4
Genome Dynamics and Stability

Series Editor: Dirk-Henner Lankenau

Transposons and the Dynamic Genome

Volume Editors: Dirk-Henner Lankenau, Jean-Nicolas Volff

With 36 Figures

Springer

Series and Volume Editor:

Priv.-Doz. Dr. Dirk-Henner Lankenau
Hinterer Rindweg 21
68526 Ladenburg
Germany
e-mail: d.lankenau@t-online.de

Volume Editor:

Prof. Dr. Jean-Nicolas Volff
Ecole Normale Supérieure de Lyon
Institute de Génomique Fonctionnelle
46 alleé d'Italie
69364 Lyon Cedex 07
France
e-mail: Jean-Nicolas.Volff@ens-lyon.fr

Cover

The cover illustration depicts two key events of DNA repair: 1. The ribbon model shows the structure of the termini of two Rad50 coiled-coil domains, joined via two zinc hooks at a central zinc ion (sphere). The metal dependent joining of two Rad50 coiled-coils is a central step in the capture and repair of DNA double-strand breaks by the Rad50/Mre11/Nbs1 (MRN) damage sensor complex. 2. Immunolocalization of histone variant γ-H2Av in γ-irradiated nuclei of *Drosophila* germline cells. Fluorescent foci indicate one of the earliest known responses to DNA double-strand break formation and sites of DNA repair.
(provided by Karl-Peter Hopfner, Munich and Dirk-Henner Lankenau, Heidelberg)

ISSN 1861-3373 e-ISSN 1861-3381
ISBN 978-3-642-02004-9 e-ISBN 978-3-642-02005-6
DOI 10.1007/978-3-642-02005-6
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009929233

Editor: Dr. Sabine Schwarz
Desk Editor: Ursula Gramm, Heidelberg
Cover figures: Prof. Karl-Peter Hopfner and Dr. Dirk-Henner Lankenau
Cover design: WMXDesign GmbH, Heidelberg
Typesetting and Production: le-tex publishing services GmbH, Leipzig

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

It will be some time before we see "slime, protoplasm, &c." generating a new animal. But I have long regretted that I truckled to public opinion, and used the Pentateuchal term of creation, by which I really meant "appeared" by some wholly unknown process. It is mere rubbish, thinking at present of the origin of life; one might as well think of the origin of matter.

Charles Darwin to James D. Hooker, March 29, 1863

Relax, there's nothing wrong with the transposition paper. People aren't ready for this yet. I stopped publishing in refereed journals in 1965 because there was no interest in the maize controlling elements.

Barbara McClintock to Mel Green, 1969

Sometimes my students and others have asked me: "what was first in evolution – retroviruses or retrotransposons?" Since Howard Temin proposed that retroviruses evolved from retrotransposons (Temin 1980; Temin et al. 1995) the other alternative that retroviruses emerged first and were the predecessors of LTR-retrotransposons has since been a controversial issue (Terzian et al., this BOOK). While DNA-transposons could not have existed in an ancestral RNA-world by definition, sure enough, some arguments definitely point towards a pre-DNA world scenario in which retroelements were the direct descendants of the earliest replicators representing the emergence of life. First, these replicators likely catalyzed their own or other's replication cycles via the catalytic properties of RNA molecules. After translation had emerged some replicators possibly encoded an RNA polymerase first. This later evolved into reverse transcriptase (RT), i.e. the most prominent key-factor at the transition into the DNA world. Simultaneously, replicators could also have encoded membrane protein-genes such as the *env* gene of recent DNA-proviruses. Membranes were likely present much earlier as prebiotic oily films that supported the evolution of a prebiotic-protometabolism (Dyson 1999; Griffiths 2007). However, how

these promiscuous communities of ancestral molecules and protocells interacted, and how the exact branching chronology of earliest events in molecular evolution led to the emergence of replicators, membrane slicks, obcells (Cavalier-Smith 2001) still remains a mystery. It still underscores Charles Darwin's statement cited top left, while Barbara McClintock's remark more than 100 years later (cited top right), represents the spirit for not giving up these most fundamental topics.

One scenario is very likely: from the geochemically dominated times of the early planet earth, prebiotic promiscuous communities including membranes, proto-peptides, metabolites, and replicators represented the ingredients of Darwin's "*wholly unknown process*." From these, we now think, life emerged in conformity with a dual definition of life based on genetics and metabolism.[1]

The platform for transposon-research is simple. Besides "genes," transposable elements evolved as indwelling entities within all cellular genomes. Thereby, they exhibited both a parasitic as well as a symbiotic double-feature that may date back to the very beginnings of life itself. Celebrating Charles Darwin's bicentenary this year, we certainly do well to honor the fact that Darwin's concept of gemmules directly led to our present day term "*genes*" (Gould 2002; Lankenau 2007b). How pleased would Darwin have been to see this idea brought onto the right track, e.g. through the works of Mendel, Weismann, deVries, or McClintock. How pleased would he have been to know how close we come today to his grand challenge: "The Origin of Species." Darwin, in fact even came as close as he could to humanities deepest concern formulating his famous statement:

"It is often said that all the conditions for the first production of a living organism are now present, which could ever have been present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, &c., present, that a protein compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed." (Charles Darwin 1871).

