG-2 like proteins with a biochemical
172 activity similar to those of the RAGs, including DNA cleavage via a nick-hairpin
173 mechanism (Huang et al. 2016) (Fig. 14.5).

174 After DNA cleavage by the jawed vertebrate RAG or *Branchiostoma belcheri*
175 RAG transposon, there are two predominant fates for the excised fragment flanked
176 by RSSs or TIRs: joining of the ends to form a signal joint or integration into a
177 new locus. The jawed vertebrate recombinase RAG strongly favors the joining of
178 the ends. Indeed, RAG in the jawed vertebrates actively directs cleaved signals and
179 coding ends into the NHEJ DNA repair pathway for joining coding segments in
180 frame. In contrast, *Branchiostoma belcheri* RAG transposon appears to strongly
181 favor integration after excision, nevertheless, it allows some TIR-TIR joints to form
182 (Huang et al. 2016; Zhang et al. 2019). It is possible that the transposase partially
183 prevents the interaction between the TIR and the NHEJ repair pathway and that the
184 jawed vertebrate RAG lost this property. Zhang et al. (2019) started to uncover the
185 mechanism beyond the domestication and evidenced important amino acid positions
186 involved in transposition/or suppressing transposition (Fig. 14.7).

187 The mechanism of co-option can be described as follows: (1) insertion of an active
188 transposon into a given genomic locus, (2) the translocation of the transposase gene;
189 in some cases, the native TIRs remain within the genomic locus. (3) The TIR like
190 sequence is recognized by the domesticated transposase (Known under the name
191 of RSSs in the case of RAG) and the TIRs together with the internal sequence are
192 excised (4) the excised sequence lost the ability to insert another genomic region and
193 the two flanking ends are joined by the non-homologous end-joining (NHEJ) DNA
194 repair pathway, to form a coding joint (CJ) (Fig. 14.6).

195 In conclusion, biochemical functions of the DDE transposon and the sequence-
196 specific recombination activating system are very similar, hence the biochemical
197 shift from a transposon to a sequence-specific recombination activating system corre-
198 sponds an easy evolutionary step requiring a very limited number of events to pass
199 form a “wild” state to a domesticated one (Fig. 14.3). This is also supported by the fact
200 that many other DDE transposons have been co-opted as recombination activating
201 site-specific endonuclease as described in the next paragraphs.

202 **Piggymac/TPB2/TPB1/TPB6** in ciliates (Baudry et al. 2009; Cheng et al. 2016;
203 Nowacki et al 2009) (Fig. 14.7).

204 Ciliates are unicellular organisms able to perform DNA rearrangements during
205 development in order to differentiate a somatic macronucleus that is metabolically
206 active from their transcriptionally silent germinal micronucleus. After duplication of
207 the germinal genome, a large proportion of their germinal micronuclear genome is
208 eliminated to differentiate a somatic macronuclear genome through the loss and in
209 particular through the excisions of thousands of internal eliminated sequences. This

