

The Dynamic Eukaryote Genome: Evolution, Mobile DNA, and the TE-Thrust Hypothesis

This Thesis is presented for the award of Doctor of Philosophy
Murdoch University, Perth, Australia

BY

Keith Robert Oliver

BA (Philosophy) Edith Cowan University
BSc (Multidisciplinary Science) Curtin University
BSc (Hons 1) (Biological Science) Murdoch University

DATE: May 23 2012

Declaration that this Thesis is My Own Work

This Thesis is entirely my own original work, and contains no previously published material, or material written by any other person, except where due reference has been clearly indicated in the text.

I have clearly stated any contribution of others to my thesis as a whole, including any research work or hypotheses not entirely attributable to me. The content of this Thesis is the result of work I have carried out since the commencement of my Research Higher Degree candidature, and does not include any significant parts of work that I have submitted to qualify for the award of any other Degree or Diploma in any University or other Tertiary Institution.

Signed: Keith Robert Oliver.....

Date.....

The Dynamic Eukaryote Genome: Evolution, Mobile DNA, and the TE-Thrust Hypothesis

Abstract

The discovery of transposable elements (TEs) by Barbara McClintock in the 1940s, triggered a new dawning in the development of evolutionary theory. However, similar to Gregor Mendel's development of the laws of heredity in the nineteenth century, it was a long time before the full significance of this discovery was appreciated. Nevertheless, by the beginning of the 21st century, the study and recognition of TEs as significant factors in evolution was well underway. However, many evolutionary biologists still choose to ignore them, to highlight the loss of fitness in some individuals caused by TEs, or concentrate on the supposed parasitic nature of TEs, and the diseases they cause.

The major concept and theme of this thesis is that the ubiquitous and extremely ancient transposable elements are not merely "junk DNA" or "selfish parasites" but are instead 'powerful facilitators of evolution'. They can create genomic dynamism, and cause genetic changes of great magnitude and variety in the genotypes and phenotypes of eukaryotic lineages.

A large variety of data are presented supporting the theme of TEs as very significant forces in evolution. This concept is formalised into a hypothesis, the TE-Thrust hypothesis, which explicitly

Abstract

presents detail of how TEs can facilitate evolution. This hypothesis opens the way to explaining otherwise inexplicable aspects of evolution, such as the mismatch between the phyletic gradualism theory, and the punctuated equilibrium concept, which is based on the fossil record.

Data from the studies of many metazoans are analysed, with a focus on the well studied mammals, especially the primates. Data from the seed plants are also included, with a strong focus on Darwin's 'abominable mystery', the rapid origin, and the extraordinary success of the flowering plants.

TEs are ubiquitous and many of them are extremely ancient, probably dating back to the origin of the eukaryotes, and some are also found in prokaryotes. TEs can build, sculpt and reformat genomes by both active and passive means. Active TE-Thrust is due to transpositions by members of the TE consortium, or their retrotransposition of retrocopy genes, or by new acquisitions of TEs, or by the endogenisation of retroviruses, and other similar phenomena. Major results of this are that the promoters carried by TEs can result in very significant alterations in gene expression, and that sequences from the TEs themselves can become exapted or domesticated as novel genes. TEs can also cause exon shuffling, possibly building novel genes. Passive TE-Thrust is due to large homogenous consortia of inactive TEs that can act passively by causing ectopic recombination, resulting in genomic deletions, duplications, and possibly karyotypic changes. TE-Thrust often works together with other facilitators of evolution, such as point mutations, which can occur in duplicated, or

Abstract

retrocopy genes, sometimes resulting in new functions for such genes.

A major concept in the TE-Thrust hypothesis is that although TEs are sometimes harmful to individuals, and can lower the fitness of a population, they endow the lineage of that population with adaptive potential and evolutionary potential. These are extremes of a continuum of intra-genomic potential, and are not separate entities. This adaptive/evolutionary potential due to the presence and activities of the TE consortium of the genomes in a lineage, greatly enhance the future survival prospects of the lineage, and its ability to undergo evolutionary transitions, and/or to radiate into a clade of multiple divergent lineages. Lineages may acquire a TE consortium by new infiltrations of TEs, either by horizontal transposon transfer, *de novo* synthesis, or endogenisation of retroviruses. Lineages lacking an effective TE consortium are likely to lack adaptive/evolutionary potential and could fail to diversify, become “living fossils”, or even become extinct, as many lineages ultimately do.

The opposite of extinction is the fecund radiation of lineages, and it is shown here that fecund species-rich lineages such as rodents (Order Rodentia) and bats (Order Chiroptera) and the angiosperms, are all well endowed with many viable active TEs. The Simian Primates which have undergone major evolutionary transitions are also well endowed with viable and periodically active TEs, and/or large homogenous populations of TEs. Data on the “living fossils” such as the coelacanth and the tuatara are very

Abstract

limited, but indicate a lack of new acquisitions of TEs, and/or the mutational decay of ancient TE families in their genomes.

Lineages are often in stasis, but a new acquisition of TEs, or other factors such as stress, hybridisation, or whole genome duplications (especially in angiosperms) may trigger a major burst of activity in the TE consortium, resulting in an evolutionary punctuation event. The TE-Thrust hypothesis thus offers an explanation for the punctuated equilibrium, frequently observed in the fossil record.

There are many other known facilitators of evolution, such as point mutations, whole genome duplications, changes in allele frequency, epigenetic changes, symbiosis, hybridisation, simple sequence repeats, karyotypic changes, drift in small populations, allopatric and sympatric reproductive isolation, co-evolution, environmental and ecological changes, and so on. In addition, there may be some as yet unknown facilitators of evolution. However, TEs usually make up between 20 to 80 percent of the genomes of eukaryotes, as against one or two percent of coding genes, and are known to be able to make genomic modifications (“mutations”) that cannot be made by other facilitators of evolution. TEs also come in many superfamilies, and in thousands of families, which make up the mobile DNA of the earth’s biota. It is apparent then that their influence on, and facilitation of, eukaryotic evolution has been very significant indeed. In this thesis data are presented, which indicate that these ubiquitous and extremely ancient TEs are powerful facilitators of change, essential to the evolution of the earth’s biota.

Abstract

The TE-Thrust hypothesis, when fully explored, developed, and tested, if confirmed, must result in an extension to the Modern Synthesis, or even become a part of a new paradigm of evolutionary theory.

Table of Contents

Acknowledgements.....	-1-
List of Published Manuscripts.....	-3-
Statement of Contributions to Jointly Authored Works	
In this Thesis.....	-5-
Abbreviations used in this Thesis.....	-7-
Chapter One: General Introduction.....	1
Chapter Two: Transposable Elements: Powerful Facilitators	
of Evolution.....	17
Appendix to Chapter Two: Biological Evidence Supporting	
Punctuated Equilibrium Type Evolution.....	51
Chapter Three: Mobile DNA and the TE-Thrust Hypothesis:	
Supporting Evidence from the Primates.....	62
Chapter Four: The TE-Thrust Hypothesis and Plants: Darwin’s	
“Abominable Mystery” and Other Puzzles.....	113
Appendix to Chapter Four: TE-Thrust and Evolution,	
Devolution, and Background Extinction in the Metazoans..	151
Chapter Five: Transposable Elements and Viruses as Factors	
in Adaptation and Evolution: an Expansion and	
Strengthening of the TE-Thrust Hypothesis.....	165

Table of Contents

Chapter Six: General Discussion.....210

Combined Bibliography.....230

Appendices Title Page.....-1-

**Appendix One: The Genomic Drive Hypothesis and
Punctuated Evolutionary Taxonations or Radiations.....-2-**

Appendix Two: Jumping Genes Drive Evolution.....-15-

**Appendix Three: Jumping Genes: How they Drove
Primate Evolution.....-23-**

Addendum.....-34-

Acknowledgements

I am very grateful for the continuing support, guidance, and encouragement I have received from my supervisors, Associate Professor Wayne Greene, Emeritus Professor Jen McComb, and the philosopher, Dr Alan Tapper, and also to Professor Stuart Bradley early in my studies, before other duties precluded him from continuing in a supervisory role.

I am especially grateful to Jen McComb who has been an encouraging friend, as well as a supervisor, and to Wayne Greene who shared his vast store of knowledge of molecular biology with me, in addition to his supervisory role. In addition I thank Wayne Greene for the PowerPoint diagrams he did for me.

I thank the talented Margot Wiburd very much for her persistence and dedication in sorting out some computer and setting out problems, which I encountered along the way.

The contributions to my rapidly growing knowledge in the subject matter of this Thesis, by overseas email contacts, have also been outstanding, and I am especially grateful to Frank Ryan and Luis Villarreal, but also to David King, and George Parris, as well as Norihiro Okada, Cedric Feschotte, and several others.

Acknowledgements

Fay Uhe, together with some of my family, unfortunately, endured a measure of neglect, while I was busy with this study. I thank Fay, and my family, for their tolerance and understanding.

I am grateful to the Council of the Royal Society of Western Australia for enabling me to organise, together with Vic Semeniuk, a Symposium on Evolutionary Biology, in 2009, which marked the 150th Anniversary of the publication of the 'Origin of Species', the 200th Anniversary of the publication of Lamarck's '*Philosophie Zoologique*', and the 200th Anniversary of the birth of Darwin. A variety of aspects of, and opinions on, evolutionary biology were presented at this very successful and well attended event.

I have long had a passionate interest in evolutionary theory, and I am grateful to all the people who have discussed this subject with me, both the orthodox, where the late Julian Ford is remembered, and the unorthodox, where Ted Steele comes to mind. I also thank the more neutral microbiologist Graham O'Hara for some very interesting discussions.

Finally I am grateful for my long association with Murdoch University, and the School of Biological Sciences and Biotechnology, with whom I have been a mature-age undergraduate student, an honorary research associate, an adjunct lecturer, and finally a post-graduate student. This has been a fruitful association indeed.

Keith Oliver, 2012

List of Published Manuscripts

Oliver K. R. and Greene W. K. (2009a) Transposable elements: powerful facilitators of evolution. *BioEssays*. 31(7): 703-714.

Oliver K. R. and Greene W. K. (2009b) The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations. *Journal of the Royal Society of Western Australia*. 92: 447-451.

Oliver K.R. and Greene W. K. (2009c) Jumping Genes Drive Evolution. *Australasian Science* 30(8): 29-31.

Oliver K. R. and Greene W. K. (2011) Mobile DNA and the TE-Thrust Hypothesis: Supporting Evidence from the Primates. *Mobile DNA* 2011, 2:8

Oliver K. R., Greene W. K. (2012a) Transposable Elements and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis. This has been submitted for publication.

Oliver K. R. and Greene W. K. (2012b) Jumping Genes: How They Drove Primate Evolution. *Australasian Science* 33 (1): 18-21.

List of Published Manuscripts

Oliver K R, McComb J A and Greene W K (2012) The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles. This is being reformatted and submitted for publication.

Statement of Contributions to Jointly Authored Works, Contained in this Thesis

The manuscripts included in this Thesis (Chapters 2 to 5), and the Appendices 1, 2 and 3 were joint contributions. Specific contributions to the work, analysis or synthesis, and/or writing and editing, of these manuscripts were as follows:

(Chapter 2) Oliver K. R. and Greene W. K. (2009) Transposable elements: powerful facilitators of evolution. *BioEssays*. 31(7): 703-714.

Keith Oliver was responsible for conception and theoretical design, analysis and interpretation of data, manuscript writing and editing. Wayne Greene, as supervisor, contributed to theoretical design, analysis and interpretation of data, manuscript writing and editing. This also applies to Chapters 3 (published) and 5 (submitted for publication), and Appendices 1, 2, and 3 (all published).

(Chapter 4) Oliver K R, McComb J. A. and Greene W K. The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles. (This is ready for reformatting for publication).

Keith Oliver was responsible for conception and theoretical design, analysis and interpretation of data, manuscript writing and editing. Jen McComb and Wayne Greene as supervisors, contributed to theoretical design, analysis and interpretation of data, manuscript writing and editing.

Statement of Contributions to Jointly Authored Work in this Thesis

(Chapter 3) Oliver K. R. and Greene W. K. (2011) Mobile DNA and the TE-Thrust Hypothesis: Supporting Evidence from the Primates. *Mobile DNA* 2011, 2:8

(Chapter 5) Oliver K. R., Greene W. K. (2012) Transposable Elements and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis.

(Appendix 1) Oliver K. R. and Greene W. K. (2009) The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations. *Journal of the Royal Society of Western Australia*. 92: 447-451.

(Appendix 2) Oliver K.R. and Greene W. K. (2009) Jumping Genes Drive Evolution. *Australasian Science* 30(8): 29-31.

(Appendix 3) Oliver K. R. and Greene W. K. (2012) Jumping Genes: How They Drove Primate Evolution. *Australasian Science* 33 (1): 18-21.

Additional work of relevance not included in this Thesis: Semeniuk V. and Oliver K. R. Proceedings of an Evolutionary Biology Symposium (various speakers and authors) held in October 2009, convened by V. Semeniuk and K. R. Oliver. Evolutionary Biology Symposium (2009) *The Journal of the Royal Society of Western Australia*. 92(4) 1-487.

Abbreviations used in this Thesis

- Alu: a SINE specific to the primates.
- ARMD: Alu-Recombination-Mediated Deletion.
- DNA-TE: DNA Transposable Element, or Transposon (Class II TE).
- EBN: Endosperm Balance Number.
- ECR-LTR: Envelope-Class Retrovirus-like LTR retro-TE.
- ERV: Endogenous RetroVirus.
- ERVs/sLTRs: Endogenous Retroviruses and/or solo Long Terminal Repeats.
- EVE: Endogenous Viral Element.
- HTT: Horizontal Transposon Transfer.
- ITR: Inverted Terminal Repeat.
- LINE: Long INterspersed Element.
- L1 or LINE-1: (autonomous) Long INterspersed Element 1.
- LTR: Long Terminal Repeat.
- LTR retro-TE: Long Terminal Repeat retro-TE.
- MIR (SINE): ancient Mammalian-wide Interspersed Repeat.
- MITE: (non autonomous) Miniature Inverted-repeat Transposable Element.
- Mya: Million years ago.
- Myr: Million years.
- retro-TE: Retrotransposable Element or Retroposon (Class I TE).
- RT: Reverse Transcriptase.
- ORF: Open Reading Frame.
- RIP: Repeat-Induced Point mutation.
- SINE: Short INterspersed Element.

Abbreviations used in this Thesis

sLTR: solo Long Terminal Repeat.

sLTR/ERV: solo LTR and/or an ERV.

SVA: SINE-VNTR-Alu chimaeric retro-TE.

TE: Transposable Element, transposon, or retroposon.

TEd-alleles: alleles deactivated or destroyed, by TE insertions.

TE_m-alleles: alleles modified in either function or regulation, or duplicated, by TE insertions.

TSD: Target Site Duplication.

UTR: UnTranslated Region.

VIMS: Variation Inducing Mobile Sequences.

VNTR: Variable Number Tandem Repeat.

Chapter 1

General Introduction

1.1 Barbara McClintock's Transposable Elements

Major discoveries, by Gregor Mendel, and by Barbara McClintock, were two very important breakthroughs that were very slow in gaining recognition (Fedoroff 1999). When Gregor Mendel carried out his experiments on crossing different lines of peas (*Pisum sativum*), and formed the concepts that have become the basis of modern genetics he was unknowingly investigating a genomic modification due to a transposable element (TE). The difference between the dominantly inherited full round seeds, and the recessively inherited wrinkled seeds, is due to the insertion of a TE, similar to Ac/Ds in maize, into the *SBEI* gene for a starch-branching enzyme, which reduces starch synthesis, and results in wrinkled yellow seeds (Bhattacharyya et al. 1990), as indicated in Figure 1-1. However, Mendel's paper was ignored for 35 years, even though it contained insights essential for an understanding of genetics. Similarly, Barbara McClintock discovered TEs in the 1940s (McClintock 1950; 1956; 1984) and it took another 30 years and more, for the significance of her finding to be appreciated, when TEs were eventually recognised as creators of genomic variation, on a large and multifaceted scale, that was unimaginable to the contemporary biologists prior to this (Kidwell and Lisch 1997).

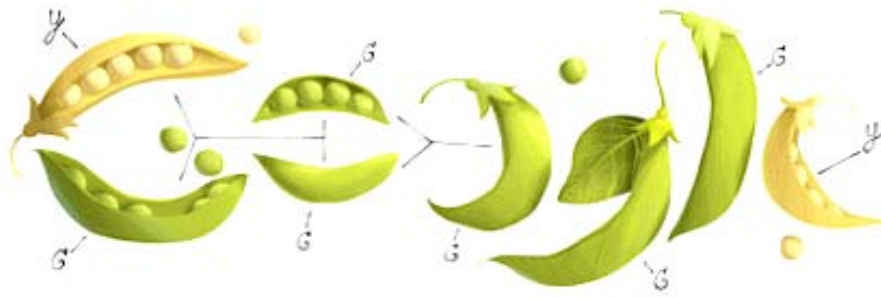


Figure 1-1. Google's 2010 Tribute to Gregor Mendel, who unknowingly used a pea plant with a TE insertion in the *SBE1* gene (Bhattacharyya 1990), in his derivation of the laws of inheritance. Although long ignored, his work eventually resulted in big advances in evolutionary theory, which are relevant to the present day.

McClintock won the Nobel Prize in 1983, largely for the introduction of a completely new concept in which chromosomes were no longer considered to be rigid structures, but to be flexible, thus allowing reorganisations of the genetic material. This reorganisation is catalysed by transposable elements (which she called 'controlling elements'), which not only jump from one place to another in the genome, but can also influence the expression of other genes (Comfort 2001a). The prescient nature of this concept has been repeatedly confirmed.

'Late in life, she synthesised her life's work into a vision of the genome as a sensitive organ of the cell, capable of rearranging itself in response to environmental cues.' (Comfort 2001b). This vision is supported by Shapiro (2010) who states that McClintock's studies taught her that maize had the ability to detect X-ray induced broken ends of chromosomes, bring them together

and fuse them to generate novel chromosome structures, including deletions, inversions, translocations, and rings.

At the conclusion of her Nobel Prize lecture McClintock said 'In the future, attention will undoubtedly be centred on the genome, with greater appreciation of its significance as a highly sensitive organ of the cell that monitors genomic activities and corrects common errors, senses unusual or unexpected events, and responds to them, often by restructuring the genome.'

However, long before McClintock's 1983 Nobel Prize, in an influential paper Britten and Davidson (1971) called TEs 'repetitive DNA sequences', as it was not known at this time that they could transpose. However, among many other things, they state that the incorporation of repeated DNA into the genome does not seem likely to be a continuous process, but occurs as sudden (on an evolutionary timescale) replication events. They give an example of a ~300 bp sequence that is present in a million copies in the mouse, but only <50 copies in the closely related rat. From this they deduce that this highly repetitive sequence family must have been produced in a relatively sudden event since the lineages leading to these species separated within the last few million years. Further to this they suggest that a possible mechanism of such saltatory replication could be the integration into the genome of many copies of a viral genome or viral-borne sequence. The prescient nature of such observations and speculations will become evident later, in the body of this thesis.

1.2 A Selection of Early Publications Suggesting or Stating the Value of TEs in Adaptation and Evolution

Chapter 1: General Introduction

As indicated by the following selection, the study and appreciation of at least a possible role for TEs in adaptation and evolution was well underway by 1999-2000, despite the negative assessments of TEs by Orgel and Crick (1980) and Doolittle and Sapienza (1980).