This statement also perfectly highlights our current technical hitches – but some have been overcome, and transposable elements have their share in approaching the solution of the grand enigma. How pleased would Darwin have been if he could have shared our modern insights into transposon-biology – as we now understand some of the inner workings of transposon activities and

[1] Life is defined synergistically as the merging of replication and metabolism. H.J. Muller wrote: *It is to define as alive any entities that have the properties of multiplication, variation and heredity* (Muller 1966). While metabolism supplies the monomers from which the replicators (i.e. genes or transposable elements) are made, replicators alter the kinds of chemical reactions occurring in metabolism. Only then can natural selection, acting on replicators, power the evolution of metabolism (Dyson 1999; Maynard Smith and Szathmary 1997).

of analogous selfish genetic elements that triggered molecular, coevolutionary chases through sequence space and the emergence of driver systems resulting in "molecular peacock's tails" such as "autosome killer-chromosomes," "selfish sex chromosomes," and "genomic imprinting machineries." Despite his surmise that present day metabolism would devour or absorb all ancient metabolic systems, we now understand that a great deal of ancient bits of information survived inside the chromosomes of all organisms in the form of sequence relicts. A lot of these ancient molecular relicts belong to the stunning, endogenous survival machines that always represented the major engines of evolution since the times of the genetic takeover – in a sense they form the pillars of life, capable of shaping the evolution of genomes and opportunistically altering genome structure and dynamics: transposable elements and viruses as their extracellular satellites, that fill our world's oceans with an unimaginable number of 10^{31} entities, or else, 10^7 virions per ml of surface seawater (Bergh et al. 1989; Williamson et al., 2008).

In fact, life began as and is driven by an emergent self-organizing property. Transposable elements seem to have played a significant role as executors of Gould's/Eldgredge's Punctuated Equilibrium[2]. How are transposable elements defined and why are they important? Transposable elements are specific segments of genomic DNA or RNA that exhibit extraordinary recombinational versatility. Treating a transposable element as an individual biological entity, it is best defined as a *natural, endogenous, genetic* ***toolbox of recombination***. This entity also overlaps with a wider definition of the term gene.[3] A transposable element is typically flanked by non-coding, direct, or inverted repeat sequences of limited length (less than 2 kb) often with promoter- and recombinational functions. These repeats flank a central core sequence, which among few other genes encodes a transposase/integrase and/or reverse transcriptase (RT). Transposable elements are the universal components of living entities that appear to come closest in resembling the presumed earliest replicators (including autocatalytic ribozymes) at the seed crystal level of the origins of life. Stuart Kauffman realized that Darwinian theory must be expanded to recognize other sources and rules of order based on the internal numeric, genetic, and developmental constraints of organisms and on the structural limits and contingencies of physico-chemical laws (Kauffman 1993). While Kauffman's approach is a step toward a deep theory of homeostasis, it is smart to define

[2]Originally Stephen Gould's and Niels Eldredges' punctuated equilibrium theory holds that most phenotypic differences occur during speciation periods but that species embedded in stable environments are remarkable stable in phenotype thereafter (Eldredge and Gould 1972). Here, the expression "phenotypic stability" is extended beyond this definition that focused on biological species. The molecular structure of genomes exhibits an analogous platform of stable order. "Genes" and "transposable elements" are examples of such a stable platform of order with emergent self-organizing properties – see also: (Kauffman 1993).

[3]In a broad context, a gene is defined as any portion of chromosomal material that potentially lasts for enough generations to serve as a unit of natural selection (Dawkins 1976).