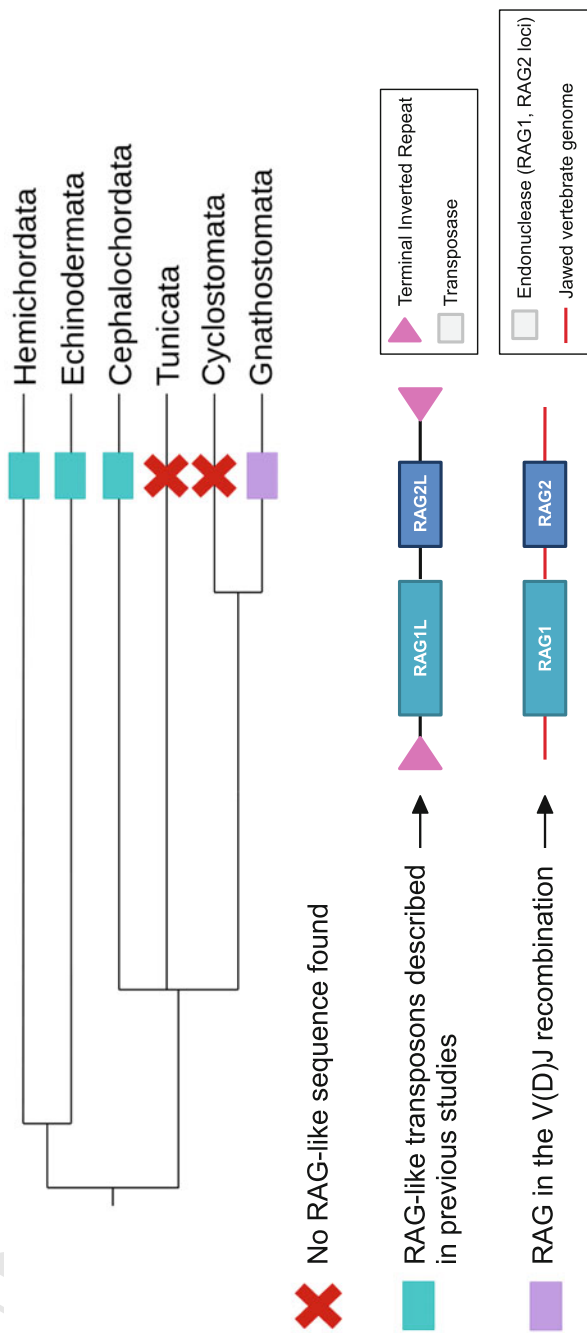


Fig. 14.5 Taxonomic distribution of the RAG and RAG-like sequences known in Deuterostomes. This is a deuterostomes consensus tree with the clades presenting RAG in the V(D)J recombination in the adaptive immune system, the clades presenting RAG-like sequences and the clades presenting no RAG-like sequences until today, described in Morales Poole et al. (2017)

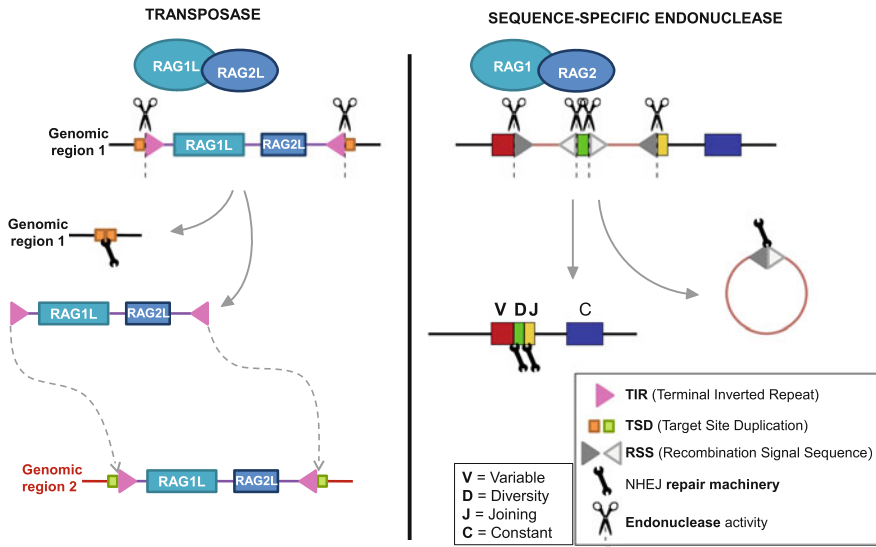


Fig. 14.6 Simple biochemical switch between the RAG-like transposase activity and the RAG endonuclease/recombinase activity in the V(D)J recombination of jawed vertebrates. Adapted from Huang et al. (2016)

210 is done by domesticated DDE transposon. Most of the eliminated sequences in the
 211 paramoecium and tetrahymena are named, respectively, by Piggymac and TPB2 and
 212 are derived from transposons belonging to the piggybac family (Fig. 14.7). In contrast
 213 to the domesticated *piggyBac* transposase genes in the macronuclear genomes of
 214 oligohymenophorea ciliates, *Oxytricha* bears thousands of active transposase genes
 215 within the *Tc1/mariner* superfamily.

216 In the case of the Tetrahymena, the domesticated transposons TPB1 and TPB6
 217 excise 12 internal eliminated sequences that disrupt exons. TPB1 and TPB6 recog-
 218 nize the TIR like sequences. TPB2 and likely a Piggymac in paramoecium seem to
 219 recognize RNA intermediates. In the case of TPB2, the TIR direct interaction with
 220 the TIR has been lost and the domesticated transposase performs its excision via
 221 scnRNA-directed heterochromatin involved specialization; thus, TPB2 to recognize
 222 heterochromatin rather than TIRs. Thus Tetrahymena TPB2 seemingly represents
 223 another level of transposase domestication. The situation is less clear in the case of
 224 Piggymac where the domesticated transposase seems to recognize both IES ends
 225 and the intermediate RNA. It is likely that in the case of TPB2 and Piggymac,
 226 the transposition mechanism was gradually grafted into a heterochromatin forma-
 227 tion pathway. In all these cases, the internal eliminated sequence form a circle that
 228 prevents reintegration in the genome as this is also the case for jawed vertebrates
 229 with the chromosomal excision product resulting from the RAG activity.