Table 1-1 Early Contributions on the Value of TEs in Adaptation and Evolution

Date	Authors	Publication details
1982	Schmidt CW & Jelinek WR	The Alu family of dispersed repetitive sequences. <i>Science</i> 216: 1065-1070.
1984	Ginzburg LR. Bingham PM & Yoo S	On the theory of speciation induced by transposable elements. <i>Genetics</i> 107: 331-341.
	Georgiev GP	Mobile genetic elements in animals and their biological significance. <i>European Journal of Biochemistry</i> 145: 203-220.
1988	Wessler SR	Phenotypic diversity mediated by the maize transposable elements Ac and Spm. <i>Science</i> 242:399-405.
1989	Jurka Ja	Subfamily structure and evolution of the human L1 family of repetitive sequences. <i>Journal of Molecular Evolution</i> 29: 496-503.
	Jurka Jb	Novel families of interspersed repetitive elements from the human genome. <i>Nucleic Acids Research</i> 18: 137-141.
1990	Bhattacharyya MK, Smith AM, Ellis THN, Hedley C & Martin C	The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. <i>Cell</i> 60: 115-112.
	Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG & Chovnick A	Evidence for horizontal transmission of the P transposable element between <i>Drosophila</i> species. <i>Genetics</i> 124: 339-335.
	Dennis ES & Brettell RI	DNA methylation of maize transposable elements is correlated with activity. <i>Transactions of the Royal Society of London Biological Science</i> 326:217-219.
	Howard BH & Sakamoto K	Alu interspersed repeats: selfish DNA or a functional gene family? <i>New Biologist</i> 2: 759-770.
1991	Brettell RI & Dennis ES	Reactivation of a silent Ac following tissue culture is associated with heritable alterations in its methylation pattern. <i>Molecular General Genetics</i> 229: 365-372.
1992	Simmons GM	Horizontal transfer of hobo transposable

Chapter 1: General Introduction

		elements within the <i>Drosophila melanogaster</i> species complex: evidence from DNA sequencing. <i>Molecular Biology and Evolution</i> 9: 1050-1060.
1993	Bailey AD & Shen CK	Sequential insertion of Alu family repeats into specific genome sites of higher primates. <i>Proceedings of the National Academy of Science of the USA</i> 90: 7205-7209.
1995	McDonald JF	Transposable elements: possible catalysts of organismic evolution. <i>Trends in Ecology and Evolution</i> 10: 123-126.
1995	Robertson HM & Lampe DJ	Recent horizontal transfer of a <i>mariner</i> transposable element among and between Diptera and Neuroptera. <i>Molecular Biology and Evolution</i> 12: 850-862.
1995	Wessler SR, Bureau TE & White SE	LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. <i>Current Opinion in Genetics & Development</i> 5: 813-821.
1996/7	Jurka J, Kapitonov VV, Klonowski P, Walichiewicz J & Smit AF	Identification of new medium reiteration frequency repeats in the genomes of Primates, Rodentia and Lagomorpha. <i>Genetics</i> 3: 235-247.
1997	Capy P, Langin T, Higuete D Maurer P & Bazin C	Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? <i>Genetica</i> 100: 63-72.
	Kidwell M G & Lisch D	Transposable elements as sources of variation in animals and plants. <i>Trends in Ecology and Evolution</i> 15: 95-99.
	McFadden J & Knowles G	Escape from evolutionary stasis by transposon mediated deleterious mutations. <i>Journal of Theoretical Biology</i> 186: 441-447.
	Villarreal L P	On viruses, sex, and motherhood. <i>Journal of Virology</i> . 71; 859-865.
1998	Capy P	A plastic genome. <i>Nature</i> 396: 522-523
	Jurka J	Repeats in genomic DNA: mining and meaning. <i>Current Opinion in Structural Biology</i> 8: 333-337.
1999	Bénit L, Lallemand J-B, Casella J-F, Philippe H & Heidmann	ERV-L elements: A family of endogenous retrovirus-Like elements active throughout the evolution of mammals. <i>Journal of Virology</i> 73: 3301-3308.
	Borodulina OR & Kramerov DA	Wide distribution of short interspersed elements among eukaryotic genomes. <i>FEBS Letters</i> 457: 409-413.
	Brosius J	Retroposons - seeds of evolution. <i>Science</i> 251:753.
	Dimitri P & Junakovic N.	Revising the selfish DNA hypothesis, new evidence on accumulation of transposable elements in heterochromatin. <i>TIG</i> 15: 123-124.
	Fedoroff N	Transposable elements as a molecular evolutionary force.
	Gilbert N & Labuda	CORE-SINES: Eukaryotic short interspersed

Chapter 1: General Introduction

	D	retroposing elements with common sequence motifs. Proceedings of the National Academy of Science of the USA 96:2869-2874.
	Jurka J & Kapitonov VVa	Sectorial muragenesis by transposable elements. Genetica 107: 239-248.
	Kumar A & Benetzen JL	Plant Retrotransposons. Annals New York Academy of Sciences 251-264.
	Matzke MA, Mette FM, Aufsatz W, Jakowitsch J & Matzke AJM	Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. Genetica 107: 271-287.
	Shapiro JA	Transposable elements as the key to a 21 st Century view of evolution. Mobile DNA 1:4.
2000	Benetzen JL	Transposable element contributions to plant gene and genome evolution. Plant Molecular Evolution 42: 251-269.
	Casavant NC, Scott LA, Cantrell MA, Wiggins LE, Baker RJ & Wichman HA	The End of the LINE?: Lack of Recent L1 Activity in a Group of South American Rodents. Genetics 154: 1809-1817.
	Fedoroff N	Transposons and gene evolution in plants. Proceedings of the National Academy of Science of the USA 97: 7002-7007.
	Kidwell MG & Lisch DR	Transposable elements and host genome evolution. Proceedings of the National Academy of Science of the USA 94: 7704-7711.
2001	Borodulina OR & Kramerov DA	Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mammalian Genome 12: 779-786.
	Kidwell MG & Lisch DR	Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55: 1-24.

1.3 More recent Papers Suggesting or Affirming Contributions to Adaptation and Evolution by TEs

In recent years there has been a veritable explosion of published papers describing TEs (Figure 1-2) and exploring the possible roles of TEs in evolution. A large number of these papers are cited throughout the major chapters (Chapters 2 to 5) of this Thesis. It is harder now to find recent authors anthropomorphising TEs as 'selfish or parasitic DNA', and 'ultimate parasites' as Orgel and Crick (1980) and Doolittle and Sapienza (1980) did, although some still emphasise their lowering of fitness (Vinogradov 2003;

Pasyukova 2004). However, there does not seem to be a great deal else published recently about TEs lowering fitness, except in asexuals (Arkhipova and Meselson 2004) and in prokaryotic lineages (Rankin et al. 2010), but only by mathematical modelling in the Rankin paper, not by experimental results, or by empirical findings. TEs do, however, have the potential to cause diseases in individuals (Deininger and Batzer 1999; Bacolla and Wells 2009; O'Donnell and Burns 2010; Baillie et al. 2011; Beck et al. 2011), but this is more than offset by the continuum of 'intra-genomic potential' ('adaptive potential', and 'evolutionary potential') benefits they bestow on the lineages to which these individuals belong (Oliver and Greene 2009a; b; 2011). Unfortunately, many evolutionary biologists appear to go on ignoring TEs altogether, and concentrate on the coding genes, in all of their concepts regarding evolutionary theory. As TEs make up 45% of the human genome, and only one or two per cent consists of coding genes, this suggests that some changes would be helpful in a reformulation of evolutionary hypotheses or theories.

1.4 The Recent Formulation of the TE-Thrust Hypothesis

Many papers have indeed been published in recent years on the possible role of TEs in evolution. However, it is only in the Chapters 2 to 5 included in this thesis that a role for TEs in evolution has been formalised into a definite hypothesis, the 'TE-Thrust hypothesis', and four modes of TE-Thrust proposed. Furthermore, 'Adaptive Potential' and 'Evolutionary Potential', as extremes of a continuum of 'intra-genomic potential' due to TE activity, have been posited, and have been assessed, with the finding of much significant data which suggest support for the TE-Thrust hypothesis.

In the Modern Synthesis it seems that empirical growth has almost always exceeded theoretical predictions and most discoveries have not been predicted by theory, but have come as complete surprises (Federoff 2000). For example, the discovery of transposable elements (TEs), the very large introns of eukaryote genes, and reverse transcription (RNA to DNA), were seemingly complete surprises. In contrast to this, the TE-thrust hypothesis, although only a hypothesis, is not based on the *a priori* assumptions of population genetics or the Modern Synthesis, but is derived purely from empirical data.

The rapidly growing knowledge of TEs and the possible role they could play in adaptation and evolution has forced a move from the concept of a static eukaryotic genome, wherein all genes occupied their specific immovable locus, to the concept of a dynamic eukaryotic genome. In these dynamic genomes TEs make a significant contribution both to phenotypic adaptation, and to phenotypic evolutionary transitions, radiations, and evolutionary novelties. Such changes often occur in a punctuated equilibrium manner (Oliver and Greene 2009a; b; Parris 2009; Zeh et al. 2009).

The concept that TEs are “selfish parasitic DNA” comes mainly from the rather speculative essays of Orgel and Crick (1980) and Doolittle and Sapienza (1980). These papers did rightly perhaps liberate us from the then prevalent notion that it was phenotypic selection that optimised genome structure. Unfortunately, however, the concept of TEs being “selfish” and “parasitic” did become an impediment to the study of the historical and

Chapter 1: General Introduction

contemporary contributions of TEs to chromosome structure (Federoff 2000) and to the processes of evolution. From humble beginnings via the groundbreaking work of Barbara McClintock (1950; 1956; 1984), and pioneers such as Ginzburg et al. (1984), it is now becoming accepted by many biologists that TEs are powerful facilitators of evolution, as we proposed in our peer reviewed published reviews and syntheses of earlier work (Oliver and Greene 2009a (Chapter 2), and 2009b (Appendix 1), 2011 (Chapter 3), and in Chapters 4, and 5. (Chapter 4 is being prepared for submission for publication, and Chapter 5 has been submitted for publication). It is hard now to see how anyone could deny that Transposable Elements are indeed powerful facilitators of evolution, but many still seem to ignore the possibility of even a small role for TEs in evolution.

1.5 A Brief Initial Outline of the TE-Thrust Hypothesis

(This brief outline of the hypothesis, shown diagrammatically in Figure 1-3, is developed much more fully in Chapter 6, from the data available in Chapters 2 to 5)

1.5.1 Posit (1): Transposable Elements (TEs) are ubiquitous and many are extremely ancient, although some of them are of recent origin (<100 Myr). They are not merely “junk”, or “parasitic DNA”, but are mostly beneficial to lineages, and are potentially, powerful facilitators of evolution.

1.5.2 Posit (2): TEs can cause genetic changes of great magnitude and variety within genomes, making genomes flexible and dynamic, so that they drive their own evolution and the evolution of their phenotypes.

Chapter 1: General Introduction

1.5.3 Posit (3): TEs can cause many genomic modifications that cannot be caused by any other mutagens.

1.5.4 Posit (4): TEs can greatly modify single gene regulation, and can modify the regulation of networks of genes.

1.5.5 Posit (5): TE-Thrust can build, sculpt, and reformat genomes by both active and passive means. Active Genomic Drive is due to the active transposition of TEs from either a heterogenous or homogenous population of TEs. Passive Genomic Drive is due to ectopic recombination between homologous TE insertions. Such ectopic recombinations are common only when there are large homogeneous populations of TEs.

1.5.6 Posit (6): TE-Thrust, via intermittent bursts of TE activity sometimes results in macro-evolutionary punctuation events in lineages in stasis, or gradualism; these often result in a drive towards novelty, diversity, or complexity and a radiation of species. This is punctuated equilibrium

1.5.7 Posit (7): Successful lineages do not destroy TEs, but there are strong genomic controls on transposition of TEs in the soma where TEs are potentially damaging. However, there is less control of TE activity in the germ line and the early embryo in mammals, where their activity can generate both potentially useful and deleterious variation in the progeny. Useful variants then increase, often to fixation, and

Chapter 1: General Introduction

Figure 1-2: Simplified Diagrammatic Representation of some Major Orders, Classes and Superfamilies of TEs (not to scale)

Class I elements (Retrotransposons)

Autonomous retrotransposons

LTR e.g. endogenous retrovirus (<10 kb)



Non-LTR e.g. LINE-1 (<6 kb)



Non-autonomous retrotransposons

SINE (7SL-related) e.g. Alu (0.3 kb)



SINE (tRNA-related) e.g. MIR (0.26 kb)



Composite e.g. SVA (<3 kb)



Class II elements (DNA Transposons)

Autonomous DNA transposons

Transposon e.g. Mariner (1.4 kb)



Non-autonomous DNA transposons

MITE e.g. Made1 (0.08 kb)



LTR, long terminal repeat; GAG, group-specific antigen; POL, polymerase; ENV, envelope protein; RT, reverse transcriptase; EN, endonuclease/integrase; UTR, untranslated region; polyA, polyA addition site; ORF, open reading frame; VNTR, variable number of tandem repeats; SINE-R, domain derived from a HERV-K; TIR, terminal inverted repeat. Black arrows, RNA polymerase II promoter (double arrows denote bidirectionality); red arrows, RNA polymerase III promoter.

deleterious variants decrease or are eliminated in future generations of the lineage, by means of natural selection.

1.5.8 Posit (8): Although sometimes harmful to some individuals, TEs can be very beneficial to lineages. There is, then, a differential survival of lineages with those lineages endowed with a suitable consortium of viable TEs being more likely to survive and to radiate or proliferate, as such lineages have enhanced adaptive potential and enhanced evolutionary potential.

1.5.9 Posit (9): Clades or lineages deficient in viable TEs, and with heterogenous populations of non-viable TEs, tend to be non-fecund, can linger in prolonged stasis, and eventually may become “living fossils” or become extinct. Conversely, clades or lineages well endowed with viable and active TEs, especially if the TEs are homogenous, tend to be fecund, or species rich, and taxonate readily.

1.5.10. There is ample evidence of punctuated equilibrium type evolution in the fossil record (Eldredge and Gould 1972; Stanley 1981; Eldredge 1986; Gould 2002), and there is independent evidence for it from extant lineages (Appendix to Chapter 2).

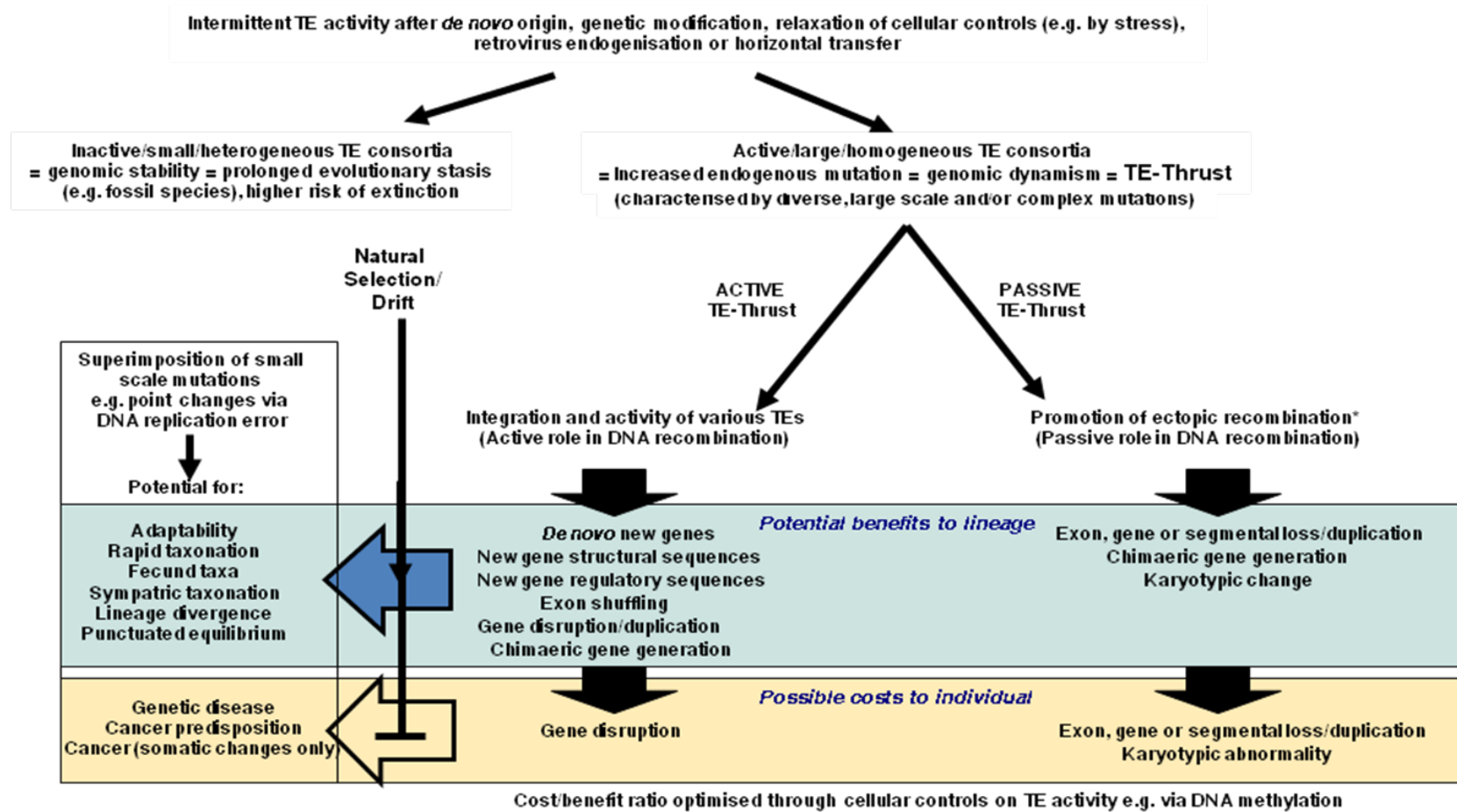
1.5.11. The TE-Thrust hypothesis has been derived from, and is supported by, the study of peer reviewed published empirical data on mammalian evolution, and to a lesser extent, angiosperm and insect evolution. The applicability, or otherwise, of this hypothesis to other Classes, and to other Phyla, needs much further study in the future.

- 1.5.12. There is also some limited support for the TE-Thrust hypothesis, from the very sparse data from the “living fossils”, such as the tuatara, and the gymnosperm *Ginkgo biloba*.
- 1.5.13. Some support for the TE-Thrust hypothesis is also found in the “enduring stasis” of “living fossil lineages” like the lobe finned lungfish and the coelacanth lineages (Appendix to Chapter 4).
- 1.5.14. The TE-Thrust hypothesis is put forward to complement and supplement other accepted, hypothesised, or possibly as yet unknown, facilitators or mechanisms of evolution, and not to diminish or deny the validity of any other such facilitators or mechanisms of evolution.
- 1.5.15. “Thrust” should not be understood in any teleological sense. There is no implication that TE-Thrust is pushing the evolution of lineages to some predetermined goal.

Non-viable and Non-functional

TEs are here designated as ‘non-viable’ when they are incapable of transposition, often due to mutations in open reading frames (ORFs) and are designated as ‘non-functional’ when they are so corrupted by mutations that they lack enough homology for ectopic recombination with others of their same kind.

Figure 1-3: Diagrammatic Representation of the TE-Thrust Hypothesis



*e.g. unequal crossover/sister chromatid exchange, chromosomal rearrangement

1.6 The Structure of This Thesis

After this introduction, the main body of this Thesis consists of four papers (Chapters 2, 3, 4, and 5), of which Chapters 2 and 3 have been peer reviewed and published, and Chapter 5 has been submitted for peer review and publication. A paper based on Chapter 4 will be submitted for peer review and publication. The body of the Thesis concludes with a General Discussion in Chapter 6. There are also three published Appendices, relevant to the thesis, the first of which has been peer reviewed and published, with the other two being published without peer review. There is also a short unpublished Addendum, of a more personal nature.

1.7 The Objectives of this Thesis

The main objective of this Thesis is to propose the TE-Thrust hypothesis in great detail, and to assess much of the available evidence as to whether or not this hypothesis is likely to be largely correct, as it is presented here, in essence, if not in all of the finer details. Investigating the specific value all of, or parts of, the TE-Thrust hypothesis should stimulate much further research, as more and more data become available. Although I believe much data suggests that the TE-Thrust hypothesis is largely correct, I abide by the statement of C. Stuart Gager (1910): 'Hypotheses are not statements of truth, but instruments to be used in the ascertainment of truth. Their value does not depend upon ultimate verification, but is to be measured by their effects upon scientific research'. It is my hope that the TE-Thrust hypothesis will make a valuable contribution to stimulating research, and to initiating new

schools of thought, that together with other newly discovered phenomena, new data, and new hypotheses, will result in an extension of the Modern Synthesis, or its replacement with a new paradigm of evolutionary theory.

Many authors have written implying the need for an extended Modern Synthesis, and many have gone much further, implying, or explicitly stating, the need for a new paradigm in evolutionary theory. A selection of examples of these are: (Dover 1982; Margulis 1991; Bussel and James 1997; Margulis and Chapman 1998; Steele et al. 1998; Shapiro 1999; Ryan 2002; Shapiro and von Sternberg 2005; Villarreal 2005; Caporale 2006; Pigliuchi and Kaplan 2006; Ryan 2006; Wessler 2006; Ryan 2007; Ryan 2009; Steele 2009; Shapiro 2010; Villarreal and Witzany 2010).

Just how big these proposed changes to the Modern Synthesis will have to be is not clear at present, but in this Thesis I present data that suggests that change will surely come.

Chapter 2

Transposable Elements: Powerful Facilitators of Evolution

2.1 Summary

Transposable elements (TEs) are powerful facilitators of genome evolution, and hence of phenotypic diversity as they can cause genetic changes of great magnitude and variety. TEs are ubiquitous and extremely ancient and although harmful to some individuals, they can be very beneficial to lineages. TEs can build, sculpt and reformat genomes by both active and passive means. Lineages with active TEs, or with abundant homogeneous inactive populations of TEs that can act passively by causing ectopic recombination, are potentially fecund, adaptable, and taxonate readily. Conversely, taxa deficient in TEs or possessing heterogeneous populations of inactive TEs may be well adapted in their niche, but tend to prolonged stasis and may risk extinction by lacking the capacity to adapt to change, or diversify. Because of recurring intermittent waves of TE infestation, available data indicates a compatibility with punctuated equilibrium, in keeping with widely accepted interpretations of evidence from the fossil record.