the starting point of life as the *catalytic closure*[4] of two elementary systems intrinsic to all forms of cellular life: (1) prebiotic protometabolism and (2) genetic inheritance[5] encompassing transposon-like replicators. Both (1) and (2) formed a duality at the emergence of life. As for Newton's second law of motion ($F = ma$) the couplet of terms metabolism and inheritance is defined in a circle; each (gene and biotic metabolism) requires the other. In fact, this circularity lay behind Poincaré's conception of fundamental laws as definitional conventions (Kauffman 1993). Further, the logical separation of the two is technical only and for argumentational, experimental purposes it is useful. On the primordial earth, ordered prebiotic proto-metabolism (Dyson 1999) likely congregated in the vicinity of geochemically formed membrane surfaces or within hemicells or obcells as Cavalier-Smith called them (Cavalier-Smith 2001; Griffiths 2007). Such earliest metabolically ordered environments perhaps were too dynamic to establish long chained replicators such as RNA. At present it appears more realistic to assume the origin and growth of long RNA molecules in sea ice (Trinks et al. 2005). Freeman Dyson unfolded a possible series of evolutionary steps establishing the modern genetic apparatus, with the evolutionary predecessors of transposable elements (i.e. replicators) at the heart of this process, establishing the modern genetic apparatus. Let us assume that the origin of life "took place" when a hemicell contained an ordered, homeostatically stable metabolic machinery (compare the similar ideas of Cavalier-Smith 2001). This system maintained itself in a stable homeostatic equilibrium. The major transition, establishing life was the integration of RNA as a self-reproducing cellular "parasite" but not yet performing a symbiotic genetic function for the hemicell. This transitional state must have been in place before the evolution of the elaborate translation apparatus linking the two systems could begin (Dyson 1999). The first replicators were not yet what we call transposable elements sensu stricto. They still had to evolve genes for proteins such as integrase and reverse transcriptase (RT). This transitional state of merging metabolism and replication represented the first of life's punctuated equilibria (Gould 2002) resulting in the inseparable affiliation of parasitic/symbiotic interactions of metabolites and replicators. The inseparable affiliation of symbiotic/parasitic features is the most typical characteristic of transposable elements active within modern genomes. After the genetic code and translation had been invented, and when the first retroelements evolved RT from some sort of RNA replicase, transposable elements (i.e. retroelements) triggered yet another punctuated equilibrium, i.e. the transition from the RNA world to an RNA/DNA world. Amazingly, the deep window into earth's most ancient past is still reflected by the vivid actions of transposable elements and viruses within all present-day genomes – it also includes the significant chimerical feature of parasitic versus symbiotic interdependencies. From time to time – typically, as evolution is

[4] *Catalytic closure* is defined as a system where every member of the autocatalytic set has at least one of the possible last steps in its formation catalyzed by some member of the set, e.g. peptides and RNA.

[5] See footnote 1

tinkering (Jacob 1977) – transposable element sequences that usually evolve under the laws of selfish and parasitic reproductive constraints became domesticated as useful integral parts of cellular genomes. One of the most forceful examples is the repeated domestication of sequence fragments from an endogenous provirus reprogramming human salivary and pancreatic salivary glands during primate evolution (Samuelson et al. 1990). The other prominent example of transposon domestication is the evolution of V(D)J recombination from the "RAG-transposon" crucial for the working of our immune system (Agrawal et al. 1998).

The above considerations force us to discern the historic rootage of transposable elements in geological deep time. The following chapters will serve sketching some of the enduring consequences of the emergence of transposable elements as inseparable constituents of modern genomes – as indwelling forces of species, populations and cells, recent and throughout evolution. The first two chapters establish key aspects of the significance of transposon dynamics as major engines of evolution on the level of genomes, populations, and species. The first chapter summarizes general theoretical approaches to transposon dynamics applicable to prokaryotes, as well as eukaryotes, with emphasis on the parasitic nature of transposable elements. Arnaud Le Rouzic and Pierre Capy point out that the evolution of a novel transposon insertion is similar to the dynamics of a single locus gene exposed to natural selection, mutations, and genetic drift. Different "alleles" can coexist at each insertion locus, e.g., a "void" allele without any insertion, a complete insertion, and multiple variants of deleted defective, inactivated alleles progressively accumulating through mutational erosion. Even though not mentioned in this context, the first chapter nicely approaches the NK model of Stuart Kauffman that forms the conceptual backbone of his grand opus the "Origins of Order" (Kauffman 1993, pp. 40–43). In the NK model N is the number of distinct genes in a haploid genome while K is the average number of other genes which epistatically influence the fitness contribution of each gene. Le Rouzic and Capy address the problem of a stable equilibrium. This, perhaps in the future promises to become congruent with Kauffman's prediction that many properties of the fitness-landscapes created with the NK model appear to be surprisingly robust and depend almost exclusively upon N and K alone (Kauffman 1993, p. 44). The second chapter merges historical aspects of transposable element dynamics at the infra- and transspecific populational level with modern approaches at the epigenetic level. While transposable elements were first discovered by Barbara McClintock in maize, Christina Vieira et al. focus and underscore the importance of *Drosophila* as a model organism in transposon research and populational studies.