230 The phylogenetic analysis (not shown) indicated that possibly a PiggyBac
 231 transposon entered the Oligohymenophorea (that include the paramoecium and
 232 tetrahymena phyla) ancestor and became later domesticated. The ancestor

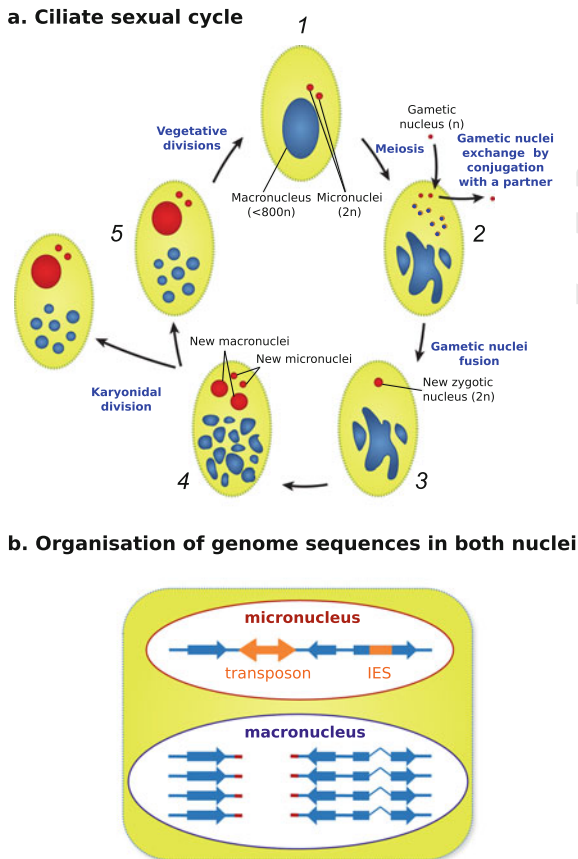


Fig. 14.7 Brief overview of biological and genomic features of the ciliate system. **a** Main steps of the sexual cycle in ciliate species belonging to the *Paramecium* genus. At the entry of the cycle in vegetative cells, the macronucleus starts disorganizing while the 2 micronuclei ($2n$) trigger meiosis in vegetative cells (1) resulting in eight gametic nuclei (n). Seven or six of these gametic micronuclei, respectively, degenerate after the acquisition of one gametic micronuclei or the exchange of one of them with another *Paramecium* cell. The fusion of both remaining gametic micronuclei (3) leads to a new zygotic nucleus ($2n$) that thereafter differentiate in new macronuclei and microneclei (4), before karyonidal division and differentiation of a new vegetative cell. **b** Organization of genome sequences in the micronucleus and the macronucleus. Here, the micronucleus genome is reduced to one chromosome containing the genes (blue arrows) that are for some of them split by insertion element sequence (IES; orange rectangle). DNA transposons (orange double arrows) can also be interspersed in chromosomes. During macronucleus differentiation, IES are precisely excised what restores split gene ORFs (the faint blue line means that the 2 ORF are fused in one ORF), transposons are eliminated what fragments chromosomes before the remaining chromosomal fragments are amplified. This figure was constructed from derived information and graphic elements in Nowacki et al. 2011

233 Piggymac/TPB2/TPB1/TPB6 domesticated transposon evolved in the two lineages:
 234 tetrahymena and paramecium Piggymac and TPB2/TPB1/TPB6. In the case of
 235 tetrahymena, a duplication gave rise to TPB2 in one part and TPB1/TPB6 ancestor
 236 in the other. TPB2 lost the possibility to recognize its TIR but gained the possibility
 237 to interact with the scan RNA. TPB1/TPB6 represent maybe the ancestral state and
 238 TPB2 could represent another state in the domestication process.

239 **Kat 1 and MAT alpha3** (Barsoum et al. 2010; Rajaei et al. 2014)

240 As we said above, some yeasts have the ability to change their mating type.
 241 In *Saccharomyces cerevisiae*, switching is initiated by homothallic switching (HO)
 242 endonuclease. In the related yeast *Kluyveromyces lactis* HO has been replaced by
 243 two domesticated DDE transposons: MATalpha 3 involved in the switch d *MAT* α to
 244 *MATa*. *Kat1* involved in the switch *de MATa* to *MAT* α .

245 *MAT* α 3 which was domesticated from transposable elements belonging to the
 246 Mutator Like Element Family MULEs. Regulated excision of this element results
 247 in a double stranded DNA break, (DSB) that stimulates recombination from the
 248 genome and initiates mating-type switching from *MAT* α to *MATa*. It has to be noted
 249 that in the case of this domesticated transposon the TIR-like sequences have been
 250 replaced by other sequences which are different on the left and right sides. The
 251 left side is a low complexity sequence with a long stretch of T and A. The right
 252 side contains a conserved motif that is conserved in sibling species-species of the
 253 *Kluyveromyces* genus. The other domesticated transposon named *Kat1* evolved from
 254 hAT (*hobo/Activator/Tam3*) transposases and is involved in the switching from *MATa*
 255 to *MAT* α . *Kat1* cleaves the *MATa* locus at two different positions, resulting in DSBs
 256 that stimulate recombination. *Kat1* recognizes a TIR like sequence. In both cases,
 257 the intervening DNA is joined into a circle.

