
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

¹Life 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/Eldredge’s Punctuated Equilibrium². 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

²Originally 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).

³In 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 inheritance⁵ 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.

⁵See 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 SINE⁶ retroelements as a model system of real novel insertions of transposable

⁶Short 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

Transposons and the Dynamic Genome

Lankenau, D.; Volff, J.-N. (Eds.)

2009, XV, 184 p., Hardcover

ISBN: 978-3-642-02004-9