The third chapter by Agnès Dettai and Jean-Nicolas Volff exemplifies the SINE[6] retroelements as a model system of real novel insertions of transposable

[6] Short interspersed nuclear elements (SINEs)

elements within variable chromosomal sites. SINES are shown as key examples for the powerful mode of evolutionary genome dynamics. Novel insertions not only create new fitness landscapes on which selection can act but if established within all germline genomes of a species they become powerful molecular morphological markers that are employed for cladistic analysis identifying unambiguous branching points in phylogenetic trees. This chapter truly represents the legacy of Willi Hennig's phylogenetic systematics (Hennig 1966; Hennig 1969) on a modern molecular platform. The chapter also lists a number of software tools making whole genome analysis feasible. Chapters 4 and 5 focus on transposable elements, and on the origin and regulation by means of double-stranded RNA and RNA interference (RNAi), another key-factor with evolutionary significance. While King Jordan and Wolfgang Miller review the control of transposable elements by regulatory RNAs and summarize general aspects of genome defense Christophe Terzian et al. in Chapter 5 present insights into the most interesting and the first example of an insect retrovirus, i.e. the endogenous *gypsy* retrotransposon of *Drosophila*. This retrovirus indeed represents an unmatched model system for multiple aspects of the biology of endogenous retroviruses as well as of an active retrotransposon. The *gypsy* provirus had been studied previously in connection with the host encoded Zn-finger protein Suppressor of Hairy Wing [Su(Hw)]. This protein turned out to be a chromatin insulator regulating chromatin boundaries and controlling enhancer-driven promoter activities. Its repetitive binding site within the *gypsy* provirus must have evolved within the *gypsy* retroelement by means of transposon evolution, perhaps in a quasispecies-like way. It is one of the most impressive examples demonstrating the emergence of the potential power of novel regulatory functions within host genomes (Gdula et al. 1996; Gerasimova and Corces 1998; Gerasimova et al. 1995). Terzian et al. (Chapter 5) advance our understanding and broaden our insights of *gypsy* driven by piRNA control mechanisms located within the heterochromatic *flamenco* locus. They further review recent findings as to the role of the envelope (Env) membrane protein serving as a model for retroviral horizontal and vertical genome transfer.

Another spectacular evolutionary example is presented in Chapter 6 by Walisko et al. It is the story of the revitalization of an ancient inactive DNA transposable element called *Sleeping Beauty*. It was reconstructed based on conserved genomic sequence-information only in the laboratory. The story is like Michael Crichton's Jurassic Park scenario, where dinosaurs were reconstructed from DNA in mosquito blood fossilized in amber. While Crichton's experiments were fiction, *Sleeping Beauty* is a real, reanimated "transposon-dinosaur." It existed for millions of years as an eroded, defective molecular fossil within a fish genome and was reactivated to study host-cell interactions in experimentally transfected human cells. Last but not least, the final chapter by Izsvák et al. describes the interactions of transposable elements with the cellular DNA repair machinery. Barbara McClintock first recognized the interdependence of chromosome breaks and transposition in her famous breakage-

fusion-bridge cycle (McClintock 1992 (reprinted)). In the early 1990s Bill Engels and co-workers discovered the fundamental, prominent double-strand break repair mechanism they called Synthesis-Dependent Strand Annealing (SDSA) as the underlying molecular mechanism repairing P-transposable element-induced double-strand breaks. This mechanism of homologous recombination is now widely recognized and its role in genome dynamics is interwoven into many volume chapters of this book series. As regards content Chapter 7 therefore closes the cycle and links this fourth book volume of the series to the first volume integrating multiple aspects of genome integrity (Lankenau 2007a).

Altogether, this book gives insight and a future perspective regarding the significance of transposable elements as selfish molecular drivers and universal features of life that exhibit in the words of Burt and Trivers "a truly subterranean world of sociogenetic interactions usually hidden completely from sight" (Burt and Trivers, 2006).

I most cordially thank all chapter authors for contributing to this volume on genome dynamics and transposable elements. Most importantly, I am deeply grateful to all the referees whose names must be kept in anonymity. At least two for each chapter were involved in commenting, shaping, and struggling with the individual scripts – I really, greatly appreciate their efforts! I thank Jean Nicolas Volff for organizing the transposable element meeting at Wittenberg some time ago and helping to invite some of the authors. I also thank the editorial staff at Springer who have always been patient with the editors and authors alike and have provided much help. I especially thank the managing editor Sabine Schwarz at Springer Life Sciences (Heidelberg) and the desk editor Ursula Gramm (Springer, Heidelberg) for their enduring assistance. I would also like to mention that le-tex publishing services oHG, Leipzig did a good job in production editing and preparing the manuscripts for print.