258 In conclusion, DDE transposons evolved as site-specific recombination activating
 259 genes many times during evolution and therefore this is a case of isoconvergent
 260 evolution of site-specific recombination activating genes.

261 **14.4.3 Simple Evolutionary Shift Can Explain Convergent**
 262 **Evolution**

263 Isoconvergent evolution, the independent evolution of similar features from the same
 264 ancestral state in different evolutionary lineages (Pontarotti and Hue 2016), could
 265 be explained in part by natural selection where the new feature gave an advantage
 266 to the individual. Isoconvergent evolution could be due to the limited number of
 267 evolutionary pathways resulting from developmental and functional constraints on
 268 the evolutionary process (Losos 2011)—Functional constraints imposing a finite
 269 number of accurate adaptations, a finite number of mechanisms can be used to answer
 270 functional problems. Finally, isoconvergent evolution can be explained in part by the
 271 ease of transition from ancestral state to a derived state. This last aspect is not really
 272 discussed in the literature.

273 Losos (2011) pointed out that the wings powering flight in vertebrates have been
274 built in different ways in birds, pterosaurs and bats. In all these cases, the wings
275 represent modified forelimbs. The combination of wings and forelimbs, in theory,
276 would be not very useful in real life. He concluded that this was due to lack of
277 constraints; however, we underline here that besides the constraints, it was easier to
278 modify forelimbs to get wings than to start from nothing. The same reasoning could
279 be applied for the DDE transposon co-option as sequence-specific recombination
280 activating systems and the biochemical shift from a transposase to a sequence-specific
281 recombination activating endonuclease is an easy evolutionary step.

282 ***14.4.4 Transposases Form the Largest Family in the Diverse*** 283 ***Genomes of Life***

284 Transposase-encoding genes are greatly over-represented in sequenced genomes
285 and metagenomes relative to other coding sequences (Aziz et al. 2010). Some of
286 the transposase coding genes could correspond to domesticated transposons, active
287 transposons and maybe to fossilized ones. From this and the chapter developed
288 above, we hypothesize that many DDE transposons have been recruited as recombi-
289 nation activating site-specific endonuclease systems. Many domesticated candidates
290 have already been described in the literature, some of them lost the endonuclease
291 domain butand conserved the DNA binding domain and they could be involved in
292 novel chromatin-modifying complexes (Feschotte 2008) while others are involved
293 in centromere binding, chromosome segregation, meiotic recombination (Sinzelle
294 et al. 2009). Some articles also described large-scale systematic analysis to search
295 for domesticated transposons including DDE transposons (Hoen and Bureau 2015;
296 Bouallégue et al. 2017). In the human genome 26 putative DDE transposases has been
297 described (Arnaoty et al. 2012), among these putative domesticated transposases as
298 far as we know other than the RAGs, only the function of one of them, PBGD5 was
299 really investigated, and its nuclease activity has been shown, however we do not
300 know whether it recognizes specific sequences and if it is really involved in recom-
301 bination (Henssen et al. 2015, 2016, 2017). Another candidate has been tested for
302 its transposase activity (Majumdar et al. 2013) where the authors show that THAP9
303 gene encodes an active DDE DNA transposase.

304 We need to test further the hypothesis that many transposons have been recruited
305 as recombination activating site-specific endonuclease. If our hypothesis is true,
306 many domesticated DDE transposons acting as specific DNA endonucleases should
307 be found throughout the life diversity.

14.4.5 *Search for Domesticated Transposons Involved in Programmed Recombination. Test of the Hypothesis*

This paragraph is a small guideline. The strategy to search for domesticated transposons involved in programmed recombination can be performed by two complementary approaches: Look for the candidate domesticated transposase and look for a somatically rearranged genomic region.

14.4.6 *Look for the Candidate Transposase Co-opted as Site-Specific Endonuclease Involved in Genomic Recombination*

Different strategies have been developed in particular by Hoen and Bureau (2015) and Bouallègue et al. (2017). The strategy described here was adapted from these publications. We first have to look for the sequence that codes for complete transposase; these can be done by profile search using an alignment with a known transposase (Eddy 2011).