2.2 Introduction

Over 50 years ago Barbara McClintock, the discoverer of transposable elements (TEs), (McClintock 1950) made the prescient suggestion that TEs have the capacity to re-pattern genomes (McClintock 1956). More recently, important work by

numerous others has provided support for this idea by characterizing TEs as potentially advantageous generators of variation upon which natural selection can act (Fedoroff 1999; Kidwell and Lisch 2001; Bowen and Jordon 2002; Deninger et al. 2003; Kazazian Jr. 2004; Brandt et al. 2005; Biémont and Vieira 2006; Volff 2006; Wessler 2006; Feschotte and Pritham 2007; Muotri et al. 2007; Böhne et al. 2008). Here, we review and add new perspectives to this data and bring many disparate strands of evidence into one holistic synthesis about how the presence of TEs within genomes makes them flexible and dynamic, so that genomes themselves are powerful facilitators of their own evolution. TEs act to increase the evolvability of their host genome and provide a means of generating genomic changes of greater variety and magnitude than other known processes. However, we acknowledge that endosymbiosis, horizontal gene transfer (especially in bacteria), endosymbiotic gene transfer, polyploidy (especially in plants), short tandem repeat slippage, point mutation, and other such phenomena are also very important in evolution. We argue that TE-generated mutations are very much complementary to these phenomena and are vital to evolution because TEs can bring about a myriad of substantial changes from sudden gene duplication events to rapid genome-wide dissemination of gene regulatory elements. TEs can even contribute coding and other functional sequences directly to the genome. Such large-scale mutations, when subjected to natural selection, can lead to major evolutionary change. This can have manifold benefits to a host lineage in terms of facilitating taxon radiation and adaptation to, or adoption of, new habitats, as well as facilitating survival when confronted with environmental or

biotic change or challenge. TEs maintain genomic evolvability in the long-term by lineage selection, that is, by the natural selection of those lineages whose genomes are endowed with a suitable repertoire of them. We thus move the focus of changes to the genome away from the fitness of the individual or the group, to the fitness or evolutionary potential of the lineage. We propose also that episodic surges of TE activity offer a ready explanation for punctuated equilibrium, as observed in palaeontology.

We accept that unfettered TE activity within a host genome could be catastrophic. However, in practice probably all organisms have evolved cellular TE control measures to minimize harm to their somatic cell DNA, while allowing some TE-generated genetic variation to be passed on to future generations via the germ line (Matzke et al. 1999; Schulz et al. 2006). As a further benefit, these cellular mechanisms appear to have been adopted to control normal gene expression on a genome-wide basis (Yoder et al. 1997; Buchon and Vaury 2006). All categories of TEs, especially in the past, have been considered to be genomic parasites (Doolittle and Sapienza 1980; Orgel and Crick 1980; Hickey 1982), but more recently, significant beneficial attributes for facilitating evolution have been recognized (Fedoroff 1999; Kidwell and Lisch 2001; Bowen and Jordon 2002; Deninger et al. 2003; Kazazian Jr. 2004; Wessler 2006; Brandt et al. 2005; Biémont and Vieira 2006; Volff 2006; Feschotte and Pritham 2007; Muotri et al. 2007; Böhne et al. 2008). In our view, TEs are almost essential for significant continuing evolution to occur in most organisms. If they are parasites then they are “helpful

parasites”, a view that is supported by a substantial and rapidly growing body of evidence.

2.3 TEs as Powerful Facilitators of Evolutionary Change

It is now generally accepted that the emergence of increasingly complex eukaryotic life forms was accompanied by a corresponding increase in genome complexity, entailing both an expansion in gene number and more elaborate gene regulation (Ohno 1970; Bird 1995; Carroll 2005). Only DNA recombination in the form of gene or segmental duplications, exon shuffling, insertions, deletions and chromosomal rearrangements, can adequately account for this increase in gene number and the complexity of their regulation (Ohno 1970; Bird 1995; Carroll 2005). TEs, which possess a number of striking features that make them suitable as general agents of genome and lineage evolution (Table 2-1), have played a crucial role by acting as mutagenic agents to either massively accelerate the rate at which such events occur, or by making them possible in the first place via *de novo* insertions, as in the origin of the jawed vertebrate adaptive immune system (Schatz 2004).

Table 2-1 Features of TEs that make them highly suitable as general agents of genotypic and phenotypic evolution, lineage selection and taxonation.

TE Feature	Comments
1) Mobile	TEs are grouped into two main classes based on their mode of transposition: Class I TEs or retrotransposons (retro-TEs)

	<p>transpose via an RNA intermediate.</p> <p>Autonomous elements: these encode reverse transcriptase (RT) e.g. endogenous retroviruses (ERVs), LTR retroelements, LINEs.</p> <p>Non-autonomous elements: e.g. SINEs, which lack an RT gene but nevertheless can retrotranspose using the transpositional machinery of LINEs.</p> <p>Class II TEs or DNA transposons (DNA-TEs) transpose directly and can do so via an encoded transposase enzyme or by utilising an alternative mechanism such as rolling-circle replication.</p> <p>Reference 1,2</p>

.. .References

<p>2) Universal</p>	<p>TEs have been found in all genomes, from bacteria to mammals. Their ubiquitous nature is due to their strong tendency to disseminate within a genome as well as colonise other genomes. Some TEs can arise spontaneously from non-transposable DNA sequences in the genome (e.g. SINEs), while others can be horizontally transferred between</p>	<p>3</p>
----------------------------	--	-----------------

	species.	
3) Ancient	<p>TEs have ancient origins that are traceable to prokaryotes.</p> <p>DNA-TEs are related to bacterial insertion sequences and retro-TEs are related to group II introns. Some TEs seem to have been present in eukaryotes from their earliest beginnings, possibly well over one billion years ago.</p>	4
4) Abundant	<p>TEs often comprise a large, if not massive, fraction of the eukaryotic genome. For example, the sequenced mammalian genomes are at least a third TE in origin in non-primates and around a half TE in primates.</p> <p>TEs appear to be an important determinant of genome size, with organisms possessing extra large genomes (<i>e.g.</i> plants) often having a very much higher TE content (>60%), compared to species with relatively small genomes, such as yeast, nematodes, insects and birds, which have a much lower TE content</p>	5,6,7-10 11,12,13

	(<20%).	
5) Beneficial	TEs are powerful mutagens that can be deleterious to some individuals – such mutations are eliminated by natural selection. However, beneficial mutations, which can often be highly advantageous to their lineage, are conserved by selection.	14-25

1 (Wicker et al. 2007); **2** (Kapitonov and Jurka 2008); **3** (Pace et al. 2008); **4** (Capy et al. 1997); **5** (Lander et al. 2001); **6** (Gibbs et al. 2007); **7** (Pontius et al. 2007); **8** (Lindblad-Toh et al. 2005); **9** (Waterston et al. 2002); **10** (International Chicken Genome Sequencing Consortium 2004); **11** (Gibbs et al. 2004); **12** (Mikkelsen et al. 2005); **13** (Kidwell 2002); **14** (Fedoroff 1999); **15** (Kidwell and Lisch 2001); **16** (Bowen and Jordon 2002); **17** (Deninger et al. 2003); **18** (Kazazian Jr. 2004); **19** (Wessler 2006); **20** (Brandt et al. 2005); **21** (Biémont and Vieira 2006); **22** (Volf 2006); **23** (Feschotte and Pritham 2007); **24** (Ohno 1970); **25** (Bhōne et al. 2008).

Table 2-2 Active mechanisms by which TEs generate the genetic novelty required for dramatic evolutionary change

Role	Comments	References
Direct Contribution to Gene/Genome Structure		
Entire genes	About 50 cases of “neogenes” whose sequences are largely TE-derived are known in the human genome <i>e.g. TERT, CENPB, RAG1/2.</i>	1,2,3

	TEs have made some extraordinarily complex evolutionary events take place that otherwise might not have occurred e.g. recombination signal sequences involved in V(D)J Ag receptor rearrangements. These, like the RAG1/2 recombinase genes themselves, appear to be derived from an ancient DNA-TE.	
Exons/partial exons	A substantial number of human genes harbour TEs within their protein-coding regions. TEs often form independent exons within genes, many of which are alternatively spliced.	4-8
Extragenic sequences		9
Direct Contribution to Gene Regulation		
Entire/partial promoters, enhancers, silencers	Many TEs act functionally to drive gene expression, often in a tissue-specific manner. Besides their effect on individual genes, TEs appear to have acted	10, 11-14

	<p>as mobile carriers of ready-made promoters/enhancers to widely disseminate discrete regulatory elements throughout the genome. This provides a plausible mechanism by which an entire suite of genes could become co-regulated to fashion new cellular pathways or build on existing ones.</p>	
<p>Regulatory (micro) RNAs</p>	<p>Many exonized TEs encode microRNAs (miRNAs).</p> <p>55 human miRNA genes derived from TEs have been identified with the potential to regulate thousands of genes.</p>	<p>4</p>
<p>Indirect: Retrotransposition/Transduction of Gene Sequences</p>		
<p>Gene duplication, exon shuffling, regulatory element seeding</p>	<p>Certain classes of retro-TEs (e.g. LINE and LTR elements) have a propensity to transduce host DNA due to their weak transcriptional termination sites. Duplication of genes can also occur via the appropriation of TE retrotranspositional machinery by host mRNA transcripts.</p>	<p>15, 16-18</p>

	<p>There are reportedly over 1000 transcribed “retrogenes” in the human genome, some have evolved highly beneficial functions e.g. <i>GLUD2</i>.</p>	
--	--	--

1 (Feschotte and Pritham 2007); **2** (Schatz 2004); **3** (Lander et al. 2001); **4** (Piriyapongsa et al. 2007b); **5** (Nekrutenko and Li 2001); **6** (Britten 2006); **7** (Bowen and Jordan 2007); **8** (Sorak et al. 2002); **9** (Sternberg and Shapiro 2005); **10** (Picard 1976); **11** (Jordan et al. 2003); **12** (Feschotte 2008); **13** (Bourque et al. 2008); **14** (Laperriere et al. 2007); **15** (Marques et al. 2005); **16** (Moran et al. 1999); **17** (Goodier et al. 2000); **18** (Burki and Kaessmann 2004).

Table 2-3 Passive mechanisms by which TEs generate the genetic novelty required for dramatic evolutionary change.

Role	Comments	References
Promotion of DNA Duplication (or Loss) by Unequal Recombination		
Gene duplication, exon duplication, segmental duplication	The mere presence of inactive, but similar, TEs in a genome, in large numbers, creates multiple highly homologous sites which tends to cause homology-driven ectopic (non-allelic) recombination of DNA. This is likely to account for most of the continuing effects of TEs in organisms with low TE activity,	1, 2

	<p>yet which have high TE content coupled to low TE diversity.</p> <p>DNA duplication events are particularly important in evolution since they create functional redundancy and the potential for gain of function and/or gene expression.</p>	
Promotion of Karyotypic Changes by Ectopic Recombination		
Intra- and inter-chromosomal DNA rearrangements	<p>TEs can passively underpin major chromosomal reorganizations by creating highly homologous sites dispersed throughout the genome that are prone to ectopic recombination. The Alu-mediated translocation t(11:22)(q23:q11) is the most frequent constitutional translocation in humans.</p>	3, 4, 5

1 (Bailey et al. 2003); **2** (Jurka 2004); **3** (Evgen'ev et al. 2000); **4** (Schwartz et al. 1998); **5** (Hill et al. 2000).

Although it is likely that our current knowledge about the impact of TEs still underestimates their true evolutionary value, there is now much specific evidence indicating that TEs can generate genetic novelty in one of two major ways: (i) actively, which can be by via

de novo insertion to directly contribute to gene sequences or alter gene regulation, or via retrotransposition of RNA transcripts to generate duplicate genes or exons (Table 2-2); and (ii) passively, by acting as homologous sequences to facilitate chromosomal rearrangements and gene duplications, or deletions (Table 2-3). In their active mode, even low levels of TEs will have a great impact on the genome, and high activity is likely to recur with every new wave of TE infiltration into a host lineage. In their passive mode, high TE content coupled with low TE diversity has a major impact on a genome through the promotion of ectopic recombination. This is the situation in primates, for example, where most TE sequence comprises either *L1* LINEs or *Alu* SINEs (Lander et al. 2001; Gibbs et al. 2007). In contrast, the sole sequenced genome of one of the 25 species of the basal chordate, amphioxus (*Branchiostoma floridae*), appears to be ill-suited for passive TE facilitated evolution since, despite having a TE content of 28%, its TEs are highly heterogeneous, belonging to over 500 families (Putnam et al. 2008).

2.4 TEs: Harmful to Some Individuals, but Beneficial to Lineages

The molecular mechanisms by which TEs act as powerful facilitators of genetic change mean that TEs can be deleterious to some individuals by very occasionally causing harmful mutations. In humans, rare germ line mutations caused by TEs underlie a number of genetic disorders (Belancio et al. 2008), but TEs pale into insignificance next to point changes and other small-scale DNA changes as an overall cause of mutation resulting in disease. Less than 0.2% of known disease-causing mutations

appear to be due to inactivation of genes by TE insertion events (Deininger and Batzer 1999; Kazazian 1999; Hedges and Batzer 2005), reflecting the present low level of TE activity in humans. A further 0.3% of pathogenic mutations in humans appear to result from gene deletions or rearrangements due to the passive effects of TEs as inducers of ectopic recombination, although this may be an underestimate (Deininger and Batzer 1999). TEs then are a minor cause of known deleterious mutations in humans, causing just over 0.5% of hereditary disease. Such costs to a small number of individuals are far outweighed by the longer-term benefits to the lineage. TE insertion is more important as a cause of DNA mutation in lineages that exhibit greater TE activity, for example in mice and *Drosophila*, where 10% and 50% of pathogenic mutations respectively are attributable to this mechanism (Eickbush and Furano 2002; Maksakova et al. 2006). Thus, there is a differential mutational burden of TEs across different taxa, which reflects the ability of their lineages to undergo adaptive radiation as we outline below.

2.5 TEs and the Evolution of Epigenetic Regulatory Mechanisms

Most TE insertions are tolerated without causing deleterious mutations, otherwise TEs could not have accumulated to such high levels in eukaryotic genomes. Excessive disruption of a genome can lead to a decline in individual and/or lineage fitness, so in practice TE activity is restricted, especially in somatic cells, but also to a lesser extent in the germ line, by multiple control mechanisms imposed by host lineages. Natural selection results in the establishment of a finely tuned balance in which the

mutagenic activity of TEs is kept at an acceptable level (Figure 1). The primary countermeasure against TEs in vertebrates involves epigenetic modifications to chromatin, most notably DNA (cytosine) methylation at CpG dinucleotides. Recent evidence also implicates interfering RNAs, which can originate from TEs themselves (Piriyapongsa et al. 2007a), as a means to counter TE activity through targeted destruction of their RNA transcripts (Smalheiser and Torvik 2006; Yang and Kazazian 2006). Attesting to the fact that cellular defences against TEs are multilayered, yet another mechanism for TE inhibition involves antiretroviral resistance factors of the APOBEC3 family (Bogerd et al. 2006).

Since TEs are primary targets for DNA methylation (Schmid 1991; Hata and Sakaki 1997; Rodriguez et al. 2008), they can bring methylation control to normal host genes that lie nearby (Yates et al 1999), and they also facilitate X chromosome inactivation (Lyon 2000) and genetic imprinting (Suzuki et al. 2007). Indeed, both DNA methylation and RNA interference (RNAi) are thought to have evolved primarily as cellular control devices for TEs, whereupon they were subsequently exapted as genome-wide regulatory mechanisms for the large-scale control of host gene expression (Yoder et al. 1997; Buchon and Vaury 2006). Therefore, TEs have seemingly not only generated a tremendous amount of genetic variation from which beneficial adaptations have been selected, but, as “helpful parasites”, TEs themselves have been a focus for regulatory innovation.

2.6 TEs, Punctuated Equilibrium, and Evolutionary Stasis

DNA sequence change, together with natural selection, underpins evolutionary change, but the widespread assumption that lineages evolve by the slow accumulation of adaptive mutations does not concur with most of the fossil record. Instead, evolutionary novelty has been observed to periodically arise fairly quickly and be interspersed with intermittent periods of very slow evolution or stasis. Periodic infestations of genomes by novel TEs predict “punctuated equilibrium” (Gould and Eldredge 1997) as a common occurrence (Figure 2-2), and therefore has the potential to reconcile evolutionary theory with the findings of paleontology, perhaps in a similar fashion as Darwinism and Mendelism were reconciled in neoDarwinism. This inference is based on evidence from multiple lineages indicating that TE activity does not generally occur at a low and uniform rate, but rather tends to occur in sudden episodic bursts (Gerasimova et al. 1985; Kim et al. 2004; Marques et al. 2005; Ray et al. 2008). Infestation of a genome by a modified or novel TE results in heightened TE activity and evolution for a time, but as cellular control mechanisms are refined and the new TEs succumb to degradation by mutation, TE activity is greatly reduced until eventually a new period of stasis can occur. Many TE families are conspicuously lineage-specific, for example *Alu* SINEs in primates, which strongly suggest that TE infiltrations have occurred contemporaneously with the divergence of lineages. TEs are thus likely to have been involved in promoting the evolutionary change that led to the origin of the lineage in the first place. The hominoid lineage provides an instructive example, where recent findings indicate that periodic explosive expansions of LINEs and SINEs, together with an exceptional burst of LTR elements 70

Mya, correspond temporally with major divergence points in primate evolution (Kim et al. 2004). Since some TEs predate the eukaryotes, it is also tempting to suggest that they help to account for other, more ancient evolutionary events, most notably the seemingly rapid speed at which the Cambrian explosion occurred, about 543 Mya.

In contrast to the rapid evolution seen in many lineages, genomic stability and evolutionary stasis is predicted in lineages that are not subject to intermittent infiltrations by TEs, either through *de novo* generation or horizontal transfer from other taxa. Absolute genome stability would appear to make a lineage unable to evolve significantly and to be non-fecund. Such a lineage could not adapt to changing requirements and ultimately would face prolonged stasis and possible eventual extinction (Figures 2-1 and 2-2). As a thought experiment, we can imagine a genome consisting almost wholly of coding sequences without any TE-derived DNA from which it can fortuitously, occasionally, engineer new functional sequences. Such a genome would seemingly have little significant evolutionary potential as it would have a greatly reduced capacity to create new genes or regulatory sequences. Insufficient active and passive TE effects may account for so-called “living fossil” species such as the coelacanth and tuatara. The coelacanth species (*Latimeria menadoensis* and *L. chalumnae*), two lobe-finned fish closely related to the tetrapods (the amphibians, reptiles, mammals and birds), were once thought to have been extinct for 63 Myr.

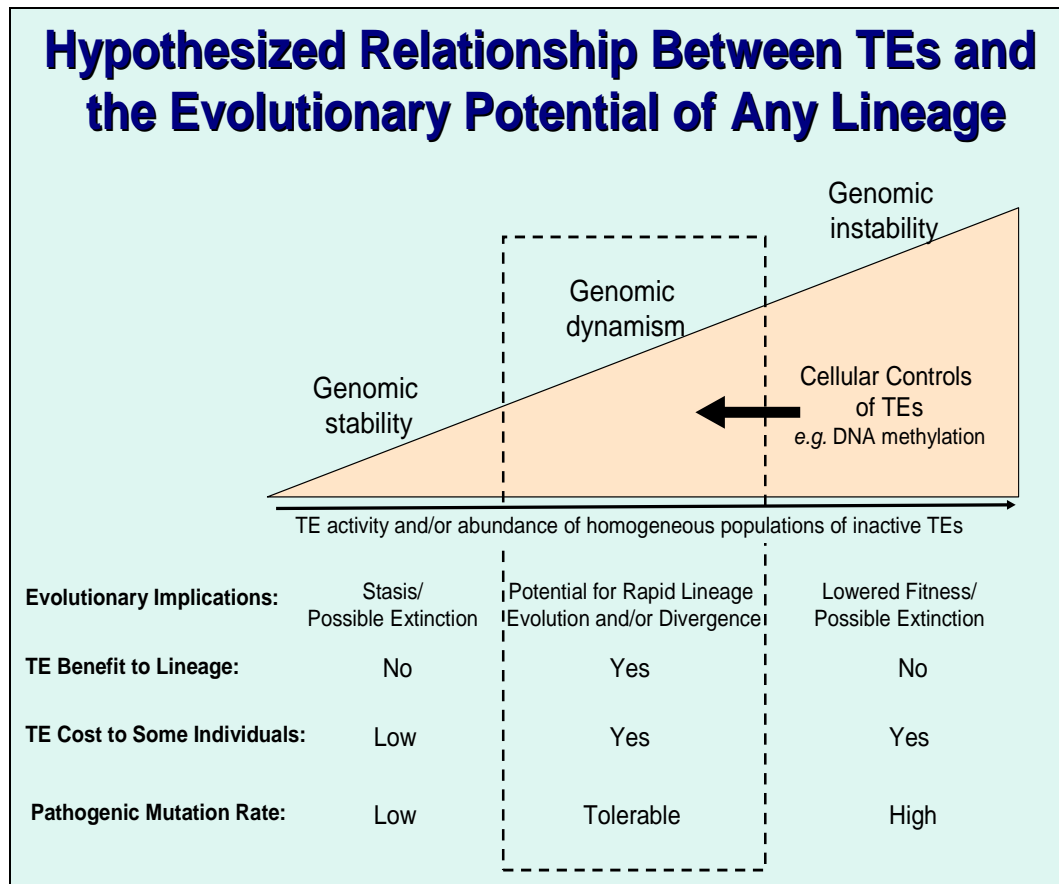


Figure 2-1. Schematic representation of the hypothesized relationship between TE activity and/or the abundance of homogeneous populations of inactive TEs on genomic variability in the germline and the evolutionary potential of any lineage. Increasing TE activity and/or abundance (with a limited diversity of inactive TEs) within a lineage promotes increased genomic variability, which, at the extreme, could result in genomic instability. In practice, most organisms have evolved strategies (such as DNA methylation) to control unfettered TE activity. Restricted, or optimum TE effectiveness (dashed box) promotes a dynamic genomic architecture that benefits the lineage at a tolerable cost to some individuals. Little or no viable TE content/activity, or the predominance of heterogeneous populations of inactive TEs, is predicted to result in stasis and the possibility of extinction. Importantly, TEs are usually much less controlled in the germ line than in the soma and TE activity is not constant, but usually occurs in intermittent waves.