Ladenburg, April 2009 Dirk-Henner Lankenau

References

Agrawal A, Eastman QM, Schatz DG (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394:744–751

Bergh O, Borsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. Nature 340:467–468

Burt A, Trivers R (2006) Genes in Conflict. The Belknap Press of Harvard University Press, Cambridge, Ma; London

Cavalier-Smith T (2001) Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. J Mol Evol 53:555–595

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dyson FJ (1999) Origins of life, Rev. edn. Cambridge University Press, Cambridge, U.K.; New York

Eldredge N, Gould SJ (1972) Punctuated equilibria: An alternative to phyletic gradualism. In: Schopf TJM (ed) Models in palaeobiology. Freeman Cooper, San Francisco, pp 82–115

Gdula DA, Gerasimova TI, Corces VG (1996) Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. Proc. Natl. Acad. Sci. U S A 93:9378–9383

Gerasimova TI, Corces VG (1998) Polycomb and Trithorax group proteins mediate the function of a chromatin insulator. Cell 92:511–521

Gerasimova TI, Gdula DA, Gerasimov DV, Simonova O, Corces VG (1995) A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell 82:587–597

Gould SJ (2002) The structure of evolutionary theory. Belknap Press of Harvard University Press, Cambridge, Mass., USA

Griffiths G (2007) Cell evolution and the problem of membrane topology. Nat Rev Mol Cell Biol 8:1018–1024

Hennig W (1966) Phylogenetic Systematics. University of Illinois Press, Illinois, USA

Hennig W (1969) Die Stammesgeschichte der Insekten. Vlg. Waldemar Kramer, Frankfurt

Jacob F (1977) Evolution and tinkering. Science 196:1161–1166

Kauffman SA (1993) The origins of order: self organization and selection in evolution. Oxford University Press, New York

Lankenau D-H (2007a) Genome integrity: Facets and perspectives. Springer, Berlin Heidelberg New York

Lankenau D-H (2007b) The legacy of the germ line – maintaining sex and life in metazoans: Cognitive roots of the concept of hierarchical selection. In: Egel R, Lankenau D-H (eds) Recombination and meiosis – Models, means and evolution, vol 3. Springer, Berlin Heidelberg New York, pp 289–339

Maynard Smith J, Szathmary E (1997) The major transitions in evolution. Oxford University Press, Oxford

McClintock B (1992 (reprinted)) Chromosome organization and genetic expression. In: Fedoroff N, Botstein D (eds) The dynamic genome: Barbara McClintock's ideas in the century of genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp 73–107

Muller HJ (1966) The gene material as the initiator and organizing basis of life. Am Nat 100:493–517

Samuelson LC, Wiebauer K, Snow CM, Meisler MH (1990) Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. Mol Cell Biol 10:2513–2520

Temin HM (1980) Origin of retroviruses from cellular moveable elements. In: Cell, vol 21, pp 599–600

Temin HM, Cooper GM, Temin RG, Sugden B (1995) The DNA provirus: Howard Temin's scientific legacy. ASM Press, Washington, D.C.

Trinks H, Schröder W, Biebricher CK (2005) Ice and the origin of life. Orig Life Evol Biosph 35:429–445

Williamson SJ et al. (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS ONE 3:e1456

Contents

Genome Dyn Stab (4)
D.-H. Lankenau, J.-N. Volff: Transposons and the Dynamic Genome
DOI 10.1007/7050_017/Published online: 15 July 2006

Theoretical Approaches to the Dynamics of Transposable Elements in Genomes, Populations, and Species

Arnaud Le Rouzic[1,2] · Pierre Capy[1] (✉)

[1] Laboratoire Évolution, Génétique et Spéciation (CNRS), Avenue de la Terrasse, Bâtiment 13, 91198 Gif sur Yvette, France
pierre.capy@legs.cnrs-gif.fr

[2] *Present address:*
Linnaeus Centre for Bioinformatics, Uppsala Universitet, 75124 Uppsala, Sweden

Abstract Transposable elements are major components of both prokaryotic and eukaryotic genomes. They are generally considered as "selfish DNA" sequences able to invade the chromosomes of a species in a parasitic way, leading to a plethora of mutations such as insertions, deletions, inversions, translocations and complex rearrangements. They are frequently deleterious, but sometimes provide a source of genetic diversity. Numerous population genetics models have been proposed to describe more precisely the dynamics of these complex genomic components, and despite a wide diversity among transposable elements and their hosts, the colonization process appears to be roughly predictable. In this paper, we aim to describe and comment on some of the theoretical studies, and attempt to define the "life cycle" of these genomic nomads. We further raise some new issues about the impact of moving sequences in the evolution and the structure of genomes.

1 Introduction

Transposable Elements (TEs) seem to be an outstanding example of evolutionary success. They are present in almost all known living species, from eubacteria and archaebacteria to the multicellular organisms. They show a huge genetic and functional diversity, and they seem to have explored during the evolution process, the most relevant ways possible to duplicate and maintain themselves in the genome of their "host". The persistence of TEs in the genome, sometimes in spite of significant deleterious effects, is generally attributed to their amplification ability. This is the basis of the "selfish DNA" theory (Orgel and Crick 1980; Doolittle and Sapienza 1980; Hickey 1982).