The ORF found needs to be checked for a bona fide catalytic site and the conservation of the DDE motif. Then in order to be sure that the transposase is likely to be active it is necessary to show that the protein evolved under constraint. This can be done by calculating the ratio of non-synonymous to synonymous substitutions (dN/dS) which is a useful measure of the strength and mode of natural selection acting on protein-coding genes (Jeffares et al. 2015).

The next step is to differentiate between transposase belonging to an active transposon from site-specific recombination activating gene (domesticated transposase). The following criteria must be present in the case of site-specific endonuclease. The domesticated transposase must be in single copy without pseudogene like sequences that could correspond to recent transposition events. The next criteria is based on the fact that the domesticated transposase should be unable to transpose thus the domesticated transposase should remain in the same genomic region in different species—conserved synteny—(Rascol et al. 2009). The higher the number of divergent species display conserved synteny, the higher is the probability that the transposase has been domesticated. For example, PGBD5 is present in all chordate genomes and belongs to a conserved synteny (Pavelitz et al. 2013) indicating that it is likely to have been domesticated.

The following step is to search the domesticated sequence recognition signal: find RSS/TIR like sequence. One way to search for the sequence recognition signal is to perform Chip-Seq experiment (Park 2009), and the other, that could be complementary to the Chip-Seq experiment is to be guided by sequence data. This could be done by looking for the active transposon which is the most similar to the domesticated transposon and to determine the TIR sequence of the transposon. The TIR sequence should be similar to the site-specific DNA sequence recognized by the domesticated



347 transposase, as this is the case for the RAG transposon and the domesticated RAG
348 (see Kapitonov and Jurka 2005; Huang et al. 2016; Morales Poole et al. 2017). The
349 TIR like sequences can then be searched in the genome(s) coming from the cells
350 where the “domesticated” transposase is transcribed. It should be noted that the
351 endonuclease could recognize RNA intermediates as this is the case for TPB2 in-
352 ciliates where the TIR recognition has been lost and instead a chromatin structure is
353 recognized.

354 ***14.4.7 Look for Somaticlly Rearranged Genomic Regions***

355 The next step is to look for rearranged genomic regions, and this can be done in
356 several ways. The first way would be to sequence the genome from the cell or
357 the tissue where the domesticated transposase is expressed (sequence 1) compare
358 this genome with a reference genome (sequence 2) from the same species (same
359 individual) in order to look for the rearranged sequence in sequence 1. Another way
360 is to use high throughput genome translocation sequencing methodology based on
361 the ability of a double-strand break to translocate to a fixed ‘bait’ double-strand break
362 generated by a nuclease (Hu et al. 2016).

363 In both cases, the genomic region, where the rearrangement occurred, should be
364 flanked by terminal repeats; these terminal repeats corresponding to the RSS should
365 be compared to the TIR of the active transposon (the most similar one).

366 Candidate regions using the second methodology have been described by Wei
367 et al. (2016). They described 27 recurrent double-strand breaks that occurred during
368 neural stem/progenitor cell differentiation and they named it as recurrent double-
369 strand break cluster. The authors hypothesized that the break region will be joined
370 to the distal break region thus potentially leading to new genetic information via for
371 example a novel exon combination that can be generated by recombination between
372 intronic regions at the DNA level via “exon shuffling.” (Alt and Schwer 2018) In
373 the case where the double-strand break followed by a recombination occurs via
374 the action of domesticated DDE transposon a possible candidate could be PGBD5.
375 Indeed PGBD5 seems to be expressed in the brain and fetal brain and therefore
376 possibly in the neural stem/progenitor cells (Pavelitz et al. 2013) and PGBD5 is able
377 to induce double-strand breaks in non-physiologic condition (Henssen et al.
378 2015, 2017).

379 ***14.4.8 Are There Other Domesticated Transposases that Are*** 380 ***Able to Generate Diversity?***

381 One important question concerns the possibility that some domesticated transposase
382 and their TIRs could create diversity through the rearrangement of distinct tandem

383 repeat paralogous gene segments (such as in V(D)J recombination). These processes
384 correspond to an excision between at least 2 sets of paralogous fragments with a
385 combinational joining. In these cases, the paralogous fragments should be flanked by
386 the TIR like sequences located 3' from each of the first paralogous fragments family
387 and 5' to each of the second fragment paralogous family. Such genomic rearrangement
388 can be searched. The corresponding transposon can be searched as described above.