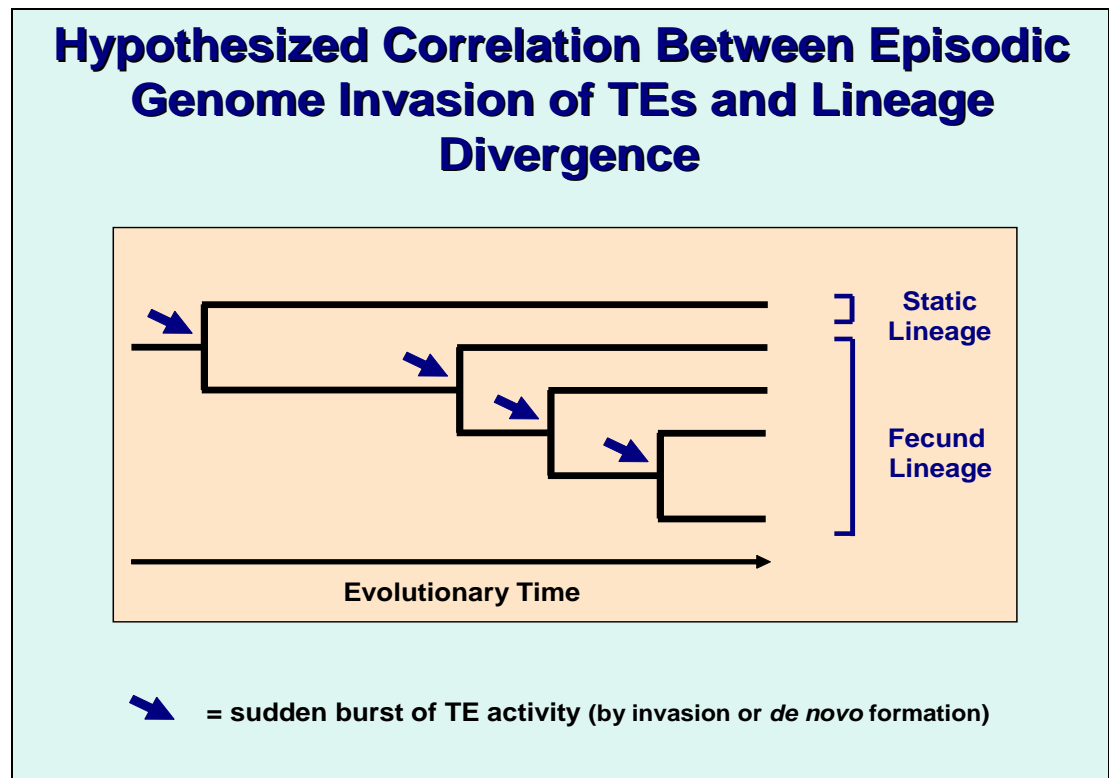


Figure 2-2. Simplified scheme illustrating the hypothesized correlation between episodic genome invasion of TEs (or their *de novo* formation) and lineage divergence. Sudden bursts of TE activity (arrowheads) following genome invasion (horizontal transfer) or *de novo* formation transiently increase genomic variability and thus the potential for lineage divergence. This may result in a fecund lineage that can undergo repeated taxonation with probable or possible punctuated equilibrium. An absence of such events within a lineage, other things being equal, is predicted to result in long-term stasis.

With a fossil record dating back to about 410 Mya, this lineage (the Sarcopterygii: coelacanths and lungfishes) gave rise to the tetrapod lineages, yet the coelacanth itself has remained little changed throughout this vast period of time. The coelacanth has SINEs that have apparently been preserved for more than 400 Myr with very little change (Bejerano et al. 2006; Nishihara et al. 2006). By contrast, the SINE families known to be active in the

tetrapods have been found to be restricted to specific clades, indicating rather recent origins and thus a rather rapid turnover of active SINE families on an evolutionary timescale (Lander et al. 2001; Waterston et al. 2002; International Chicken Genome Sequencing Consortium 2004; Lindblad-Toh 2005; Gibbs et al. 2007; Pontius et al. 2007). Although only a small percentage of its genome has so far been sequenced, the evidence suggests that evolution has stalled in the coelacanth due to a lack of intense intermittent activity by transient TE families. Thus, the coelacanth has become, in a sense, a molecular fossil (Bejerano et al. 2006), or a “living fossil,” possibly well adapted to a seemingly stable habitat.

Another living fossil is the tuatara of New Zealand, (Figure 2.3.) with two closely related species (*Sphenodon punctatus* and *S. guntheri*) These are the last remaining representatives of a previously abundant and highly diverse reptilian lineage known as sphenodontids that co-existed with the early dinosaurs, around 220 Mya. The retro-TE content of 121 kb of tuatara genomic DNA sequence was found to be only 2.7%, comprised of 0.11% SINEs and 2.59% LINEs, the latter being heterogeneous (Wang et al. 2006). In a separate study, one DNA-TE was identified in the tuatara (*S. punctatus*), but the coding regions contained several stop codons indicating that it has been immobile for a very long time (Kapitonov and Jurka 2006). It is difficult to make much of such limited data except to note that Wang et al. (2006) reported surprise to us as the low TE content, and the relatively high diversity of retro-TE families in a living fossil are very compatible

Figure 2-3 Tuatara, living fossil reptiles (Photos J McComb)

A: *Sphenodon punctatus*



B: *Sphenodon guntheri*



with the proposal of TEs as essential facilitators of evolution. However, more data are needed to reach any firm conclusions.

2.7 TEs as a Prerequisite for Evolutionary Radiation

Why certain lineages are incredibly diverse and others are species-poor remains an enigmatic question in evolutionary biology. Many factors have been proposed to contribute to the variability in species-richness observed among different taxa, although none seem applicable to all taxa. A key factor, we contend, are TEs, which serve to dramatically increase the rate of molecular evolution and thus the probability of speciation, depending on ecological and other factors. In the absence of TEs, it is difficult to envisage how significant taxon events could occur, given that members of species tend to become genotypically trapped at local optima metaphorically termed “adaptive peaks” (Kauffman and Johnsen 1991). However, the extent of the genetic change wrought by TEs permits the emergence of new genes, the alteration of gene expression patterns, and the structural rearrangements of chromosomes, all of which are thought to be fundamental to the evolution of lineage or species-specific traits (Barrier et al. 2001; Riesberg 2001; Orr et al. 2004). Changes of this magnitude are important for taxon events because they permit rapid crossings from one adaptive peak to another, a phenomenon difficult to explain by gradualism.

The idea that TEs could promote the appearance of new species has been proposed previously (Ginzburg et al. 1984; McClintock 1984; McFadden and Knowles 1997), but has received little attention, largely due to a lack of strong evidence. While there are

few complete genome sequences and insufficient data on TE activity in a comprehensive range of taxa, some data is available from diverse examples: In *E.coli* B there has been a high level of Insertion Sequence (IS type TE) activity (including transpositions, deletions, and horizontal gene transfer) since K12 and O157:H7 diverged from a common ancestor (Schneider et al. 2002). The *virilis* group within the speciose genus *Drosophila* possesses rich karyotypic variety that is significantly correlated with the position of DNA-TE insertion sites (Evgen'ev et al. 2000). In plants, TEs are the most significant factor in determining the structure of complex genomes, often comprising the majority of the genome (Bennetzen 2000).

In recent years a number of vertebrate, mainly mammalian, genomes have been fully sequenced, which have yielded comparative data on TE types and content (Lindblad-Toh et al 2005; Waterson et al. 2002; Gibbs et al. 2004; Lander et al. 2001; Mikkeisen et al. 2006; Gibbs et al. 2007; Mikkeisen et al. 2007; Pontius et al. 2007). In our view, genomes from species-rich lineages would necessarily exhibit high TE activity and/or high infiltration by homogeneous populations of TEs. We see such a plastic genome architecture, if other factors are equal, as a prerequisite for adaptive radiation as it provides the required genetic variation for a lineage to exploit ecological opportunities when old niches are emptied by extinctions or when new niches are, or can be, created. The mammalian order, Chiroptera, provides a good example of the creation of a new niche as exemplified by the rapid and fecund radiation of the microbats (suborder Microchiroptera) that began with an Eocene (55-34

mya) “big bang” (Simmons 2005). A flying mammal that could feed on flying insects was probably a hard niche to occupy, but once occupied it allowed a massive radiation of such magnitude that bats now account for over 22% of all mammalian species (Wilson and Reeder 2005). The most fecund genus in the microbats is *Myotis* (the mouse-eared bats) with 103 species. Recent data for representatives of *Myotis*, most particularly *Myotis lucifugus*, have revealed that many of the microbats are richly endowed with TEs, both retro-TEs and DNA-TEs (Pritham and Feschotte 2007; Ray et al. 2007; Ray et al. 2008). Previously, DNA-TEs were thought to have been inactive in mammals for at least 37 Myr (Lander et al. 2001; Gibbs et al. 2004; Pace and Feschotte 2007). Remarkably, the *Myotis* DNA-TEs appear to be still active and there have evidently been sequential waves of DNA-TE activity in this lineage, resulting in massive amplifications of individual elements (Pritham and Feschotte 2007; Ray et al. 2007; Ray et al. 2008). DNA-TEs have been implicated in the duplication and shuffling of host genetic material (Pritham and Feschotte 2007). Thus, it seems probable that the ability of the genus *Myotis* to adapt to a range of niches, and thereby spectacularly diversify is being underpinned by natural selection acting on the dynamic genomes created, at least in part, by the activity of DNA-TEs.

Among other mammals, genomic plasticity engendered by TEs should also be found in the large orders Rodentia (~2,000 species) and Primates (~235 species). TEs in representative species of two large rodent genera, *Mus* and *Rattus* are not only quite homogeneous (Waterston et al 2002; Gibbs et al, 2004),

being dominated by just a handful of elements, namely LINE-1, ERV/sLTR, and B1, B2, and B4-SINEs, but significantly, have remained highly active (DeBerardinis et al. 1998; Gibbs et al. 2004; Maksakova et al. 2006). Viable ERVs are nearly extinct in humans, but are particularly active in the rodents (Smit 1993; Costas 2003; Gibbs et al. 2004) and LINE-1 elements have also remained highly active (DeBerardinis et al. 1998; Brouha et al. 2003; Goodier et al. 2001). Although ecological and other factors have likely contributed to the success of the Rodentia, the high content, homogeneity and activity of TEs evident in rodent genomes so far sequenced is strongly consistent with TEs being an essential force in the Rodentia radiation.

TEs have also been highly active in the primate lineage (Lander et al. 2001; Mikkeisen et al. 2005; Gibbs et al. 2007). This activity has not been consistent over time; rather, primate evolution has been marked by periodic explosions of TE activity with mobility now having been largely curtailed from its peak about 40 Mya (Kim et al. 2004; Khan et al. 2006; Mills et al. 2007; Pace and Feschotte 2007). Even so, significant residual TE activity persists (Mills et al. 2006). A critical feature of primate TEs is not only their abundance but their remarkable homogeneity, with just two elements, *L1* LINES and *Alu* SINEs, accounting for over 60% of all interspersed repeat DNA in this lineage (Lander et al. 2001; Mikkelsen et al. 2005; Gibbs et al. 2007). This makes primate genomes virtually ideal for the passive effects wrought by TEs. By promoting homology-driven ectopic recombination of DNA, *L1* and *Alu* repeats can bring about both inter- and intra-chromosomal rearrangements that underlie lineage- and species-specific

genetic changes. Comparisons between the human and chimpanzee genomes have revealed the significant extent to which TEs have passively exerted their effects in the recent evolutionary history of primates (Sen et al. 2006; Han et al. 2007). The observed high enrichment of TE elements at low copy repeat junctions indicates that TEs have also been an important factor in the generation of segmental duplications that are uniquely abundant in primate genomes (Bailey et al. 2003; Bailey and Eichler 2006; Johnson et al. 2006a; Wooding and Jorde 2006).

Confirmation that TE-facilitated evolution is responsible for much taxonation will require much more data from different taxa, including a large increase in DNA sequence information. Sequenced representatives of the fecund orders of rodents, bats and primates certainly support the concept that TEs are powerful facilitators of evolution. Importantly, we could find no counterevidence in the form of any mammalian species-rich lineages lacking significant TE content and/or active or passive effects. We therefore see the effects of TEs as having the potential to explain why certain other animal clades are anomalously large. These are prime targets for future research.

2.8 TE Activity Increases under Conditions of Stress

The deleterious effects of TEs are minimised by mechanisms that involve TE repression or excision, depending on the host taxon. However, as first proposed by McClintock (1984) the cost/benefit ratio of TE-facilitated variation in a host may shift during periods of greater evolutionary stress. Under stress, increased levels of TE transposition might be advantageous, accelerating the rate of

genome restructuring and promoting potentially useful genetic variability. Lineages able to diversify in this manner are more likely to remain extant, by virtue of at least some progeny inheriting a favourable adaptation to enable survival in the face of biotic or environmental challenges. Much evidence indicates that a TE-host relationship has indeed evolved whereby normally innocuous TEs become much more active through transcriptional de-repression at times of stress. This has been documented in a wide range of taxa, from yeast to mammals. Cellular stressors known to activate TEs, particularly SINEs, include heat shock (Liu et al. 1995; Kimura et al. 2001; Li and Schmid 2001), genetic damage (Rudin and Thompson 2001; Hagan et al. 2003), oxidative stress (Cam et al. 2008), translational inhibition (Liu et al. 1995; Kimura et al. 2001; Li and Schmid 2001) and viral infection (Kimura et al. 2001; Li and Schmid 2001). An appealing possibility is that TEs may actually be part of the normal physiological response to cellular stress, with putative functional roles in DNA double-strand break repair (Eickbush 2002; Olivares et al. 2003) and the regulation of protein translation (Chu et al. 1998; Häslér and Strub 2006). Such roles would accord with SINEs being disproportionately located within gene-rich areas of the genome, a distribution that is possibly explicable if these elements provide some benefit and are therefore subject to positive selection (Lander et al. 2001). In any case, it would appear that successful taxa specifically permit more TE activity under conditions of stress, which would enhance genetic change at times when it would provide most benefit to the lineage. As a consequence, stress may be a contributing factor to punctuated equilibrium by promoting sudden evolutionary bursts in lineages

following periods of stasis.

2.9 TEs are Differentially Active in the Germline and During Early Embryogenesis

Uncontrolled transposition of TEs in the soma cannot benefit the lineage, but can be deleterious to individuals, for example, by leading to cancers if oncogenically-relevant genes are affected (Bannert and Kirth 2004). However, TEs can only be useful facilitators of evolution if they are allowed some leeway to propagate within the germ line, or at least within the very early embryo, since only such genetic change can be inherited and be of potential benefit to the lineage. Dramatic hypomethylation of DNA within primordial germ cells and their descendants in the germ line is a well-known phenomenon in mammals (Allegrucci et al. 2005; Morgan et al. 2005; Reik 2007). This opens a “window of opportunity” to allow some TE activity as the gametes are formed, and thus inheritance of altered genomes by the zygote. A second window of opportunity occurs in the preimplantation embryo, where massive genomic demethylation occurs following fertilization (Allegrucci et al. 2005; Morgan et al. 2005; Reik 2007). During these hypomethylation windows, TEs temporarily become transcriptionally active (Dupressoir and Heidmann 1996; Evsikov et al. 2004; Peaston et al. 2004). This is reflected in enormously high reverse transcriptase levels in mouse oocytes and preimplantation embryos compared with somatic cells (Evsikov et al. 2004), as well as in TE transposition activity, with several *de novo* insertion events having been documented in the human germ line or early embryo (Wallace et al. 1991; Brouha et al. 2002; van den Hurk et al. 2007). In support of the data that these

hypomethylation episodes provide fertile ground for TE-mediated evolutionary change, genes expressed during gametogenesis and early development have a much greater chance of being retrotransposed than genes expressed exclusively in somatic tissues (Kleene et al. 1998; Evsikov et al. 2004). TEs thus powerfully facilitate evolution and the reach of natural selection extends to the gamete, the zygote, and the embryo. Intriguingly, TE-mediated transcription during mouse embryogenesis actually appears to be essential for normal preimplantation development (Peaston et al. 2004; Beraidi et al. 2006), illustrating how TE biology and normal host physiology are often heavily entwined.

Since some TE activity in the germ line, but not the soma, is beneficial to lineages, the restriction of TE activity to the germ line should be a widespread adaptation. Indeed, eukaryotes have evolved many mechanisms for reducing harm to the soma while allowing TEs to generate diversity in the germ line genome. In *Drosophila melanogaster* and related species, the mobility of the *P* element DNA-TE is limited to the germ line by an alternate splicing mechanism, which generates two different TE-encoded proteins: a repressor of transposition in somatic cells and a germ line specific transposase responsible for genomic mobility (Rio 1990). Similarly, *I* element LINE retrotransposition is restricted to the female germline (Picard 1976), whereas *gypsy* and *ZAM* LTR-TEs are specifically expressed in follicle cells of the developing oocyte; both then invade the oocyte before the vitelline membrane forms (Song et al. 1997). Nuclear dimorphism in the ciliated protozoan *Tetrahymena thermophila* provides an alternative mechanism for differential TE activity. TEs are restricted to the

germ line micronucleus since the somatic macronucleus specifically undergoes programmed DNA rearrangements and deletions to remove potentially deleterious TEs (Fillingham et al. 2004). These examples indicate that successful lineages, from protozoa to mammals, permit TE activity in the germ line and strictly minimise it in the soma, where it is potentially damaging.

2.10 The Differential Impact of TEs on Distinct Regions of the Genome

TEs promote genome plasticity, but to be of most benefit to lineages should selectively operate on genes whose evolvability needs to be high for host fitness, and be excluded from highly conserved genes with critical functions. Consistent with this, TEs have been found to be enriched within and near rapidly evolving genes with roles that demand flexibility, such as responses to external stimuli, immunity, cellular signaling, transport and metabolism (Grover et al. 2003; van de Lagemaat et al. 2003; Chen and Li 2007). The accumulation of TEs to high densities near such genes can subsequently promote ectopic recombination. Accordingly, it has been proposed that TE-rich regions, which undergo relatively frequent unequal crossover, can be considered to be “gene factories”, that is, genomic sites where gene clusters are preferentially formed (Mallon et al. 2004). Gene families generated in this manner can benefit the host lineage by undergoing subfunctionalization following on from sequence divergence. However, highly conserved genes with crucial roles in cell structure or the regulation of development have been found to be TE-poor (Simons et al. 2006; Chen and Li 2007). Most notable are the tightly linked vertebrate HOX gene clusters (four in

mammals, seven in most fish) which are virtually devoid of TEs. A paucity of TEs among the clusters of vertebrate HOX genes may reflect extensive functional and organizational constraints that strongly select against insertion events (Wagner et al. 2003; Fried et al. 2004; Simons et al. 2006). It is also consistent with their strict requirement for stability, since once HOX genes had evolved distinct roles as master regulators of the complex vertebrate body plan, it became unlikely that any significant flexibility engendered by TEs would be tolerated by natural selection. Interestingly, invertebrates, with over 30 phyla are radically more diverse in body plan than vertebrates, and possess a more evolvable single HOX gene cluster that is permissive to TE insertions (Wagner et al. 2003; Fried et al. 2004). That TEs associate with dynamic gene regions adds weight to the view that they have been a major force in gene evolution in a wide range of taxa.