Selfish DNA sequences appear to be submitted to several antagonistic multi-level forces, driving them along various evolutionary pathways. These depend on multiple factors, such as the biology of the host species, the features of the TE family, or simply chance. TE dynamics can be quite com-

plex such that further analysis rests on mathematical models of population genetics. At the molecular level, the more efficient the transposition process, the more likely the colonization of the genome will be. However, if the elements are deleterious for the host, individuals carrying too many copies will be eliminated through natural selection. Evolution of genomes would also certainly lead to the appearance of systems controlling or regulating replication, and elements are likely to evolve towards a way of bypassing such systems. Recurrent genomic mutations lead to partial or complete deletions or inactivations of TE copies, while some elements or fragments of elements may remain integrated in the genome and participate in an adaptive function of the organism. In this chapter, we propose to review the interactions existing between a genome and such internal parasites from a population genetics point of view. These interactions can change radically between the several successive stages of the invasion, from the active colonization of the genome by elements, to the probable loss of the transposition activity.

2
Genome Colonization

Theoretical studies of TE dynamics are generally challenged by the complexity of the process (see Charlesworth et al. 1994; Le Rouzic and Deceliere 2005 for review). The evolution of each TE insertion is actually similar to the dynamics of a single locus gene exposed to natural selection, mutations, and genetic drift. Different "alleles" can coexist at each insertion locus (e.g., a "void" allele without any insertion, a complete insertion, and multiple deleted, defective, inactivated alleles progressively appearing through mutations), and each of them might have different transposition rates and different impacts on the fitness in heterozygous or homozygous states. Depending on the stage in the invasion and on the features of the element, several insertions, often a few dozens and sometimes much more, have to be considered simultaneously. Finally, the total number of insertion sites is thought to vary, each transposition event leading to a new insertion locus.

2.1
Copy Number Dynamics

Except for complex computer simulations, modelling such a system must be achieved through approximations. For instance, the initial invasion of the element in a void population can be modelled in the same way as segregation distortion, considering only one insertion locus (Hickey 1982). However, this approach does not give us the opportunity to explore the subsequent steps of the invasion, when TEs accumulate in the genome, and it therefore becomes

necessary to consider average copy numbers. Charlesworth and Charlesworth (1983), for example, proposed to describe the variation of the average copy number $\bar{n}$ by $\Delta\bar{n} \simeq \bar{n} \cdot (u - v)$, where u is the transposition rate and v the deletion rate. This transposition (respectively deletion) rate corresponds to the mean number of transposition (or deletion) events for one copy in one generation. "Transposition" and "deletion" have to be understood here as generic terms aiming to include multiple kinds of molecular events, since only the resulting state is considered: a transposition (or, more precisely, a duplication) event leads to the appearance of a copy at a new insertion site, while a deletion results in the lost of a copy from its original insertion site[1]. This model is supposed to be approximately universal (i.e., all known TEs can fit with this model provided u and v are set accurately). If $u > v$, the element is able to invade, and the copy number increasing is exponential (Fig. 1). However, such dynamics do not appear realistic, since an infinite multiplication of a TE in a genome probably leads to its destruction. Two main evolutionary forces are supposed to be able to counterbalance this invasion: transposition regulation and natural selection (Fig. 2).

Transposition regulation consists in a decrease of the transposition rate during the invasion[2]. It can be roughly modelled by a transposition rate (i.e., duplication rate) $u_{\bar{n}}$ which is dependent on the mean copy number in the population $\bar{n}$: the higher the copy number, the lower the transposition rate. When the transposition rate $u_{\bar{n}}$ is equivalent to the deletion rate v, then $\Delta\bar{n} = 0$ and an equilibrium state is achieved (Fig. 1). However, this equilibrium situation supposes that $u = v$, which is generally not verified in natural populations, where transposition rates are usually at least one order of magnitude higher than the deletion rates (Nuzhdin and Mackay 1995; Suh et al. 1995; Maside et al. 2000). It, therefore, appears unlikely that transposition regulation is the only evolutionary force implied in TE copy number control.

Due to their activity, TEs represent a potential source of a large spectrum of mutations and chromosomal rearrangements. These mutations have been shown to be generally deleterious (Eanes et al. 1988; Mackay et al. 1992; Charlesworth 1996; Houle and Nuzhdin 2004), and natural selection is

[1] Class I elements (retrotransposons) transpose by a replicative mechanism, often referred as "copy and paste"; they can, however, be lost – or duplicated (Lankenau et al. 1994) – through other mechanisms, such as recombination between the terminal repeats of LTR retrotransposons (Vitte and Panaud 2003), or by synthesis dependant strand annealing (SDSA) (Lankenau and Gloor 1998). On the contrary, class II transposons move through a "cut and paste" mechanism; they are excised from the donnor site and reinserted at a new locus. They are, however, frequently duplicated through a homologous template dependant process (Brookfield 1995). Even if these mechanisms are not related, the overall dynamics of a TE family can be described by a transposition rate and a deletion rate, and interestingly, the order of magnitude of these parameters do not appear to be very different across TE classes (Hua-Van et al. 2005).