389 14.5 Conclusion

390 We discussed in this chapter the possibility that DDE transposons have been recruited
391 several times as systems involved in programmed DNA elimination and perhaps in
392 the generation of receptor diversity. This is due to the fact that the functional shift is an
393 easy one and that the huge number of transposase domains are present and across the
394 tree of life. Therefore, DDE transposons along with other so-called selfish element
395 encoding nucleases (homing nuclease and prophage) could have been major players
396 in the evolution of biodiversity. The co-option of DDE transposons as the regulatory
397 element is also very important. This has been largely discussed, Justin Goodrich this
398 Issue) (Sinzelle et al. 2009; Jangam et al. 2017). Another important role of DDE
399 transposons is the one in the horizontal gene transfer (HGT). Horizontal transfer
400 of DDE transposons (HTT) has been widely reported in eukaryotes (El Baidouri
401 and Panaud 2015). However, in most cases, the literature discusses HGT and HTT
402 in a non-integrated manner. However, the role of DDE transposons as the carrier
403 of additional antibiotic resistance genes is well known in bacteria (Babakhani and
404 Oloomi 2018), and at least the role of DDE transposon as carrier one case in yeast
405 is also known. (McDonald et al. 2019). It is likely that most of the HGT are driven
406 by transposons, but as most of the studied cases HGT corresponds to ancient events
407 the transposase and the corresponding TIR has been lost so an effort in the analysis
408 of recent HGTs should be done. Therefore, DDE transposons could have been major
409 players in organismal evolution because it helps in genetic exchange between species.

410 The DDE transposon evolutionary trajectory should be included in the public
411 goods hypothesis for the evolution of life on Earth (McInerney et al. 2011). According
412 to this hypothesis, nucleotide sequences are simply seen as goods, passed from one
413 organism to another through both vertical and horizontal transfer. The interesting
414 things about DDE transposons are that they evolved in vertical and horizontal manner
415 and they should be seen as goods since they increase the possibility of transfer and
416 can be co-opted for example as recombination activating site-specific endonucleases
417 involved in programmed DNA elimination.

418 **Acknowledgements** This work was supported by the French Government under the «Investisse-
419 ments d'avenir» (Investments for the Future) program managed by the Agence Nationale de
420 la Recherche (ANR, fr: National Agency for Research), (reference: Méditerranée Infection
421 10-IAHU-03).

References

- 422
- 423 Abe K, Shimizu SY, Tsuda S, Sato T (2017a) A novel non prophage(-like) gene-intervening element
424 within *gerE* that is reconstituted during sporulation in *Bacillus cereus* ATCC10987. *Sci Rep*
425 7:11426. <https://doi.org/10.1038/s41598-017-11796-8>
- 426 Abe K, Takamatsu T, Sato T (2017b) Mechanism of bacterial gene rearrangement: SprA-catalyzed
427 precise DNA recombination and its directionality control by SprB ensure the gene rearrangement
428 and stable expression of *spsM* during sporulation in *Bacillus subtilis*. *Nucleic Acids Res* 45:6669–
429 6683. <https://doi.org/10.1093/nar/gkx466>
- 430 Alt FW, Schwer B (2018) DNA double-strand breaks as drivers of neural genomic change, function,
431 and disease. *DNA Repair (Amst)* 71:158–163
- 432 Arakawa H, HauschiLd J, Buerstedde JM (2002) Requirement of the activation-induced deaminase
433 (AID) gene for immunoglobulin gene conversion. *Science (80-)* 295:1301–1306. <https://doi.org/10.1126/science.1067308>
- 434
- 435 Arnaoty A, Pitard B, Bateau B, Bigot Y, Lecomte T (2012) Novel approach for the development
436 of new antibodies directed against transposase-derived proteins encoded by human neogenes.
437 *Methods Mol Biol* 859:293–305. https://doi.org/10.1007/978-1-61779-603-6_17
- 438 Aziz RK, Breitbart M, Edwards RA (2010) Transposases are the most abundant, most ubiquitous
439 genes in nature. *Nucleic Acids Res* 38:4207–4217. <https://doi.org/10.1093/nar/gkq140>
- 440 Babakhani S, Oloomi M (2018) Transposons: the agents of antibiotic resistance in bacteria. *J Basic*
441 *Microbiol* 58:905–917
- 442 Barsoum E, Martinez P, Åström SU (2010) $\alpha 3$, a transposable element that promotes host sexual
443 reproduction. *Genes Dev* 24:33–44. <https://doi.org/10.1101/gad.557310>
- 444 Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M (2009) PiggyMac,
445 a domesticated piggyBac transposase involved in programmed genome rearrangements in the
446 ciliate *Paramecium tetraurelia*. *Genes Dev* 23:2478–2483. <https://doi.org/10.1101/gad.547309>
- 447 Boehm T, Hirano M, Holland SJ, Das S, Schorpp M, Cooper MD (2018) Evolution of alternative
448 adaptive immune systems in vertebrates. *Annu Rev Immunol* 36:19–42. <https://doi.org/10.1146/annurev-immunol-042617-053028>
- 449
- 450 Bouallègue M, Rouault JD, Hua-Van A, Makni M, Capy P (2017) Molecular evolution of piggyBac
451 superfamily: from selfishness to domestication. *Genome Biol Evol* 9:323–339. <https://doi.org/10.1093/gbe/evw292>
- 452
- 453 Cheng CY, Young JM, Lin CYG, Chao JL, Malik HS, Yao MC (2016) The piggyBac transposon-
454 derived genes TPB1 and TPB6 mediate essential transposon-like excision during the develop-
455 mental rearrangement of key genes in *Tetrahymena thermophila*. *Genes Dev* 30:2724–2736.
456 <https://doi.org/10.1101/gad.290460.116>
- 457 Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1002195>
- 458
- 459 El Baidouri M, Panaud O (2015) Horizontal transfers and the new model of TE-driven genome
460 evolution in Eukaryotes. In: *Evolutionary biology: biodiversification from genotype to phenotype*.
461 Springer International Publishing, pp 77–92
- 462 Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA (2015) A new perspective on
463 lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* 13:641–650
- 464 Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*
465 9:397–405
- 466 Flajnik MF (2016) Evidence of G.O.D.'s miracle: unearthing a RAG transposon. *Cell* 166:11–12
- 467 Henssen AG, Henaff E, Jiang E, Eisenberg AR, Carson JR, Villasante CM, Ray M, Still E, Burns
468 M, Gandara J, Feschotte C, Mason CE, Kentsis A (2015) Genomic DNA transposition induced
469 by human PGBD5. *Elife*. <https://doi.org/10.7554/elife.10565>
- 470 Henssen AG, Jiang E, Zhuang J, Pinello L, Succi ND, Koche R, Gonen M, Villasante CM, Armstrong
471 SA, Bauer DE, Weng Z, Kentsis A (2016) Forward genetic screen of human transposase genomic
472 rearrangements. *BMC Genom* 17:548. <https://doi.org/10.1186/s12864-016-2877-x>