2.11 Evolutionary Potential is Compromised in Organisms That Fully Control TEs.

Eukaryotes mostly suppress TEs rather than eliminate them, but at least one organism, the ascomycete fungus *Neurospora crassa*, has been able to totally rid itself of intact TEs by means of a novel repeat-induced point mutation (RIP) mechanism (Galagan and Salker 2004). A critical characteristic of RIP is that it not only eliminates TEs, but also any newly formed gene duplicates. As gene duplication is thought to be almost essential for effective evolution (Ohno 1970), fungi such as *N. crassa* seem to have destroyed their evolutionary potential in the interests of genome defence. Just 0.1% of its 10,082 genes share greater than 80% similarity (Galagan et al. 2003). Thus, RIP has seemingly come at

the cost of deactivating TE facilitation of evolution almost completely. If *N. crassa* has been able to evolve RIP then why haven't other taxa taken such extreme measures to combat TEs? We would expect that selection for the destruction of TEs would result in stasis and possible eventual extinction, and that selection for very strong suppression in the soma, with some activity of TEs maintained in the germ line, would promote evolvability, with minimum fitness costs to individuals. This suggests a ready explanation as to why nearly all extant taxa do not destroy their TEs.

2.12 Conclusions

Mutational variability is vital to evolution because it provides the raw material upon which natural selection (and/or random drift) can act to lead to evolution and taxonation. Here, we bring together seemingly overwhelming evidence that supports a major evolutionary role for TEs as an irreplaceable source of genetic novelty. Far from being junk, TEs have established dynamic genome architectures and have an evolutionary legacy that is breathtaking in scope, ranging from the creation of novel genes and gene families to the establishment of complex gene regulatory networks. In humans, TE sequences have directly contributed to about a quarter of transcribed gene sequences, including a significant number of protein-coding regions, and a quarter of gene promoter regions. TEs are not curious and infrequent causes of genomic change but have repeatedly made very significant contributions to genome evolution. Moreover, as "helpful parasites" TEs appear to have prompted the emergence of genome-wide regulatory processes such as DNA methylation

and RNAi. In many cases, TE biology and normal host physiology have become interwoven, with TEs being directly implicated in critical processes such as epigenetic gene silencing, embryonic development, X-chromosome inactivation, DNA repair, stress response and antigen-specific immunity. Nevertheless, given their ancient origins, the biological and evolutionary impact of TEs is probably still underestimated, since much of their influence may now be untraced, and untraceable.

A very important feature of TEs is that they produce a bewildering variety of mutations, including highly complex ones that are very unlikely to arise by any other means. Functioning actively, or passively as homologous sites for ectopic recombination, TEs can suddenly duplicate, modify, or remove coding regions, greatly alter gene expression patterns, or rearrange chromosomes. Although such changes can be detrimental, natural selection, and the evolution of host TE-control mechanisms such as DNA methylation ensure that a balance is achieved between a tolerable level of deleterious effects on individuals and long-term beneficial effects for the lineage.

TEs possess all the qualities needed to be powerful facilitators of evolution. They are seemingly universally distributed, incredibly ancient, highly abundant and actively mobile. Since TE activity tends to occur in sudden episodic bursts, for example following horizontal transfer, or *de novo* derivation, TEs predict punctuated equilibrium as a common feature of evolution. An abundance of TE activity and/or passive homogeneity of TEs, provides a prerequisite explanation, other things being equal, as to how

certain lineages are able to diversify spectacularly. In contrast, genomes without significant TE activity or homogeneity of inactive TEs are predicted to be liable to evolutionary stasis. Much more data will be available in the near future and if this data supports the hypothesis clearly emerging from this review, it could become a new paradigm in evolutionary theory.

Supplementary note for Chapter 2

Non-coding RNAs

The overwhelming complexity of the non-protein coding transcriptome, and its function in gene regulation is rapidly becoming apparent (Werner & Swan 2010). piRNAs (piwi interacting RNAs) are germ line specific RNAs, and male mice that lack key enzymes in piRNA become sterile (Lau et al. 2006). piRNAs derive from distinct non-coding regions of the genome and suppress TE activity by transcriptional silencing. Thus in this way they impact on the TE-Thrust hypothesis. In plants siRNAs (short interfering RNAs) represent a powerful defence strategy against viruses, and plant cells produce virus-derived siRNAs upon infection whereas animal cells do not (Ding & Voinnet 2007). This is because defence against viruses is largely covered by the immune system in animals, so the biological role of siRNAs in animals is speculative (Okamura & Lai 2008). The siRNAs are noted here as in future some of their actions in plants may be important in relation to the TE Thrust hypothesis.

Reference: Werner & Swan 2010 and the references therein.

Appendix to Chapter Two

Biological Evidence Supporting Punctuated Equilibrium Type Evolution

A2.1 Summary

In contrast to the concept of gradualism in evolution current among many biologists (in line with Darwin's proposals), paleontologists found a pattern of punctuation events interspersed with relative stasis in the fossil record, which they named punctuated equilibrium. Of late, some biologists have sought to determine which model was the correct one, or whether gradualism and punctuated equilibrium both occurred on occasions, and some have also proposed hypotheses to possibly explain this type of evolution. An explanation for both gradualism and punctuated equilibrium is an important component of the TE-Thrust hypothesis. Some recent studies by biologists support the occurrence of both punctuated equilibrium and gradualism. In addition, karyotypic changes associated with bursts of TE activity appear to be a potent source of reproductive isolation and speciation, although it is noted that there are multiple other causes of reproductive isolation which may result in speciation.

A2.2. Introduction

Gradualism and Punctuated Equilibrium are two possible modes of evolution, or perhaps two extremes of a continuum. Recent

evolutionary thought has been dominated by an assumption that biological lineages evolve by the slow and gradual accumulation of adaptive mutations, that is, by gradualism, and that macroevolution (the origin of higher taxa) can be explained by an extrapolation of microevolution (the origin of races, varieties and species) into the distant past (Kutschera and Niklas 2004; and many others). This line of thought has been mostly dominant since Charles Darwin who, influenced by Lyell's concept of very slow changes in geology, regarded gradualism as fundamental to his theory. Darwin unreservedly said *Natura non facit saltum* (nature does not make a leap). Despite a number of early dissenters such as Bateson in the 1890s who strongly advocated evolution by discontinuous variation or sudden leaps, gradualism was eventually incorporated into neoDarwinism and the Modern Synthesis (Bowler 2003). However, many palaeontologists have found that gradualism does not concur with the majority of the fossil record. Instead, new species are found to arise abruptly and periodically and there are intermittent and often long periods of stasis, punctuated by periods of rapid change and branching speciation. These punctuations often occur during different periods in diverse lineages, so are apparently not always related to environmental changes. The observed persistence of ancestors in stasis, following the abrupt appearance of a descendant, is an indicator of punctuated equilibrium (Eldredge and Gould 1972; Gould and Eldredge 1977; Stanley 1981; Eldredge 1986, 1995; Gould 2002). Punctuated equilibrium, as detailed by the palaeontologists cited above, has been observed in certain very fine grained strata, and entails intermittent periods of rapid evolutionary change, over an estimated 15,000 to 40,000 years

(Gould 2002), which gives birth to a new taxon that remains little changed (i.e. in a period of stasis, or gradualism) until it becomes extinct, usually four to ten million years later (Appendix to Chapter 4: A4.1 to A4.8). This taxon, while it is extant, is often the progenitor of other taxa in the same lineage. Contemporaneous, or successor taxa, in the same lineage eventually suffer the same fate. Mass extinction events can interrupt this pattern, but they account for less than 5% of all extinct species and recovery from them tends to be slow, about 5 million years in the Early Triassic, after the end of Permian great mass extinction (Erwin 2001). This seems to make the “Cambrian explosion” seem all the more remarkable.

That gradualism occurs sometimes is not denied. Fortey (1985), from a study of Ordovician trilobites, estimated that the ratio of punctuated equilibrium type speciation to gradualist speciation is 10:1, while Ridley (2004) posits that although both occur, and punctuated equilibrium appears to be the more common, they may be extremes of a continuum. It seems, therefore, that the ratio of these types of speciation events, one to the other, is somewhat uncertain. Gould (2002) states that punctuated equilibrium should not be confused with the hypothesised evolution of “hopeful monsters” by saltations (Goldschmidt 1940). Many palaeontologists have observed punctuated equilibrium, but they could not explain it in terms of the Modern Synthesis. Now, however, intermittent waves of transposable element activity have very recently been hypothesised to be a major causal factor of punctuated equilibrium (Oliver and Greene 2009a;b Zeh et al. 2009; Parris 2009). This finally reconciles evolutionary theory to

punctuated equilibrium and the fossil record. However, whereas Zeh et al. (2009) place heavy emphasis on environmental stress as a trigger for TE activity, we additionally consider recent acquisitions of TEs as intermittent events that can trigger new waves of TE activity (Oliver and Greene, 2009a). Parris (2009) also proposes that intermittent germ line endogenisations by retroviruses, possibly in concert with environmental change, are an example of a trigger for intermittent rapid taxonation, and I agree that this also occurs

A2.3 Biological Evidence for Punctuated Equilibrium

There are data, independent of that from paleontological data, which suggest support for the occurrence of punctuated equilibrium type evolution. These data further support the TE-Thrust hypothesis modes, which predict punctuated equilibrium type evolution due to intermittent bursts of TE activity. These can occur as either punctuation events interrupting stasis, or as punctuation events interrupting gradualism (Table 3-1).

A2.4 to A2.7 below, are also quite independent of any reliance on the consideration of studies of TE activity.

A2.4 Cubo (2003) in a study of extant ratites (Aves: Palaeognathae) finds evidence for speciation change, rather than gradual change in extant ratites. In speciation models morphological change is assumed to occur during or just after cladogenesis in both daughter species, and the resultant morphologies remain in stasis over long periods of time until the next cladogenetic event.

A2.5 In their study of various sub-clades of extant mammals the Bayesian estimates of Matilla and Bokma (2008) suggest that gradual evolution is responsible for only a small part of body size variation between mammal species. They conclude that as gradual evolution only explains perhaps one-third of interspecific variation, gradual evolution seems to be grossly overvalued.

A2.6 In a study of extant ratites Laurin et al. (2011) confirm the conclusion of Cubo (2003) and find that ratite evolution has been mostly speciational (close to the punctuated equilibrium model) for shape related characters. However, their data suggest that it has been mostly gradual for size related characters.

The TE-Thrust hypothesis offers an explanation for both punctuated equilibrium and gradualism (Table 3-1), so these data are in accord with the predictions of this hypothesis, and suggest that it may be correct.

A2.7 Co-evolution

Independent support for evolution by punctuated equilibrium has also come from an example of co-evolution (Togu and Sota 2009).

A2.8 Archaeogenomic Angiosperm Studies

In extant and ancient plant genomes of *Gossypium* (cotton) species there is archaeogenomic evidence of punctuated genome evolution due to intermittent TE activity. It was found that an apparent TE-consortium enlargement (punctuation) event in *G. herbaceum* has occurred in far less than 2,000 years, as

comparative analyses of retro-TE profiles from archeological (1,600 years old) and modern *G. herbaceum* genomes show significantly different genomic TE composition. This suggests that the TE activity and evolutionary development of this domesticated lineage has been occurring at a high level. In contrast to this, there was minimal differentiation in the TE consortia between some recent and archaic samples of *G. barbadense* genomes. This was despite the samples being separated by over 2,000 miles in distance and 3,000 years in time, thus suggesting stability or stasis within this lineage (Palmer et al. 2012).

A2.9 Whole Genome Duplication (Polyploidy)

Whole genome duplication (WGD or polyploidy), which is abundant in angiosperms and which can result in punctuated equilibrium type evolution (4.11 & 4.12), has not entirely ceased in vertebrates, where an allotetraploid rodent species has been identified (Gallardo et al. 2004; Gallardo et al. 2006), although most occurrences of whole genome duplication in vertebrates are thought to be of very ancient origin. The duplication of the HOX clusters is thought to have occurred via WGD (Kassahn et al. 2009). Retained ohnologs are highly biased towards those for signalling proteins and transcription factors, suggesting that this large pool of new genes could have enabled the complex regulation required for the vertebrate body plan. There have been two rounds of WGD in all jawed vertebrates, and this accounts for the genesis of almost one third of human genes. Most fish have undergone a third WGD more recently (Manning and Scheeff 2010) These duplications of Hox clusters, by WGD, from one to two, and from two to four, or more, followed by eliminations of

some genes in some of the duplicated clusters clearly marks a major feature of vertebrate evolution (Gould 2002).

A2.11 Further Evidence for the TE-thrust Hypothesis as an Explanation for Punctuated Equilibrium

Several lines of investigation suggest that various types of genomic disruption can play an important role reproductive isolation and speciation, and that such speciation may owe more to ephemeral and essentially arbitrary events than to a gradual response to natural selection. Almost 80% of new species appear to originate due to rare stochastic events (Venditti et al. 2010).

There are an astounding 529 species in 122 genera in the Rodentia, Muridae subfamily Murinae (Michaux 2001). The pattern of karyotype evolution in the *Mus* subgenera of this Old World subfamily Murinae, supports punctuation event type evolution. This pattern is not consistent with evolutionary gradualism, and suggests that taxonation is due to rare abrupt events.

The four subgenera of *Mus* (Murinae) diverged nearly simultaneously within a million years during which karyotypic changes occurred at the rate of about 13 per Myr. In contrast to this the karyotypic change rate was change of only about one per Myr during the next 6-7 Myr. That is, the pattern of karyotypic change exhibited a short phase of intensive change, followed by a stage of much slower karyotypic change. In *Mus* the period of high karyotypic change coincided with cladogenetic events: the separation of the four subgenera occurred during this period.

Compared to the high rates of chromosome change in the very speciose Muridae, other mammalian lineages generally display a low rate of karyotypic change of about 0.1 to 0.2 changes per Myr (Veyrunes et al. 2006). This suggests that it may be no coincidence that the very speciose Muridae (26% of extant mammal species) are highly infested with persisting retroviruses and endogenised retroviruses (Maksakova 2006) which are causal to karyotypic changes via ectopic recombination of the large-sized, and abundant, ERV insertions (Feschotte and Gilbert 2012). It also suggests a link between punctuated karyotypic changes and punctuated speciation in the Muridae.

A similar pattern of rapid evolution and karyotypic change has been found in many of the rodents of the very speciose New World Muridae, subfamily Sigmodontinae (6.16.2) which have very numerous MysTR ERVs (Cantrell et al. 2005; Erickson et al. 2011). The Sigmodontinae contains an extraordinary 79 genera and 432 species (Michaux et al. 2001).

Bush et al. (1977) in a study of extant and extinct species in 225 vertebrate genera, found that speciation rate strongly correlated with the rate of chromosomal evolution, and that average rates of speciation in lower vertebrate genera were only one fifth of those in the mammalian genera.

A2.12 Other factors in Reproductive Isolation and Speciation

I do not suggest that karyotypic changes are the only source of reproductive isolation, which often precedes the divergence of species, or of higher taxa, as there are many other causes of reproductive isolation. These include subdivision of a population

into semi-isolated demes with a small effective population size (Wright 1931; Bush et al. 1977; Eldredge 1995; Jurka et al. 2011). This can be due to patchy distribution, organisation into clans or harems, low adult mobility and juvenile dispersal, or strong individual territoriality as in some rodents, e.g. *Mus musculus* (Bush et al. 1977), and to many other factors, e.g. environmental changes, behavioural changes, and genetic or other genomic changes (Venditti et al. 2010). Mating preference can also be a cause of reproductive isolation due to differing fly microbiota, as has been demonstrated experimentally in *Drosophila melanogaster* (Sharon et al. 2010). In plants, reproductive isolation can be caused by ploidy differences, or to endosperm balance number (EBN) differences in angiosperms (Box 4-1) as well as many other factors.

A2.13 TE-Activity, Adaptation, and Speciation

In their short review Rebollo et al. (2010) argue that:

- Some bursts of TE activity are able to induce speciation through karyotypic changes.
- Generation of multiple *L1* LINE families occurred concurrently with intense speciation in <0.3 Myr, within *Rattus* (Murinae; Rodentia).
- Bursts of transposition are not always associated with speciation, e.g. the *P* (DNA-TE) and *I* (LINE, retro-TE) TEs have recently been acquired by *Drosophila melanogaster*, without any observed speciation occurring. However, in this regard, I would note that *D. melanogaster*, has very recently (<400 years) colonised the world, from its sub-Saharan origin,

suggesting a recent increase in, or realisation of its adaptive potential (5.10, 5.11). This further suggests that this occupation of new habitats, possibly combined with other factors (A.2.12) may eventually result in speciation, as estimates of the time required for a punctuation event are 15,000 to 40,000 years (A2.2; 5.17) and other estimates of the time needed for speciation are much higher (e.g.100,000 years: Byron Lamont, personal communication).

- Bursts of transposition may be driven by selective release of viable TEs as the result of epigenetic response to the environment, as TEs are subject to epigenetic regulation. This appears to have a Lamarckian flavour (4.6.1).
- Bursts of transposition result in a renewal of genetic diversity, that is, they result in adaptive potential, and/or evolutionary potential (Box 5-1), which may be realised in the present or at some time in the future (5.10, 5.11).

A2.12 Conclusion

Punctuated Equilibrium type evolution has long been recognised by paleontologists (Eldredge and Gould 1972; Gould and Eldredge 1977; Stanley 1981; Eldredge 1986, 1995; Gould 2002). In agreement with this there are data, independent of paleontology and also in many cases of studies of TEs and TE activity, which support both the occurrence of punctuated equilibrium type evolution and gradualism. These data, combined

Appendix to Chapter 2: Biological Evidence supporting Punctuated Equilibrium Type Evolution

with the paleontological data, suggest that the rigid gradualism of Darwin, and of the Modern Synthesis, although it sometimes occurs, is very unlikely to be able to account for the whole of the evolution of life on earth. Intermittent TE activity is likely to be a major contributor to punctuated equilibrium.

Chapter 3

Mobile DNA and the TE-Thrust Hypothesis: Supporting Evidence from the Primates

3.1 Summary

Transposable elements (TEs) are increasingly being recognised as powerful facilitators of evolution. We propose the TE-Thrust hypothesis to encompass TE-facilitated processes by which genomes self-engineer coding, regulatory, karyotypic or other genetic changes. Although TEs are occasionally harmful to some individuals, genomic dynamism caused by TEs can be very beneficial to lineages. This can result in differential survival and differential fecundity of lineages. Lineages with an abundant and suitable repertoire of TEs have enhanced evolutionary potential and, if all else is equal, tend to be fecund, resulting in species-rich adaptive radiations, and/or they tend to undergo major evolutionary transitions. Many other mechanisms of genomic change are also important in evolution, and whether the evolutionary potential of TE-Thrust is realised is heavily dependent on environmental and ecological factors. The large contribution of TEs to evolutionary innovation is particularly well documented in the primate lineage. In this paper, we review numerous cases of beneficial TE-caused modifications to the genomes of higher primates, which strongly support our TE-Thrust hypothesis.

3.2 Introduction

Building on the groundbreaking work of McClintock (1956) and numerous others (Georgiev 1984; Brosius 1991; Fedoroff 1999; Kidwell and Lisch 2001; Bowen and Jordan 2002; Deininger et al. 2003; Kazazian Jr 2004; Wessler 2006; Biémont and Vieira 2006; Volff 2006; Feschotte and Pritham 2007; Muotri et al. 2007; Böhne et al. 2008), we further advanced the proposition of transposable elements (TEs) as powerful facilitators of evolution (Oliver and Greene 2009a) and now formalise this into 'The TE-Thrust hypothesis'. In this paper, we present much specific evidence in support of this hypothesis, which we suggest may have great explanatory power. We focus mainly on the well-studied higher primate (monkey, ape and human) lineages. We emphasise the part played by the retro-TEs, especially the primate-specific non-autonomous Alu short interspersed element (SINE), together with its requisite autonomous partner long interspersed element (LINE)-1 or L1 (Figure 3-1A). In addition, both ancient and recent endogenisations of exogenous retroviruses (endogenous retroviruses (ERVs)/solo long terminal repeats (sLTRs) have been very important in primate evolution (Figure 3-1A). The Alu element has been particularly instrumental in the evolution of primates by TE-Thrust. This suggests that, at least in some mammalian lineages, specific SINE-LINE pairs have a large influence on the trajectory and extent of evolution on the different clades within that lineage.