[2] This phenomenon has been described for many elements in several species (Labrador and Corces 1997). It is particulary well documented in intensively studied systems, such as *P* element in *Drosophila melanogaster* and its *KP* repressor (Jackson et al. 1988; Simmsons et al. 1990; Corish et al. 1996).

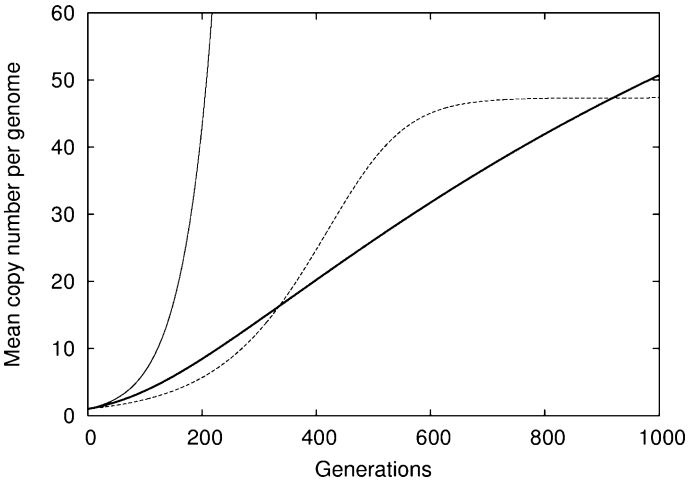

Fig. 1 Basic transposable element dynamics. If the transposition rate (frequency of a duplication event per copy and per generation) as well as the deletion rate (probability for a copy of being lost by various processes – see text) are constant, without any selection, the copy number increases exponentially ($\Delta n = n \cdot (u - v)$, with $u = 0.02$ and $v = 0.001$, *thin continuous line*). This probably does not correspond to a realistic situation, and several hypotheses have been proposed to explain the limitation of TE amplification (Charlesworth and Charlesworth 1983): (i) a regulation system, which supposes that the transposition rate decreases with the copy number: $\Delta n = n \cdot (u_n - v)$, with $u_n = u/(1 + k \cdot n)$, k being a factor that quantifies the intensity of regulation (here, $k = 0.2$, *thick line*); (ii) natural selection that eliminates, in each generation, a part of the insertions from the genome; $\Delta n = n \cdot (u - v - \partial \log w_n / \partial n)$. The *dotted line* represents the dynamics of such a system, with $w_n = 1 - s \cdot n$ (additive effects of insertions), and $s = -0.01$ (i.e., each insertion decreases the fitness by 1%)

also likely to restrain the TE proliferation. In a polymorphic population, the individuals carrying the lower number of copies are more likely to reproduce, leading to a slight decrease, each generation, in the mean copy number. Charlesworth and Charlesworth (1983) proposed to model this process by $\Delta \bar{n} = \bar{n} \cdot (u - v - s_{\bar{n}})$, where $s_{\bar{n}} = |\partial \log w_{\bar{n}} / \partial \bar{n}|$, w_n representing the fitness of an individual carrying n copies (and $w_{\bar{n}}$ being the fitness of a virtual individual having the average number of copies $\bar{n}$, which is reasonably close to the average fitness of the population). This model does not always lead to a stable equilibrium (Fig. 1), depending on the shape of the fitness curve w_n (Fig. 3).

The two processes (i.e., regulation and selection) are not mutually exclusive, and one can easily imagine that the TE amplification can be subject to both of them. Well-known TE families, such as *P* element in *Drosophila*, indeed appear to be both regulated (Lemaitre et al. 1993; Coen et al. 1994) and selected against (Snyder and Doolittle 1988; Eanes et al. 1988). A sim-