- 473 Hissen AG, Koche R, Zhuang J, Jiang E, Reed C, Eisenberg A, Still E, Macarthur IC, Rodríguez-
474 Fos E, Gonzalez S, Puiggròs M, Blackford AN, Mason CE, De Stanchina E, Gönen M, Emde
475 AK, Shah M, Arora K, Reeves C, Socci ND, Perlman E, Antonescu CR, Roberts CWM, Steen H,
476 Mullen E, Jackson SP, Torrents D, Weng Z, Armstrong SA, Kentsis A (2017) PGBD5 promotes
477 site-specific oncogenic mutations in human tumors. *Nat Genet* 49:1005–1014. <https://doi.org/10.1038/ng.3866>
478
- 479 Hilton JA, Meeks JC, Zehr JP (2016) Surveying DNA elements within functional genes of
480 heterocyst-forming cyanobacteria. *PLoS One*. <https://doi.org/10.1371/journal.pone.0156034>
481
- 482 Hoen DR, Bureau TE (2015) Discovery of novel genes derived from transposable elements using
483 integrative genomic analysis. *Mol Biol Evol* 32:1487–1506. <https://doi.org/10.1093/molbev/msv042>
484
- 485 Hu J, Meyers RM, Dong J, Panchakshari RA, Alt FW, Frock RL (2016) Detecting DNA double-
486 stranded breaks in mammalian genomes by linear amplification-mediated high-throughput
487 genome-wide translocation sequencing. *Nat Protoc* 11:853–871. <https://doi.org/10.1038/nprot.2016.043>
488
- 489 Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, Zhang Y, Yu W, Pontarotti P, Escriba H,
490 Le Petillon Y, Liu X, Chen S, Schatz DG, Xu A (2016) Discovery of an active RAG transposon
491 illuminates the origins of V(D)J recombination. *Cell* 166:102–114. <https://doi.org/10.1016/j.cell.2016.05.032>
492
- 493 Jangam D, Feschotte C, Betrán E (2017) Transposable element domestication as an adaptation to
494 evolutionary conflicts. *Trends Genet* 33:817–831
- 495 Jeffares DC, Tomiczek B, Sojo V, dos Reis M (2015) A beginners guide to estimating the non-
496 synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods Mol Biol* 1201:65–90. https://doi.org/10.1007/978-1-4939-1438-8_4
497
- 498 Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived
499 from Transib transposons. *PLoS Biol* 3:0998–1011. <https://doi.org/10.1371/journal.pbio.0030181>
500
- 501 Keeling PJ, Roger AJ (1995) The selfish pursuit of sex. *Nature* 375:283
- 502 Klar AJS, Ishikawa K, Moore S (2014) A unique DNA recombination mechanism of the mating/cell-
503 type switching of fission yeasts: a review. *Microbiol Spectr*. <https://doi.org/10.1128/microbiol-spec.mdna3-0003-2014>
504
- 505 Koufopanou V, Burt A (2005) Degeneration and domestication of a selfish gene in yeast: molecular
506 evolution versus site-directed mutagenesis. *Mol Biol Evol* 22:1535–1538. <https://doi.org/10.1093/molbev/msi149>
507
- 508 Losos JB (2011) Convergence, adaptation, and constraint. *Evolution (N Y)* 65:1827–1840. <https://doi.org/10.1111/j.1558-5646.2011.01289.x>
509
- 510 Majumdar S, Singh A, Rio DC (2013) The human THAP9 gene encodes an active P-element DNA
511 transposase. *Science (80-)* 339:446–448. <https://doi.org/10.1126/science.1231789>
512
- 513 McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfordorfer S, Milgate A, Solomon
514 PS (2019) Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA
515 between three fungal wheat pathogens. *MBio*. <https://doi.org/10.1128/mbio.01515-19>
516
- 517 McInerney JO, Pisani D, Baptiste E, O'Connell MJ (2011) The public goods hypothesis for the
518 evolution of life on Earth. *Biol Direct* 6:41. <https://doi.org/10.1186/1745-6150-6-41>
519
- 520 Morales Poole JR, Huang SF, Xu A, Bayet J, Pontarotti P (2017) The RAG transposon is active
521 through the deuterostome evolution and domesticated in jawed vertebrates. *Immunogenetics*
522 69:391–400. <https://doi.org/10.1007/s00251-017-0979-5>
523
- 524 Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF (2009) A functional
525 role for transposases in a large eukaryotic genome. *Science* 324:935–938
- 526 Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*
527 10:669–680
- 528 Pavelitz T, Gray LT, Padilla SL, Bailey AD, Weiner AM (2013) PGBD5: a neural-specific intron-
529 containing piggyBac transposase domesticated over 500 million years ago and conserved from
530 cephalochordates to humans. *Mob DNA* 4:23. <https://doi.org/10.1186/1759-8753-4-23>
531