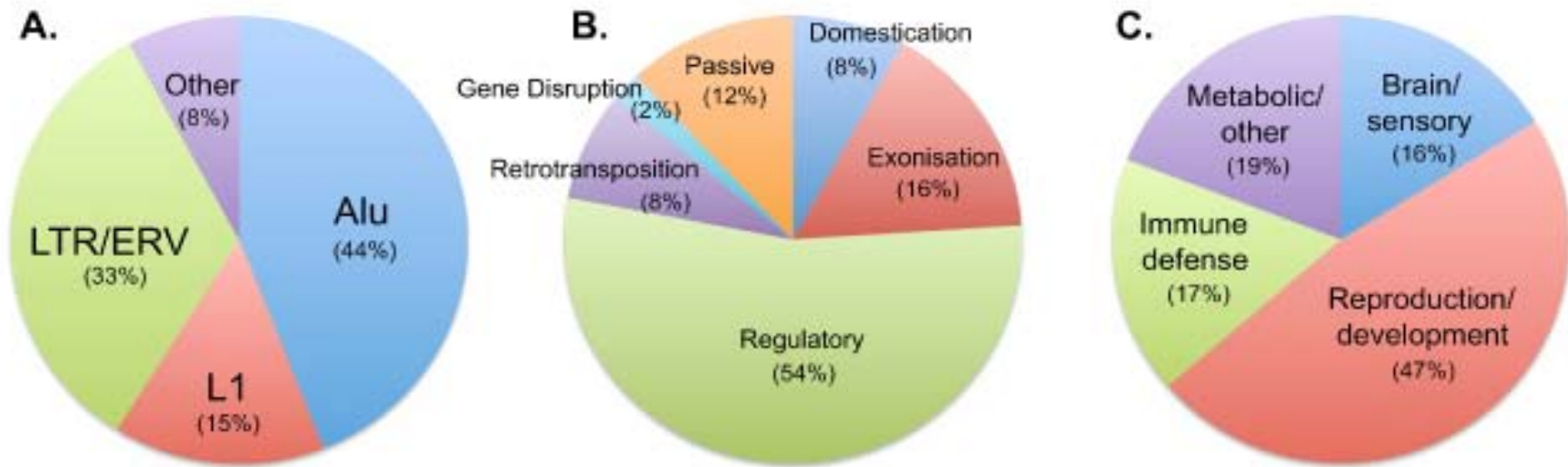


Figure 3-1: Summary of the effect of TES on primate evolution. (A) Transposable elements (TEs) implicated in the generation of primate-specific traits. **(B)** Types of events mediated by TEs underlying primate-specific traits. Passive events entail TE-mediated duplications, inversions or deletions. **(C)** Aspects of primate phenotype affected by TEs. Based on the published data shown in Tables 3-3 to 3-6.

3.3 The TE-Thrust Hypothesis

The ubiquitous, very diverse, and mostly extremely ancient TEs are powerful facilitators of genome evolution, and therefore of phenotypic diversity. TE-Thrust acts to build, sculpt and reformat genomes, either actively by TE transposition and integration (active TE-Thrust), or passively, because after integration, TEs become dispersed homologous sequences that facilitate ectopic DNA recombination (passive TE-Thrust). TEs can cause very significant and/or complex coding, splicing, regulatory and karyotypic changes to genomes, resulting in phenotypes that can adapt well to biotic or environmental challenges, and can often invade new ecological niches. TEs are usually strongly controlled in the soma, where they can be damaging (Matzke et al. 1999; Schulz et al. 2006), but they are allowed some limited mobility in the germline and early embryo (Dupressoir and Heidmann 1996; Brouha et al. 2002; van den Hurk et al. 2007), where, although they can occasionally be harmful, they can also cause beneficial changes that can become fixed in a population, benefiting the existing lineage, and sometimes generating new lineages.

There is generally no Darwinian selection for individual TEs or TE families, although there may be exceptions, such as the primate-specific Alu SINEs in gene-rich areas (Lander et al. 2001; Walters et al. 2009). Instead, according to the TE-Thrust hypothesis, there is differential survival of those lineages that

contain or can acquire suitable germline repertoires of TEs, as these lineages can more readily adapt to environmental or ecological changes, and can potentially undergo, mostly intermittently, fecund radiations. We hypothesise that lineages lacking a suitable repertoire of TEs are, if all else is equal, liable to stasis, possibly becoming “living fossils” or even becoming extinct.

TE activity is usually intermittent (Haring et al. 2000; Gerasimova et al. 1985; Kim et al. 2004; Marques et al. 2005; Ray et al. 2008), with periodic bursts of transposition due to interplay between various cellular controls, various stresses, *de novo* syntheses, *de novo* modifications, new infiltrations of TEs (by horizontal transfer), or new endogenisations of retroviruses. However, the vast majority of viable TEs usually undergo slow mutational decay and become non-viable (incapable of activity), although some super-families have remained active for more than 100 Myr. Episodic TE activity and inactivity, together with differential survival of lineages, suggests an explanation for punctuated equilibrium, evolutionary stasis, fecund lineages, and adaptive radiations, all found in the fossil record, and for extant “fossil species” (Oliver and Greene 2009a,b; Zeh et al. 2009).

TE-Thrust is expected to be optimal in lineages in which TEs are active and/or those that possess a high content of homogeneous TEs, both of which can promote genomic dynamism (Oliver and Greene 2009a). We hypothesise four main modes of TE-Thrust

(Table 3-1), but as these are extremes of continuums, many intermediate modes are possible.

- Mode 1: periodically active heterogeneous populations of TEs result in stasis with the potential for intermittent punctuation events.
- Mode 2: periodically active homogenous populations of TEs result in: 1) gradualism as a result of ectopic recombination, if the TE population is large, with the potential for periodic punctuation events, or 2) stasis with the potential for periodic punctuation events if the TE population is small.
- Mode 3: non-viable heterogeneous populations of TEs, in the absence of new infiltrations, result in prolonged stasis, which can sometimes result in extinctions and/or “living fossils”.
- Mode 4: non-viable homogenous populations of TEs, in the absence of new infiltrations, can result in: 1) gradualism as a result of ectopic recombination, if the TE population is large or 2) stasis if the TE population is small.

These modes of TE-Thrust are in agreement with the findings of palaeontologists (Gould 2002) and some evolutionary biologists (Ridley 2004) that punctuated equilibrium is the most common mode of evolution, but that gradualism and stasis also occur. Many extant “living fossils” are also known.

We acknowledge that TE-Thrust acts by enhancing evolutionary potential, and whether that potential is actually realized is heavily influenced by environmental, ecological and other factors.

Moreover, there are many other “engines” of evolution besides TE-Thrust, such as point mutation (Pollard et al. 2006), simple sequence repeats (Kashi and King 2006), endosymbiosis (Margulis and Chapman 1998), epigenetic modification (Monk 1995) and whole-genome duplication (McLysaght et al. 2002), among others. These often complement TE-Thrust; for example, point mutations can endow duplicated or retrotransposed genes with new functions (Dulai et al. 1999; Burki and Kaessmann 2004). There may also be other, as yet unknown, or hypothesised but unconfirmed, “engines” of evolution.

3.4 Higher Primate Genomes are very suited to TE-Thrust as they Possess Large Homogeneous Populations of TEs

Human and other extant higher primate genomes are well endowed with a relatively small repertoire of TEs (Table 3-2). These TEs, which have been extensively implicated in engineering primate-specific traits (Table 3-3; Table 3-4; Table 3-5; Table 3-6), are largely relics of an evolutionary history marked by periodic bursts of TE activity (Kim et al. 2004; Batzer and Deininger 2002; Bailey et al. 2003). TE activity is presently much reduced, but extant simian lineage genomes remain well suited for passive TE-Thrust, with just two elements, Alu and L1, accounting for over 60% of the total TE DNA sequence (Lander et al. 2001; Mikkelsen et al. 2005; Gibbs et al. 2007). In humans, there are 10 times as many mostly

Table 3-1: Hypothesised Major Modes of Transposable Element (TE) Thrust

Mode	TE Activity	TE homogeneity	TE population size	Evolutionary outcome	Type of TE thrust
1	Viable and intermittently active	Heterogeneous	Large	Stasis with punctuation events	Active
			Small	Stasis with punctuation events	Active
2	Viable and intermittently active	Homogeneous	Large	Gradualism with punctuation events	Active and passive
			Small	Stasis with punctuation events	Active
3	Non-viable/inactive	Heterogeneous	Large	Stasis ^{a,b}	Minimal ^c
			Small	Stasis ^{a,b}	Minimal ^c
4	Non-viable/inactive	Homogeneous	Large	Gradualism ^a	Passive ^c
			Small	Stasis ^{a,b}	Minimal ^c

^aUnless new infiltrations or reactivation of TEs occur.

^bFossil taxa are a possible outcome of prolonged stasis

^cInactive/non-viable TEs can be exapted in a delayed fashion, which could cause some resumption of active TE-Thrust.

homogeneous class I retro-TEs as there are very heterogeneous class II DNA-TEs (Lander et al. 2001). Only L1, Alu, SVA (SINE-R, variable number of tandem repeats (VNTR), Alu) and possibly some ERVs, remain active in humans (Mills et al. 2007).

L1 and the primate-specific Alu predominate in simians (Lander et al. 2001; Mikkelsen et al. 2005; Gibbs et al. 2007), and thus strongly contribute to TE-Thrust in this lineage (Figure 3-1A). The autonomous L1 is almost universal in mammals, whereas the non-autonomous Alu, like most SINEs, is conspicuously lineage-specific, having been synthesized *de novo*, extremely unusually, from a 7SL RNA-encoding gene. The confinement of Alu to a single mammalian order is typical of younger SINEs, whereas ancient SINEs, or exapted remnants of them, may be detectable across multiple vertebrate classes (Gilbert and Labuda 1999). Alu possesses additional unusual characteristics: extreme abundance (1.1 million copies, occurring every 3 kb on average in the human genome), frequent location in gene-rich regions, and a lack of evolutionary divergence (Lander et al. 2001; Labuda and Striker 1989). Their relatively high homology is most easily explained as being the result of functional selection helping to prevent mutational drift. Thus, Alus have been hypothesised to serve biological functions in their own right, leading to their selection and maintenance in the primate genome (Walters et al. 2009). For example, A-to-I RNA editing, which has a very high prevalence in the human genome, mainly occurs within Alu elements (Levanon et al. 2004), which would seem to provide primates with a genetic sophistication beyond that of other

mammals. Alus may therefore not represent a peculiar, evolutionary neutral invasion, but rather positively selected functional elements that are resistant to mutational degradation (Mattick and Mehler 2008). This has significance for TE-Thrust, as it would greatly prolong the usefulness of Alus as facilitators of evolution within primate lineages. Other human retro-TEs include the fossil tRNA mammalian-wide interspersed repeat (MIR) SINE, which amplified approximately 130 Mya (Lander et al. 2001; Krull et al. 2007) and the much younger SVA, a non-autonomous composite element partly derived from ERV and Alu sequences, which is specific to the great apes and humans (Ostertag et al. 2003). Like Alus, SVAs are mobilised by L1-encoded enzymes and, similar to Alu, a typical full-length SVA is GC-rich, and thus constitutes a potential mobile CpG island. Importantly, ERVs are genome builders/modifiers of exogenous origin (Mayer and Meese 2005). Invasion of ERVs seems to be particularly associated with a key mammalian innovation, the placenta (Table 3-4). The endogenisation of retroviruses and the horizontal transfer of TEs into germlines clearly show that the Weismann Barrier is permeable, contrary to traditional theory.

The DNA-TEs, which comprise just 3% of the human genome, are extremely diverse, but are now completely inactive (Lander et al. 2001; Pace and Feschotte 2007). Although some have been exapted within the simian lineage as functional coding sequences (Table 3-3; to Table 3-6), DNA-TEs, it seems, cannot now be a significant factor for TE-Thrust in primates, unless there are new infiltrations.

3.5 TE-Thrust Influences Evolutionary Trajectories

A key proposal of our TE-Thrust hypothesis is that TEs can promote the origin of new lineages and drive lineage divergence through the engineering of specific traits. Ancestral TEs shared across very many lineages can, by chance, lead to the delayed generation of traits in one lineage but not in another. For example, more than 100 copies of the ancient amniote-distributed AmnSINE1 are conserved as non-coding elements specifically among mammals (Nishihara et al. 2006). However, as they often show a narrow lineage specificity, we hypothesise that younger SINEs (with their partner LINEs) may have a large influence upon the trajectory and the outcomes of the evolution within clades, as is apparent with the Alu/L1 pair in primates (Figure 3-1A). Probably not all SINEs are equal in this ability; it seems that some SINEs are more readily mobilised than others, and when mobilised, some SINEs are more effective than others at facilitating evolution by TE-Thrust. The extremely abundant primate Alu dimer seems to illustrate this. Whereas the overwhelming majority of SINEs are derived from tRNAs, Alus may have proliferated so successfully because they are derived from the 7SL RNA gene (Ullu and Tschudi 1984), which is part of the signal recognition particle (SRP) that localises to ribosomes. Alu RNAs can therefore bind proteins on the SRP and thus be retained on the ribosome, in position to be retrotransposed by newly synthesized proteins encoded by their partner L1 LINEs (Dewannieux et al. 2003).

Among the primates, the simians have undergone the greatest evolutionary transitions and radiation. Of the approximately 367 extant primate species, 85% are simians, with the remainder being prosimians, which diverged about 63 Mya. Significantly, large amplifications of L1, and thus of Alus and other sequences confined to simians, offer a plausible explanation for the lack of innovation in the trajectory of evolution in the prosimian lineages, compared with the innovation in the simian lineages. Since their divergence from the basal primates, the simians have experienced repeated periods of intense L1 activity that occurred from about 40 Mya to about 12 Mya (Khan et al. 2006). The highly active simian L1s were responsible for the very large amplification of younger Alus and of many gene retrocopies (Ohshima et al. 2003). Possibly, differential activity of the L1/Alu pair may have driven the trajectory and divergence of the simians, compared with the prosimians. The greater endogenisation of some retroviruses in simians compared with prosimians (Bénit et al. 1999) may also have played a part. These events may also explain the larger genome size of the simians compared with prosimians (Liu et al. 2003).

A significant feature of Alus is their dimeric structure, involving a fusion of two slightly dissimilar arms (Quentin 1992). This added length and complexity seems to increase their effectiveness as a reservoir of evolutionarily useful DNA sequence or as an inducer of ectopic recombination. It may therefore be no coincidence that simian genomes are well endowed with dimeric Alus. Viable SINEs in the less fecund and less evolutionary

innovative prosimians are heterogeneous, and include the conventional dimeric Alu, Alu-like monomers, Alu/tRNA dimers and tRNA SINEs (Schmid 1998). This distinctly contrasts with simian SINEs; in simians, viable SINEs are almost entirely dimeric Alus. Thus, both qualitatively and quantitatively, the Alu dimer seems to represent a key example of the power of a SINE to strongly influence evolutionary trajectory.

Although these coincident events cannot, by themselves, be a clear indication of cause and effect, distinct Alu subfamilies (*AluJ*, *AluS*, *AluY*) correlate with the divergence of simian lineages (Batzer and Deininger 2002; Bailey et al. 2003). Whereas the *AluJ* subfamily was active about 65 Mya when the separation and divergence between the simians and the prosimians occurred, the *AluS* subfamily was active beginning at about 45 Mya, when the Old World monkey proliferation occurred, followed by a surge in *AluY* activity and expansion beginning about 30 Mya, contemporaneous with the split between apes and Old World monkeys (Batzer and Deininger 2002; Bailey et al. 2003). Thus, periodic expansions of Alu subfamilies in particular seem to correspond temporally with major divergence points in primate evolution. More recent Alu activity may be a factor in the divergence of the human and chimpanzee lineages, with Alus having been three times more active in humans than in chimpanzees (Mikkelsen et al. 2005; Mills et al. 2006). Moreover,

Table 3-2: Summary of the Major Transposable Elements (TEs) found in Humans

	Family	Percentage of genome	Number in genome	Average length, bp	Maximum length, kb	Viable	Potentially autonomous
Type II: retro-TEs	LTR ^a /ERV ^b	8.3	443,000	510	10	No	Yes (via reverse transcriptase)
	LINE1 ^c	16.9	516,000	900	6	Some	Yes (via reverse transcriptase)
	LINE2	3.2	315,000	280	5	No	Yes (via reverse transcriptase)
	Alu SINE ^d	10.6	1,090,000	270	0.3	Yes	No
	MIR ^e SINE	2.2	393,000	150	0.26	No	No
	SVA ^f SINE-like composite	0.2	3,000	1,400	3	Yes	No
Type II: DNA-TEs	Many	2.8	294,000	260	3	No	Some (via transposase)

^aLTR = Long terminal repeat

^bERV = endogenous retrovirus

^cLINE = short interspersed nuclear element

^dSINE = short interspersed nuclear element

^eMIR = mammalian-wide interspersed repeat

^fSVA = SINE-VNTR-Alu

Table 3-3: Specific Examples of Transposable Elements (TEs) Implicated in Primate-specific Traits: Brain and Sensory

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>snaRs</i>	Cell growth and translational regulation	Alu	Afr. great ape/human	Domestication	Novel genes	Brain, testis	Active	Parrott and Mathews, 2009
	<i>BCYRN1</i>	Translational regulation of dendritic proteins	Alu	Simian	Domestication	Novel gene	Brain	Active	Watson and Sutcliffe, 1987
	<i>FLJ33706</i>	Unknown	Alu	Human	Domestication	Novel Gene	Brain	Active	Li et al., 2010
Neuronal stability?	<i>SETMAR</i>	DNA repair and replication	Hsmar1	Simian	Exonization	Novel fusion gene	Brain, various	Active	Cordaux et al., 2006
	<i>Survivin</i>	Anti-apoptotic/ brain development	Alu	Ape	Exonization	Novel Isoform	Brain, spleen	Active	Mola et al., 2007

Table 3-3 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>ADARB1</i>	RNA editing/ neurotransmitter receptor diversity	Alu	>Human	Exonization	Novel isoform	Brain, various	Active	Lai et al., 1997
	<i>CHRNA1</i>	Synaptic transmission	MIR ^b	Great ape	Exonization	Novel isoform	Neuro-muscular	Active	Krull et al., 2007
	<i>ASMT</i>	Melatonin synthesis	LINE-1 ^c	>Human	Exonization	Novel isoform	Pineal gland	Active	Rodriguez et al., 1994
	<i>CHRNA3</i>	Synaptic transmission	Alu	Great ape	Regulatory	Major promoter	Nervous system	Active	Fornasari et al., 1997
	<i>CHRNA6</i>	Synaptic transmission	Alu	>Human	Regulatory	Negative regulation	Brain	Active	Ebihara et al., 2002

Table 3-3 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>NAIP</i>	Anti-apoptosis (motor neuron)	Alu	>Human	Regulatory	Alternative promoters	CNS, various	Active	Romanish et al., 2009
	<i>CNTNAP4</i>	Cell recognition/adhesion	ERV ^d	>Human	Regulatory	Alternative promoter	Brain, testis	Active	van de Lagemaat et al., 2003
	<i>CCRK</i>	Cell cycle-related kinase	Alu	Simian	Regulatory	CpG island	Brain	Active	Farcas et al., 2009
Enhanced cognitive capacity/memory?	<i>GLUD2</i>	Neurotransmitter recycling	Unknown	Ape	Retrotransposition	Novel gene	Brain	Active	Burki and Kaessman 2004
Altered auditory perception?	<i>CHRNA9</i>	Cochlea hair development/modulation of auditory stimuli	Alu	Human	Deletion	Exon loss	Cochlea, sensory ganglia	Passive	Sen et al., 2006

Table 3-3 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Trichromatic colour vision	<i>OPN1LW</i>	Cone photoreceptor	Alu	Old World primate	Duplication	Novel gene	Retina	Passive	Dulai et al., 1999

^a > = Maximum known distribution

^bMIR = mammalian-wide interspersed repeat

^cLINE = long interspersed nuclear element

^dERV = endogenous retrovirus

Table 3-4: Specific Examples of Transposable Elements (TEs) Implicated in Primate-specific Traits: Reproduction and Development

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Placental morphogenesis	<i>Syncytin-1</i>	Trophoblast cell fusion	ERV ^b	Ape	Domestication	Novel gene	Placenta	Active	Mi et al., 2000
Placental morphogenesis	<i>Syncytin-2</i>	Trophoblast cell fusion	ERV	Simian	Domestication	Novel gene	Placenta	Active	Blaise et al., 2003
	<i>HERW1</i>	Unknown	ERV	Simian	Domestication	Novel gene	Placenta	Active	Kjeldbjerg et al., 2008
	<i>HERW2</i>	Unknown	ERV	Simian	Domestication	Novel gene	Placenta	Active	Kjeldbjerg et al., 2008
	<i>ERV3</i>	Development and differentiation?	ERV	Old World primate	Domestication	Novel gene	Placenta, various	Active	Larsson et al., 1994

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>DNMT1</i>	DNA methylation	Alu	>Afr. great ape	Exonization	Novel isoform	Fetal, various	Active	Hsu et al., 1999
	<i>LEPR</i>	Leptin receptor	SVA	Human	Exonization	Novel isoform	Fetal liver	Active	Damert et al., 2004
	<i>IL22RA2</i>	Regulation of inflammatory responses/ interleukin-22 decoy receptor	LTR ^c	Great ape	Exonization	Novel isoform	Placenta	Active	Piriyapongsa et al., 2007b
	<i>PPHLN1</i>	Epithelial differentiation/ nervous-system development	ERV/Alu/ LINE-1 ^d	Ape	Exonization	Novel isoforms	Fetal, various	Active	Huh et al., 2006
	<i>CGB1/2</i>	Chorionic gonadotropin	Alu (snaR-G1/2)	Afr. great ape	Regulatory	Major promoter	Testis	Active	Parrott and Mathews, 2009