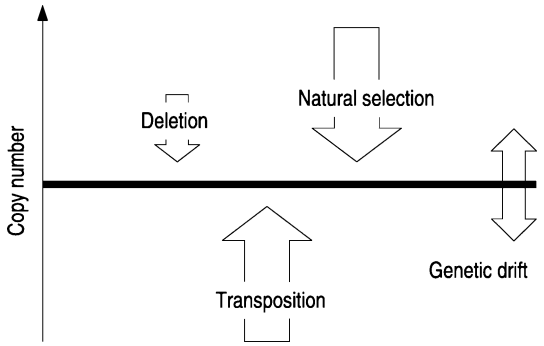

Fig. 2 Simple representation of the different evolutionary forces implied in the dynamics of TE copy number in the genome of a species. Transposition (or, more exactly, duplication) will increase the average copy number, while various kinds of transposition-related or unrelated deletions or excisions will eliminate copies from the genome. If the insertions are deleterious, the individuals carrying fewer copies will reproduce better than the others, and natural selection will decrease the mean copy number in the population. Several processes can be involved in this fitness loss: direct effect of insertions in genes or regulatory regions, repetitions leading to deleterious ectopic recombinations, or straight deleterious effect of the transposition activity (Nuzhdin 1999). Finally, in small populations, random genetic drift can shift the copy number below or above the expected value. At the beginning of the invasion process, the transposition rate is probably high, and the genomic copy number increases. A further equilibrium state can be achieved when increasing and decreasing forces are balanced; a decay in the transposition rate (recurrent mutations of active copies, transposition regulation …) or an intensification of the selective strengths can lead to this situation

ple model that combines both natural selection and transposition regulation shows that the effects of both evolutionary forces are cumulative (Fig. 4): if the transposition regulation is too weak to induce a realistic stabilization of the copy number, and if the selection strength alone is not sufficient to lead to an equilibrium (even if the fitness function does not match the conditions detailed in Fig. 3), then a perfectly realistic equilibrium copy number can be achieved when both control mechanisms overlap.

2.2 The Birth of a New TE Invasion

All these models describe the colonization of a TE family as a deterministic process. The spread of a TE in a population, and the progressive increase in the copy number does indeed appear as a predictable mechanism (e.g., Biémont 1994), provided the population size is large, thus limiting the influence of genetic drift (for the role of genetic drift in TE dynamics, see Brookfield and Badge 1997). However, regardless of the population size, an element cannot escape from randomness at the beginning of its invasion.

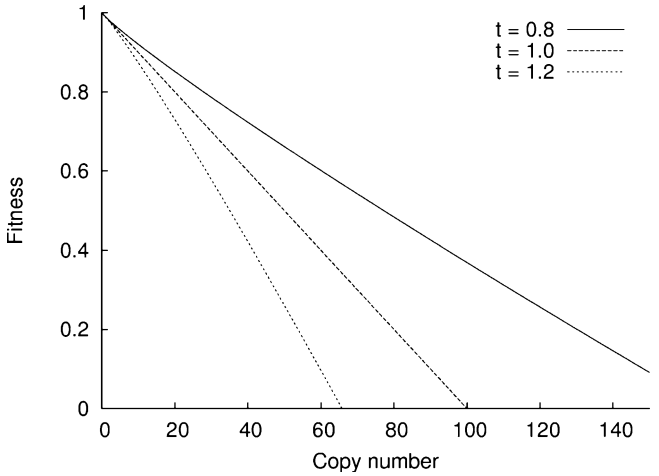

Fig. 3 The existence of a potential equilibrium state depends on the shape of the fitness curve (Charlesworth and Charlesworth 1983). The accumulation of TEs is supposed to be deleterious, and the fitness of an individual depends on the number of copies carried by its genome: the higher the copy number, the lower the fitness. However, an equilibrium can be achieved only if the fitness function is log-concave, i.e., if $\partial \log w_n / \partial n > 0$. The graph presents the shape of three different fitness functions, all based on the formula $w_n = 1 - s \cdot n^t$, which has been often used because its shape depends only on the parameter t: each insertion decreases the fitness by the same value ("additive model" with $t = 1$, *thick dotted line*), the absolute effect of insertion decreases during the invasion ($t = 0.8$, *continuous line*), or each new insertion is more deleterious than the previous ones ("multiplicative model", $t = 1.2$, *thin dotted line*). These different selection models may correspond to different mechanisms known to be related to TE-mediated mutations (Nuzhdin 1999). If the main cause of the deleterious effects of TEs relies in insertion effects (e.g., disruption of coding or regulatory sequences), the linear model could be likely. On the other hand, if the major part of the TE-induced genetic load correspond to chromosomal abnormalities due to ectopic recombinations between TE copies, the multiplicative model could be more appropriate, since the frequency of recombinations probably increases with the square of the copy number (Langley et al. 1988). The respective weights of these different factors are still poorly known (see Le Rouzic and Deceliere 2005 for review)

Each new element that colonizes the genome of a species derives from a closely related TE sequence coming from the same genome or from the genome of another species. Genomes are full of inactive or deleted TE copies, which can potentially recombine and generate a new, functional TE sequence. However, most TE invasions seem to be related to interspecific horizontal transfers (HTs), which remain anecdotal for eukaryotic "standard" genes (Davis and Wurdack 2004; Kurland et al. 2003), but much more frequent in TE evolution. Indeed, TEs are generally thought to show an amazing ability to "jump" between species (Kidwell 1992), whatever the phylogenetic distances between them (closely related *Drosophila*, Silva et al. 2004; Sanchez-Gracia et al. 2005, or different lineages of vertebrates, Leaver 2001).