- 526 Pontarotti P, Hue I (2016) Road map to study convergent evolution: A proposition for evolutionary
527 systems biology approaches. *Evolutionary biology: convergent evolution, evolution of complex*
528 *traits, concepts and methods*. Springer International Publishing, Heidelberg, pp 3–21
- 529 Rajaei N, Chiruvella KK, Lin F, Åström SU (2014) Domesticated transposase Kat1 and its fossil
530 imprints induce sexual differentiation in yeast. *Proc Natl Acad Sci U S A* 111:15491–15496.
531 <https://doi.org/10.1073/pnas.1406027111>
- 532 Rascol VL, Levasseur A, Chabrol O, Grusea S, Gouret P, Danchin EGJ, Pontarotti P (2009)
533 CASSIOPE: an expert system for conserved regions searches. *BMC Bioinf* 10:284. <https://doi.org/10.1186/1471-2105-10-284>
- 534 Sinzelle L, Izsvák Z, Ivics Z (2009) Molecular domestication of transposable elements: from
535 detrimental parasites to useful host genes. *Cell Mol Life Sci* 66:1073–1093
- 536 Teng G, Schatz DG (2015) Regulation and evolution of the RAG recombinase. In: *Advances in*
537 *immunology*. Academic Press Inc, pp 1–39
- 538 Wang J, Davis RE (2014) Programmed DNA elimination in multicellular organisms. *Curr Opin*
540 *Genet Dev* 27:26–34
- 541 Wei PC, Chang AN, Kao J, Du Z, Meyers RM, Alt FW, Schwer B (2016) Long neural genes harbor
542 recurrent DNA break clusters in neural stem/progenitor cells. *Cell* 164:644–655. <https://doi.org/10.1016/j.cell.2015.12.039>
- 543 Yu K, Lieber MR (2019) Current insights into the mechanism of mammalian immunoglobulin class
544 switch recombination. *Crit Rev Biochem Mol Biol* 54:333–351
- 545 Zhang Y, Cheng TC, Huang G, Lu Q, Surleac MD, Mandell JD, Pontarotti P, Petrescu AJ, Xu A,
547 Xiong Y, Schatz DG (2019) Transposon molecular domestication and the evolution of the RAG
548 recombinase. *Nature* 569:79–84. <https://doi.org/10.1038/s41586-019-1093-7>
- 549 Zufall RA, Robinson T, Katz LA (2005) Evolution of developmentally regulated genome
550 rearrangements in eukaryotes. *J Exp Zool Part B Mol Dev Evol* 304:448–455