Table3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>GSDMB</i>	Epithelial development	Alu	Ape	Regulatory	Major promoter	Stomach	Active	Komiyama et al., 2010
	<i>HYAL4</i>	Hyaluronidase	LINE-1/Alu	>Human	Regulatory	Major promoter	Placenta	Active	van de Lagemaat et al., 2003
Placental oestrogen synthesis	<i>HSD17B1</i>	Oestrogen synthesis	ERV	>Human	Regulatory	Major promoter	Ovary, placenta	Active	Cohen et al., 2009
Placental development	<i>INSL4</i>	Regulation of cell growth and metabolism	ERV	Old World primate	Regulatory	Major promoter	Placenta	Active	Bieche et al., 2003
	<i>DSCR4</i>	Unknown reproductive function	ERV	Ape	Regulatory	Major promoter	Placenta, testis	Active	Dunn et al., 2006

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>DSCR8</i>	Unknown reproductive function	ERV	>Ape	Regulatory	Major promoter	Placenta, testis	Active	Dunn et al., 2006
	<i>CGA</i>	Common subunit of chorionic gonadotropin, luteinizing, follicle-stimulating and thyroid-stimulating hormones	Alu	>Simian	Regulatory	Negative regulation	Placenta, pituitary gland	Active	Scofield et al., 2000
Globin switching	<i>HBE1</i>	Embryonic oxygen transport	Alu	>Human	Regulatory	Negative regulation	Fetal	Active	Wu et al., 1990
	<i>GH</i>	Growth hormone	Alu	>Human	Regulatory	Negative regulation	Pituitary gland	Active	Trujillo et al., 2006

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>WT1</i>	Urogenital development	Alu	>Human	Regulatory	Negative regulation	Urogenital	Active	Hewitt et al., 1995
Efficient placental gas exchange	<i>HBG1</i>	Fetal oxygen transport	LINE-1	Old World primate	Regulatory	Tissue-specific enhancer	Fetal	Active	Johnson et al.b, 2006
Placental leptin secretion	<i>LEP</i>	Metabolic regulatory hormone	LTR	>Human	Regulatory	Tissue-specific enhancer	Placenta	Active	Bi et al., 1997
	<i>MET</i>	Hepatocyte growth-factor receptor	LINE-1	> Afr. great ape	Regulatory	Alternative promoter	Liver, Pancreas, Lung	Active	Nigumann et al., 2002
	<i>BCAS3</i>	Embryogenesis/erythropoiesis	LINE-1	> Afr. great ape	Regulatory	Alternative promoter	Fetal, various	Active	Wheelan et al., 2005

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>CHRM3</i>	Synaptic transmission	LINE-1	Human	Regulatory	Alternative promoter	Placenta	Active	Huh et al., 2009
	<i>CLCN5</i>	Chloride transporter	LINE-1	>Human	Regulatory	Alternative promoter	Placenta	Active	Matlik et al., 2006
	<i>SLC01A2</i>	Organic anion transporter	LINE-1	>Human	Regulatory	Alternative promoter	Placenta	Active	Matlik et al., 2006
	<i>CHRM3</i>	Synaptic transmission	LTR	Human	Regulatory	Alternative promoter	Testis	Active	Huh et al., 2009
	<i>IL2RB</i>	Growth-factor receptor	LTR	>Human	Regulatory	Alternative promoter	Placenta	Active	Cohen et al., 2009
Placental development	<i>ENTPD1</i>	Thromboregulation	LTR	>Human	Regulatory	Alternative promoter	Placenta	Active	van de Lagemaat et al., 2003

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>MKKS</i>	Molecular chaperone	LTR/LINE-2	>Human	Regulatory	Alternative promoter	Testis, fetal	Active	van de Lagemaat et al., 2003
	<i>NAIP</i>	Anti-apoptosis	ERV	>Human	Regulatory	Alternative promoter	Testis	Active	Romanish et al., 2007
	<i>EDNRB</i>	Placental development/circulation	ERV	>Human	Regulatory	Alternative promoter	Placenta	Active	Medstrand et al., 2001
Placental development	<i>PTN</i>	Growth factor	ERV	Ape	Regulatory	Alternative promoter	Trophoblast	Active	Schulte et al., 1996
	<i>MID1</i>	Cell proliferation and growth	ERV	Old World primate	Regulatory	Alternative promoter	Placenta, fetal kidney	Active	Landry et al., 2002

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Placental oestrogen synthesis	<i>NOS3</i>	Endothelial nitric oxide synthesis	ERV	>Human	Regulatory	Alternative promoter	Placenta	Active	Huh et al., 2008b
	<i>GSDMB</i>	Epithelial development	ERV	Ape	Regulatory	Alternative promoter	Various	Active	Sin et al., 2006
	<i>CYP19</i>	Oestrogen synthesis	ERV	Simian	Regulatory	Alternative promoter	Placenta	Active	van de Lagemaat et al., 2003
	<i>AMACs</i>	Fatty-acid synthesis	SVA	Afr. great ape	Retrotransposition	Novel genes	Placenta, testis	Active	Xing et al., 2006
	<i>POTEs</i>	Pro-apoptosis/spermatogenesis	LINE-1	Ape	Retrotransposition	Novel fusion genes	Testis, ovary, prostate, placenta	Active	Lee et al., 2006

Table3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>PIPSL</i>	Intracellular protein trafficking	LINE-1	>Great ape	Retrotransposition	Novel fusion gene	Testis	Active	Babushok et al., 2007
	<i>CDYs</i>	Chromatin modification	Unknown	Simian	Retrotransposition	Novel genes	Testis	Active	Lahn and Page, 1999
	<i>ADAM20/21</i>	Membrane metalloprotease	Unknown	>Human	Retrotransposition	Novel genes	Testis	Active	Betran and Long, 2002
Placental growth hormone secretion	<i>GH</i>	Placental growth hormone	Alu	Simian	Duplication	Novel genes	Placenta	Passive	De Mendoza et al., 2004
	<i>Chr19 miRNAs</i>	Unknown	Alu	Simian	Duplication	Novel genes	Placenta	Passive	Zhang et al., 2008

Table 3-4 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Enhanced immune tolerance at fetal-maternal interface	<i>LGALS13/14/16</i>	Carbohydrate recognition/immune regulation	LINE-1	Simian	Duplication	Novel genes	Placenta	Passive	Than et al., 2009
Efficient placental gas exchange	<i>HBG2</i>	Fetal oxygen transport	LINE-1	Simian	Duplication	Novel gene	Fetal	Passive	Fitch et al., 1991

^a > = Maximum known distribution

^bERV = endogenous retrovirus

^cLTR = long terminal repeat

^dLINE = long interspersed nuclear element

Table: 3-5 Specific Examples of Transposable Elements (TEs) Implicated in Primate-specific Traits: Immune Defence

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Soluble CD55	<i>CD55</i>	Complement regulation	Alu	>Human	Exonization	Novel isoform	Various	Active	Caras et al., 1987
Intracellular TNFR	<i>P75TNFR</i>	Tumour necrosis factor receptor	Alu	Old World primate	Exonization	Novel isoform	Various	Active	Singer et al., 2004
Altered infectious-disease resistance?	<i>IRGM</i>	Intracellular pathogen resistance	ERV ^b	Afr. Great Ape	Regulatory	Major promoter	Various	Active	Bekpen et al., 2009
Altered infectious-disease resistance?	<i>IL29</i>	Antiviral cytokine	Alu/LTR ^c	Human	Regulatory	Positive regulation	Dendritic cells, epithelial cells	Active	Thomson et al., 2009

Table 3-5 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>FCER1G</i>	IgE/IgG Fc receptor/ T cell antigen receptor	Alu	Ape	Regulatory	Positive/negative regulation	T cells, basophils	Active	Brini et al., 1993
	<i>CD8A</i>	T cell interaction with class I MHC	Alu	Ape	Regulatory	Tissue-specific enhancer	T cells	Active	Hambor et al., 1993
Red cell ABH antigen	<i>FUT1</i>	Fucosyltransferase	Alu	Ape	Regulatory	Alternative promoter	Erythrocytes	Active	Apoil et al., 2000
	<i>TMPRSS3</i>	Membrane serine protease	Alu/LTR	>Human	Regulatory	Alternative promoter	Peripheral blood leukocytes	Active	van de Lagemaat et al., 2003
Colon Le antigen expression	<i>B3GALT5</i>	Galactosyltransferase	ERV	Old World primate	Regulatory	Alternative promoter	Colon, small intestine, breast	Active	Dunn et al., 2003

Table 3-5 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Prolactin potentiation of the adaptive immune response	<i>PRL</i>	Regulation of lactation and reproduction	ERV	Old World primate	Regulatory	Alternative promoter	Lymphocytes, endometrium	Active	Gerlo et al., 2006
	<i>ST6GAL1</i>	Sialyltransferase	ERV	>Human	Regulatory	Alternative promoter	B lymphocytes	Active	van de Lagemaat et al., 2003
Vitamin D regulation of cathelicidin antimicrobial peptide gene	<i>CAMP</i>	Antimicrobial peptide	Alu	Simian	Regulatory	Vitamin D responsiveness	Myeloid cells, various	Active	Gombart et al., 2009

Table 3-5 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>MPO</i>	Myeloperoxidase/ microbicidal enzyme	Alu	>Human	Regulatory	Thyroid hormone/ retinoic acid responsiveness	Myeloid cells	Active	Piedrafita et al., 1996
Altered infectious disease resistance?	<i>IFNG</i>	Antiviral/ immunoregulatory factor	Alu	Old World primate	Retrotransposition	Novel positive regulatory element	Natural killer cells, T cells	Active	Ackerman et al., 2002
Absence of N-glycolylneuraminic acid/ altered infectious-disease resistance?	<i>CMAH</i>	N-glycolylneuraminic acid synthesis	Alu	Human	Gene disruption	Gene loss	Various	Active	Hayakawa et al., 2001

Table 3-5 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>IRGM</i>	Intracellular pathogen resistance	Alu	Old and New World monkey	Gene disruption	Gene loss	Various	Active	Bekpen et al., 2009
Altered malaria resistance?	<i>HBA2</i>	Oxygen transport	Alu	>Ape	Duplication	Novel gene	Erythrocyts	Passive	Hess et al., 1983

^a > = Maximum known distribution

^bERV = endogenous retrovirus

^cLTR = long terminal repeat

Table 3-6: Specific Examples of Transposable Elements (TEs) Implicated in Primate-specific Traits: Metabolic and Other

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>RNF19A</i>	Ubiquitin ligase	Alu	> Human	Exonization	Novel isoform	Various	Active	Huh et al., 2008
	<i>BCL2L11</i>	Pro-apoptotic	Alu	> Human	Exonization	Novel isoform	Various	Active	Wu et al., 2007
	<i>BCL2L13</i>	Pro-apoptotic	Alu	> Human	Exonization	Novel isoform	Various (cytosolic instead of mitochondrial)	Active	Yi et al., 2003
	<i>SFTPB</i>	Pulmonary surfactant	Alu/ERV ^b	Primate	Exonization	Novel isoform	Various	Active	Lee et al., 2009
Efficiency of ZNF177 transcription and translation	<i>ZNF177</i>	Transcriptional regulator	Alu/LINE-1 ^c /ERV	> Human	Exonization	Novel isoform	Various	Active	Landry et al., 2001

Table 3-6 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
Production of salivary amylase	<i>AMY1s</i>	Starch digestion	ERV	Old World primate	Regulatory	Major promoter	Salivary gland	Active	Ting et al., 1992
	<i>BAAT</i>	Bile metabolism	ERV	> Human	Regulatory	Major promoter	Liver	Active	van de Lagemaat et al, 2003
	<i>CETP</i>	Cholesterol metabolism	Alu	> Human	Regulatory	Negative regulation	Liver	Active	Le Goff et al., 2003
Absence of FMO1 in adult liver/ altered drug metabolism?	<i>FMO1</i>	Xenobiotic metabolism	LINE-1	> Human	Regulatory	Negative regulation in liver	Kidney	Active	Shephard et al., 2007
	<i>RNF19A</i>	Ubiquitin ligase	LTR ^d	> Human	Regulatory	Alternative promoter	Various	Active	Huh et al., 2008a

Table 3-6 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>APOC1</i>	Lipid metabolism	ERV	Ape	Regulatory	Alternative promoter	Various	Active	Medstrand et al, 2001
	<i>KRT18</i>	Epithelial keratin	Alu	> Human	Regulatory	Retinoic acid responsiveness	Various	Active	Vasant and Reynolds, 1995
	<i>PTH</i>	Parathyroid hormone	Alu	> Old World primate	Regulatory	Negative calcium responsiveness	Parathyroid gland	Active	McHaffie and Ralston, 1995
	<i>PRKACG</i>	cAMP signalling/regulation of metabolism	Unknown	> Old World primate	Retrotransposition	Novel gene	Various	Active	Reinton et al., 1998
	<i>NBR2</i>	Unknown	Alu	Old World primate	Duplication	Novel gene	Various	Passive	Jin et al., 2004

Table 3-6 (continued)

TE generated trait	Gene affected	Gene function	TE responsible	Distribution ^a	Type of event	Effect	Tissue expression	Type of TE-Thrust	Reference
	<i>LRRC37A</i>	Unknown	Alu	Old World primate	Duplication	Novel genes	Various	Passive	Jin et al., 2004
	<i>ARF2</i>	GTPase/vesicle trafficking	Alu	Great ape	Inversion	Novel fusion gene	Various	Passive	Jin et al, 2004
Altered arterial wall function?	<i>ELN</i>	Elastin	Alu	> Old World primate/human	Deletion	Exon losses	Various	Passive	Szabo et al., 1999
Low body mass?	<i>ASIP</i>	Energy metabolism/pigmentation	Alu	Lesser ape (gibbon)	Deletion	Gene loss	Various	Passive	Nakayama and Ishida, 2006

^a > = Maximum known distribution

^bERV = endogenous retrovirus

^cLINE = long interspersed nuclear element

^dLTR = long terminal repeat

at least two new Alu subfamilies (*AluYa5* and *AluYb8*) have amplified specifically within the human genome since the human-chimpanzee split (Mikkelsen et al. 2005; Mills et al. 2006; Hedges et al. 2004).

Passive TE-Thrust mediated by the Alu/L1 pair has also been evident as a force contributing to lineage divergence in the primates. Ectopic recombinations between Alus, in particular, are a frequent cause of line-age-specific deletion, duplication or rearrangement. Comparisons between the human and chimpanzee genomes have revealed the extent to which they have passively exerted their effects in the relatively recent specific evolutionary history of primates. An examination of human-Alu recombination-mediated deletion (ARMD) identified 492 ARMD events responsible for the loss of about 400 kb of sequence in the human genome (Sen et al. 2006). Likewise, Han et al. (2007) reported 663 chimpanzee-specific ARMD events, deleting about 771 kb of genomic sequence, including exonic sequences in six genes. Both studies suggested that ARMD events may have contributed to the genomic and phenotypic diversity between chimpanzees and humans. L1-mediated recombination also seems to be a factor in primate evolution, with Han et al. (2005) reporting 50 L1-mediated deletion events in the human and chimpanzee genomes. The observed high enrichment of TEs such as Alu at low-copy-repeat junctions indicates that TEs have been an important factor in the generation of segmental duplications that are uniquely abundant in primate genomes (Bailey et al. 2003).

Such genomic duplications provide a major avenue for genetic innovation by allowing the functional specialisation of coding or regulatory sequences. Karyotypic changes are thought to be an important factor in speciation (Rieseberg 2001). Major differences between the human and chimpanzee genomes include nine pericentric inversions, and these have also been linked to TE-mediated recombination events (Kehrer-Sawatzki et al. 2005). It thus seems that both the active and passive effects of Alu and L1 have greatly facilitated and influenced the trajectory of simian evolution by TE-Thrust. Transfer RNA-type SINEs, with suitable partner LINEs, probably perform this role in other lineages.

3.6 TE-Thrust Affects Evolutionary Trajectory by Engineering Lineage-specific Traits

TEs can act to generate genetic novelties and thus specific phenotypic traits in numerous ways. Besides passively promoting exon, gene or segmental duplications (or deletions) by unequal recombination, or by disruption of genes via insertion, TEs can actively contribute to gene structure or regulation via exaptation. On multiple occasions, TEs have been domesticated to provide the raw material for entire genes or novel gene fusions (Volf 2006). More frequently, TEs have contributed partially to individual genes through exonization after acquisition of splice sites (Sela et al. 2010; Shen et al. 2011). Independent exons generated by TEs are often alternatively spliced, and thereby result in novel expressed isoforms that increase the size

of the transcriptome (Sorek et al. 2002). The generation of novel gene sequences during evolution seems to be heavily outweighed by genetic or epigenetic changes in the transcriptional regulation of pre-existing genes (Monk 1995; Carroll 2005). Consistent with this, much evidence indicates that a major way in which TEs have acted to functionally modify primate genomes is by actively inserting novel regulatory elements adjacent to genes, thus silencing or enhancing expression levels or changing expression patterns, often in a tissue-specific manner (Nigumann et al. 2002; Jordan et al. 2003; van de Lagemaat et al. 2003). Moreover, because they are highly repetitious and scattered, TEs have the capacity to affect gene expression on a genome-wide scale by acting as distributors of regulatory sequences or CpG islands in a modular form (Feschotte 2008). Many functional binding sites of developmentally important transcription factors have been found to reside on Alu repeats (Polak and Domany 2006). These include oestrogen receptor-dependent enhancer elements (Norris et al. 1995) and retinoic acid response elements, which seem to have been seeded next to retinoic acid target genes throughout the primate genome by the *AluS* subfamily (Vansant and Reynolds 1995). As a consequence, TEs are able to contribute significantly to the species-specific rewiring of mammalian transcriptional regulatory networks during pre-implantation embryonic development (Xie et al. 2010). Similarly, primate-specific ERVs have been implicated in shaping the human p53 transcriptional network (Wang et al. 2007) and

rewiring the core regulatory network of human embryonic stem cells (Kunarso et al. 2010).

Certain classes of retro-TEs can actively generate genetic novelty using their retrotranspositional mechanism to partially or fully duplicate existing cellular genes. Duplication is a crucial aspect of evolution, which has been particularly important in vertebrates, and constitutes the primary means by which organisms evolve new genes (Ohno 1970). LINEs and SVAs have a propensity to transduce host DNA due to their weak transcriptional termination sites, so that 3' flanking regions are often included in their transcripts. This can lead to gene duplication, exon shuffling or regulatory-element seeding, depending on the nature of the sequence involved (Burki and Kaessmann 2004; Moran et al. 1999; Goodier et al. 2000). Duplication of genes can also occur via the retrotransposition of mRNA transcripts by LINEs. Such genes are termed retrocopies, which, after subsequent useful mutation, can sometimes evolve into retrogenes, with a new, related function. There are reportedly over one thousand transcribed retrogenes in the human genome (Vinckenbosch et al. 2006), with about one new retrogene per million years having emerged in the human lineage during the past 63 Myr (Marques et al. 2005). Some primate retrogenes seem to have evolved highly beneficial functions, such as *GLUD2* (Burki and Kaessmann 2004).

3.7 Specific Evidence for TE-Thrust: Examples of Traits Engineered by TEs in the Higher Primates

TEs seem to have heavily influenced the trajectories of primate evolution and contributed to primate characteristics, as the simians in particular have undergone major evolutionary advancements in cognitive ability and physiology (especially reproductive physiology). The advancement and radiation of the simians seems to be due, in part and all else being equal, to exceptionally powerful TE-Thrust, owing to its especially effective Alu dimer, partnered by very active novel L1 families, supplemented by ERVs and LTRs. These have engineered major changes in the genomes of the lineage(s) leading to the simian radiations and major transitions. We identified more than 100 documented instances in which TEs affected individual genes and thus were apparently implicated at a molecular level in the origin of higher primate-specific traits (Table 3-3 to Table 3-6). The Alu SINE dominated, being responsible for nearly half of these cases, with ERVs/sLTRs being responsible for a third, followed by L1-LINEs at 15% (Figure 3-1A). Just 2% were due to the young SVAs, and 1% each to ancient MIR SINEs and DNA-TEs. More than half the observed changes wrought by TEs were regulatory (Figure 3-1B). As discussed below, TEs seem to have influenced four main aspects of the primate phenotype: brain and sensory function, reproductive physiology, immune defence, and metabolic/other (Figure 3-1C, and Table 3-3 to Table 3-6). Notably, ERVs, which are often highly transcribed in the germline and placenta (Prudhomme et al. 2005), were

strongly associated with reproductive traits, whereas Alus influenced these four aspects almost equally (Figure 3-2).

3.7.1 Brain and Sensory Function

The large brain, advanced cognition and enhanced colour vision of higher primates are distinct from those of other mammals. The molecular basis of these characteristics remains to be fully defined, but from evidence already available, TEs (particularly Alus) seem to have contributed substantially via the origination of novel genes and gene isoforms, or via altered gene transcription (Table 3). Most of the neuronal genes affected by TEs are restricted to the apes, and they seem to have roles in synaptic function and plasticity, and hence learning and memory. These genes include multiple neurotransmitter receptor genes and *glutamate dehydrogenase 2 (GLUD2)*, a retrocopy of *GLUD1* that has acquired crucial point mutations. *GLUD2* encodes glutamate dehydrogenase, an enzyme that seems to have increased the cognitive powers of the apes through the enhancement of neurotransmitter recycling (Burki and Kaessmann 2004). The cell cycle-related kinase (*CCRK*) gene represents a good example of how the epigenetic modification of TEs can be mechanistically linked to the transcriptional regulation of nearby genes (Farcas et al. 2009). In simians, this gene possesses regulatory CpGs contained within a repressor Alu element, and these CpGs are more methylated in the cerebral cortex of human compared with chimpanzee.

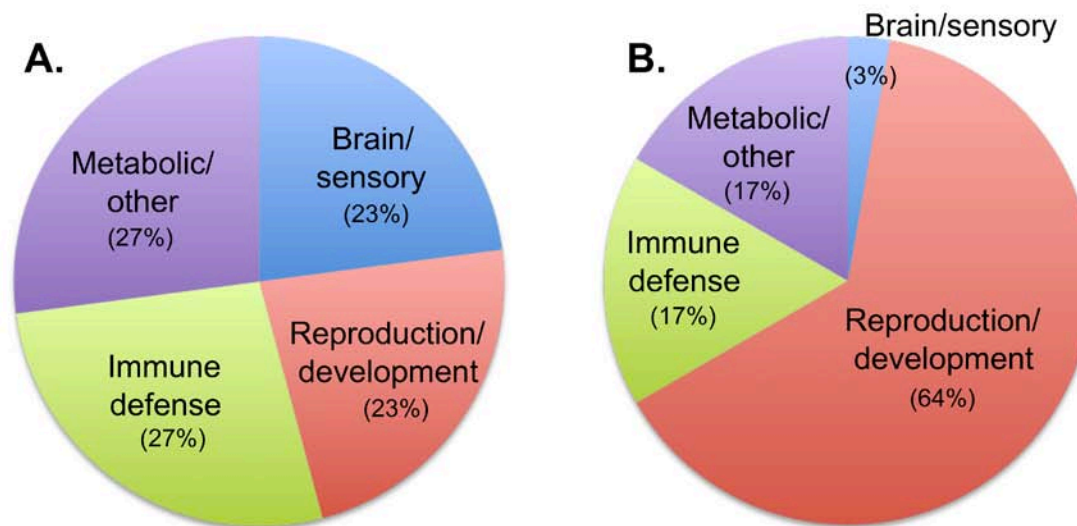


Figure 3-2 Comparison of aspects of primate phenotype affected by (A) Alu elements and (B) LTR/ERVs.

Based on the published data shown in Tables 3 to 6

Concordantly, *CCRK* is expressed at higher levels in the human brain (Farcas et al. 2009). TEs may also affect the brain at a somatic level, because embryonic neural progenitor cells have been found to be permissive to L1 activity in humans (Coufal et al. 2009). This potentially provides a mechanism for increasing neural diversity and individuality. As our human lineage benefits from a diversity of additional individual talents, as well as shared talents, this phenomenon, if confirmed, could increase the 'fitness' of the human lineage, and is entirely consistent with the concept of differential survival of lineages, as stated in our TE-Thrust hypothesis.

The trichromatic vision of Old World monkeys and apes immensely enhanced their ability to find fruits and other foods, and probably aided them in group identity. This trait evidently had its origin in an Alu-mediated gene-duplication event that occurred about 40 Mya, and subsequently resulted in two separate cone photoreceptor (opsin) genes (Dulai et al. 1999), the tandem *OPN1LW* and *OPN1MW*, which are sensitive to long- and medium-wave light respectively. Other mammals possess only dichromatic vision.

3.7.2 Reproductive Physiology

Compared with other mammals, simian reproduction is characterized by relatively long gestation periods and by the existence of a hemochorial-type placenta that has evolved additional refinements to ensure efficient fetal nourishment. Available data suggests that TE-Thrust has contributed much of the uniqueness of the higher primate placenta, which seems

Chapter 3: Mobile DNA and the TE-Thrust Hypothesis: Supporting Evidence from the Primates

to be more invasive than that of other mammals, and releases a large number of factors that modify maternal metabolism during pregnancy. These characteristics appear to be due to the generation of novel placenta genes and to various TEs having been exapted as regulatory elements to expand or enhance the expression of pre-existing mammalian genes in the primate placenta (Table 3-4). The growth hormone (*GH*) gene locus is particularly notable for having undergone rapid evolution in the higher primates compared with most other mammals. A crucial aspect of this evolutionary advance was a burst of gene-duplication events in which Alu-mediated recombination is implicated as a driving force (De Mendoza et al. 2004). The simians thus possess between five and eight *GH* gene copies, and these show functional specialisation, being expressed in the placenta, in which they are thought to influence fetal access to maternal resources during pregnancy (De Mendoza et al. 2004; Lacroix et al. 2002). Longer gestation periods in simians were accompanied by adaptations to ensure an adequate oxygen supply. One key event was an L1-mediated duplication of the *HBG* globin gene in the lineage leading to the higher primates, which generated *HBG1* and *HBG2* (Fitch et al. 1991). *HBG2* subsequently acquired expression specifically in the simian fetus, in which it ensures the high oxygen affinity of fetal blood for more efficient oxygen transfer across the placenta. Old World primates additionally express *HBG1* in the fetus, owing to an independent LINE insertion at the beta globin locus (Johnson et al. 2006b). Thus, the important process of placental gas exchange has been extensively improved by TEs in simians, in

contrast to that of many mammals, including prosimians, in which fetal and adult haemoglobins are the same.

Two prominent examples of functionally exapted genes whose sequences are entirely TE-derived are *syncytin-1* (*ERVWE1*) and *syncytin-2* (*ERVWE2*). Both of these primate-specific genes are derived from ERV envelope (*env*) genes (Mi et al. 2000; Blaise et al. 2003). The syncytins play a crucial role in simian placental morphogenesis by mediating the development of the fetomaternal interface, which has a fundamental role in allowing the adequate exchange of nutrients and other factors between the maternal bloodstream and the fetus. In a remarkable example of convergent evolution, which attests to the importance of this innovation, two ERV *env* genes, *syncytin-A* and *syncytin-B*, independently emerged in the rodent lineage about 20 Mya (Dupressoir et al. 2005), as did *syncytin-Ory1* within the lagomorphs 12-30 Mya, and these exhibit functional characteristics analogous to the primate syncytin genes (Heidmann et al. 2009). This example, as well as many others (Tables 3-3 to Table 3-6) suggests the possibility that TE-Thrust may be an important factor in convergent evolution, a phenomenon that can be difficult to explain by traditional theories.

3.7.3 Immune Defence

Immune-related genes were probably crucial to the primate lineage by affording protection from potentially lethal infectious diseases. TEs have been reported to contribute to higher primate-restricted transcripts, or to the expression of a wide variety of

Chapter 3: Mobile DNA and the TE-Thrust Hypothesis: Supporting Evidence from the Primates

immunologically relevant genes (Table 3-5). One example is the insertion of an AluY element into intron 1 of the fucosyltransferase (*FUT*)1 gene in an ancestor of humans and apes. This enabled erythrocytic expression of *FUT*1, and thus the ABO blood antigens (Apoil et al. 2000), an adaptation linked to the selective pressure by malarial infection (Cserti and Dzik 2007). A particularly good example of a primate-specific adaptation that can be accounted for by a TE is the regulation of the cathelicidin antimicrobial peptide (*CAMP*) gene by the vitamin D pathway. Only simians possess a functional vitamin D response element in the promoter of this gene, which is derived from the insertion of an AluSx element. This genetic alteration enhances the innate immune response of simians to infection, and potentially counteracts the anti-inflammatory properties of vitamin D (Gombart et al. 2009).

3.7.4 Metabolic/Other

TEs seem to underlie a variety of other primate adaptations, particularly those associated with metabolism (Table 3-6). A striking example, related to dietary change, was the switching of the expression of certain α -amylase genes (*AMY1A*, *AMY1B* and *AMY1C*) from the pancreas to the salivary glands of Old World primates. This event, which was caused by the genomic insertion of an ERV acting as a tissue-specific promoter (Ting et al. 1992), facilitated the utilization of a higher starch diet in some Old World primates. This included the human lineage, in which consumption of starch became increasingly important, as evidenced by the average human having about three times

more *AMY1* gene copies than chimpanzees (Perry et al. 2007). Another example was the loss of a 100 kb genomic region in the gibbons, due to homologous recombination between AluSx sites (Nakayama and Ishida 2006), resulting in gibbons lacking the *ASIP* gene involved in the regulation of energy metabolism and pigmentation, which may help to account for their distinctive low body mass, so beneficial for these highly active arboreal primates.

3.8 TE-Thrust and divergence of the human lineage

Human and chimpanzee genomes exhibit discernable differences in terms of TE repertoire, TE activity and TE-mediated recombination events (Lander et al. 2001; Mikkelsen et al. 2005; Khan et al. 2006; Mills et al. 2006; Hedges et al. 2004; Sen et al. 2006; Han et al. 2007; Han et al. 2005). Thus, although nucleotide substitutions to crucial genes are important (Pollard et al. 2006), TE-Thrust is likely to have made a significant contribution to the relatively recent divergence of the human lineage (Cordaux and Batzer 2009; Britten 2010). In support of this, at least eight of the examples listed (Table 3; Table 4; Table 5; Table 6) are unique to humans. A notable example of a human-specific TE-mediated genomic mutation was the disruption of the *CMAH* gene, which is involved in the synthesis of a common sialic acid (Neu5Gc), by an AluY element over 2 Mya (Hayakawa et al. 2001). This may have conferred on human ancestors a survival advantage by decreasing infectious risk from microbial pathogens known to prefer Neu5Gc as a receptor.

3.9 Conclusion

A role for TEs in evolution has long been recognized by many, yet its importance has probably been underestimated. Using primates as exemplar lineages, we have assessed specific evidence, and conclude that it points strongly to an instrumental role for TEs, via TE-Thrust, in engineering the divergence of the simian lineage from other mammalian lineages. TEs, particularly Alu SINEs, have essentially acted as a huge primate-restricted stockpile of potential exons and regulatory regions, and thereby have provided the raw material for these evolutionary transitions. TEs, including Alu SINEs, L1 LINEs, ERVs and LTRs have, through active TE-Thrust, contributed directly to the primate transcriptome, and even more significantly by providing regulatory elements to alter gene expression patterns. Via passive TE-Thrust, homologous Alu and L1 elements scattered throughout the simian genome have led to both genomic gain, in the form of segmental and gene duplications, and genomic loss, by promoting unequal recombination events. Collectively, these events seem to have heavily influenced the trajectories of primate evolution and contributed to characteristic primate traits, as the simian clades especially have undergone major evolutionary advancements in cognitive ability and physiology. Although as yet incompletely documented, the evidence presented here supports the hypothesis that TE-Thrust may be a pushing force for numerous advantageous features of higher primates. These very beneficial features apparently include enhanced brain function, superior fetal nourishment, valuable trichromatic colour vision, improved

Chapter 3: Mobile DNA and the TE-Thrust Hypothesis: Supporting Evidence from the Primates

metabolism, and resistance to infectious-disease agents. Such large evolutionary benefits to various primate clades, brought about by various TE repertoires, powerfully demonstrate that if TEs are “junk” DNA then there is indeed much treasure in the junkyard, and that the TE-Thrust hypothesis could become an important part of some future paradigm shift in evolutionary theory.

Chapter 4

**The TE-Thrust Hypothesis and Plants:
Darwin's "Abominable Mystery"
and Other Puzzles**

4.1 Summary

The origin and the extremely rapid diversification of the angiosperms is one of the most extraordinary phenomena in evolutionary history. In this chapter a possible connection between this and the TE-Thrust hypothesis is explored. A causal role seems likely, as up to 80% of an angiosperm genome is comprised of TEs suggesting that TEs are potentially effective facilitators of such rapid evolution. The high frequency of hybridisation and polyploidy in angiosperms, both in the wild and in cultivation, compared to the gymnosperms and cycads is notable. The continuing evolution of resprouter angiosperms in fire prone areas, together with other data is taken to indicate that TEs can effectively facilitate somatic evolution, in addition to germ line evolution, in plants. TE activity due to polyploidy and hybridisation is posed as a major factor in the evolution of the angiosperms, which originated and diversified rapidly during the Cretaceous, and have continued to evolve up to the present. The gymnosperms, however, despite their continuing large biomass in some areas, have been in retreat, and most are apparently in stasis.

Box 4-1: Reproduction in Gymnosperms and Angiosperms

Gymnosperm Fertilisation

In gymnosperms the male gametophyte (pollen) develops a number of cells and eventually produces two sperm cells. The pollen tube penetrates the neck of the archegonium (the female sex organ) and releases the sperm cells (and some other nuclei) adjacent to the egg cell. One sperm fertilises the egg and all other nuclei disintegrate. Exceptions are found in *Ephedra* and *Gnetum* (Gnetales) in which the second sperm nucleus may fuse with another of the female gametophyte cells. However no further development occurs (Friedman and Carmichael 1996).

Angiosperm Double Fertilisation

In angiosperms the male gametophyte also produces two sperm cells in the pollen tube. This grows through the style and in the ovule enters the haploid female gametophyte cell adjacent to the egg cell. The pollen tube tip bursts releasing the two sperm cells and the tube nucleus. The sperm cells are naked, and there is only partial development of cellulose cell walls in the gametophyte allowing one sperm to fuse with the egg cell (forming the diploid zygote) and the second to unite with another adjacent cell, (the binucleate central cell) forming the triploid endosperm, which thus has one male genome and two maternal genomes. This is termed 'double fertilisation'. There are variations on the details given above, but the process is unique to angiosperms. The tube nucleus disintegrates. The endosperm nucleus usually starts divisions before the zygote does so, forming a nutrient rich tissue that nourishes the developing embryo (Berger et al. 2008). It is not known if this unusually complex reproductive system played a part in the formidable success of the angiosperms (Berger 2008).

Endosperm Balance Number (EBN) in Angiosperms

In many angiosperm taxa imprinted genes can result in failure of the endosperm to develop, and thus failure of the zygote to survive, if their EBNs are not in the approximate ratio of two maternal, to one paternal, in the developing endosperm. Variation of EBN in different species, may either inhibit or enable inter-ploidy crosses and/or interspecific hybridisation. Variation of EBN in different cytotypes within a species, can also result in intraspecific reproductive differences. EBN is thought to have been a factor in angiosperm evolution (Tate et al. 2005).

4.2 Introduction

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

Transposable elements (TEs) were first discovered in an angiosperm, namely maize, by Barbara McClintock (McClintock 1950; 1953; 1987), although her discovery was largely ignored until about 50 years later (Federoff 2000). There are many TEs in angiosperms, with up to 80% or more of the genome being made up by them. However, there are some fundamental differences between angiosperms and the mammalian metazoans (Box 4-1, and reviewed by Kejnovsky et al. 2009) from which the TE-Thrust hypothesis was largely derived, so here we investigate whether or not this hypothesis has equal relevance to the evolution of the angiosperms.

In a letter to J D Hooker on July 22 1879, Darwin described the rapid rise and early diversification within the angiosperms as an "abominable mystery" (Davies et al. 2004). Darwin's abominable mystery was mostly about his abhorrence that evolution could be both rapid and potentially even saltational because of his strongly held notion *natura non facit saltum* or, 'nature does not take a leap' (Friedman 2009).

The major angiosperm lineages originated 130-90 Mya, and they dramatically rose to ecological dominance 100-70 Mya (Davies et al. 2004) despite being comparative latecomers in the evolution of life on earth, as the Cambrian "explosion" of animal phyla began 570 Mya (Keary 1996) and the angiosperm "explosion" did not begin until about 400 million years later. Soltis et al. (2008) give the origin of the angiosperms as ~140-180 Mya, but their origin is 130 Mya according to Masterson (1994). Soltis et al. (2008) estimate that the angiosperms now have at least ~250,000 extant

species. Davies et al. (2004) agrees, but suggests that the final figure may be double this. The gymnosperms, which enjoyed dominance prior to the evolution of the angiosperms have been very largely over-run, and the angiosperms now dominate the earth.

4.3 Some Major Principles of the TE-Thrust Hypothesis

TE-Thrust can cause many genomic modifications (genmods) that cannot be caused by other "mutagens", such as exon-shuffling, retro-copies of genes, exaptation of potentially beneficial TE sequences, and many regulatory changes. Additionally, through ectopic recombination between multiple similar copies, TEs may give rise to duplications and deletions, and karyotypic changes (Oliver and Greene 2009a, 2011). Novel TEs can be acquired by germ lines by endogenous *de novo* modifications to resident TEs, *de novo* synthesis, e.g. SINEs, SVAs (Wang et al. 2005), by endogenisation of exogenous retroviruses resulting in ERVs and solo-LTRs, and perhaps rather rarely, by horizontal transposon transfer (HTT), often between completely unrelated taxa (Schaack et al. 2010). Such acquisitions of TEs by germ line genomes, can result in intermittent bursts of TE activity (Marques et al. 2005; Gerasimova et al. 1985; Kim et al. 2004; Ray et al. 2008). Various stresses experienced by an organism can also induce TE activity (Hagan et al. 2003; Li and Schmid 2001; Kimura et al. 2001). These intermittent bursts of TE activity were critical to the evolution of gene regulation during speciation in animals (Jurka 2008) and such bursts of TE activity, we additionally propose, can result in intermittent evolutionary transitions or radiations within lineages. TE-Thrust has given evidence to support the proposition

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

of punctuated equilibrium, and has offered an explanation for the great fecundity of some lineages, and for the "living fossils" in others (Oliver and Greene 2009a,b; 2011).

Although sometimes harmful to some individuals, TEs can be very beneficial to lineages. This longer term benefit results in the lineage selection of those lineages endowed with suitable consortia of TEs. Taxa or lineages of many clades that are deficient in viable (capable of activity) and active TEs, and with heterogenous populations of non-viable (incapable of activity) or inactive TEs, tend not to radiate into new species, and tend to prolonged stasis, which may eventually result in extinction, or lingering on as "fossil species". Conversely, lineages well endowed with viable and active, but suitably (incompletely) controlled TEs, tend to be fecund, or species rich, as they taxonate readily. In short, TEs, which constitute the major facilitator of evolution by TE-Thrust, can result in the generation of widely divergent new taxa, fecund lineages, lineage selection, and punctuated equilibrium (Oliver and Greene 2009a,b; 2011).

The above applies only if all other factors, such as environmental and ecological factors, are equal, and this may not always be so. There are often semi-isolated demes (Eldredge 1995), or disjunct sub-populations (Macfarlane et al. 1987) in the population of a single species. Drift of TE families can occur in these, either to fixation or extinction, and these demes may be the founders of new species (Jurka et al. 2011). A gain of TE superfamilies or families of TEs is also possible (Schaack et al. 2010) by HTT (horizontal transposon transfer). *De novo* synthesis of new

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

families, or modifications to old superfamilies or families can also occur (Wang et al. 2005; Oshima et al. 2003). These demes or disjunct populations may be prominent in plants, which are non-motile. However, in some cases their seeds may be transported over significant distances, by a wide variety of means (Howe and Smallwood 1982).

There are many other facilitators of evolution in addition to TE-Thrust which may facilitate adaptation and evolution, such as polyploidy or whole genome duplication, hybridisation, epigenetic changes, mycorrhizal fungi associations and other ecological changes which particularly impact on angiosperm evolution, either alone or in combination with each other, and/or in combination with TE-Thrust.

In the longer term, plate tectonics, climate changes, variation in CO₂ and O₂ levels, evolving or migrating pathogens and/or herbivores, and/or competitors, may affect plant radiations or extinctions.

4.3 Features of Angiosperms

Major features of angiosperms are: (1) as in other Plantae angiosperms have an alternation of generations, resulting in two distinct life phases, the haploid gametophyte i.e. the haploid embryo sac and the pollen, and the diploid sporophyte, (2) double fertilisation (Box 4-1) forming a zygote (sporophyte) and a triploid endosperm to nourish the zygote, and (3) the absence of a sequestered germ line that is continuous throughout life, as is present in some animals (Walbot and Evans 2003). Many genes,

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

then, must function in a single dose (gametophyte), and in a double dose (zygote), and in a triple dose (endosperm). This is further complicated by multiple rounds of natural polyploidy, both ancient and recent, and by recent human induced polyploidy in many crop and ornamental cultivars. Such tolerance to different ploidy levels, such as between the zygote and the endosperm, may help to explain the notable tolerance of angiosperms to polyploidy, both euploid and aneuploid. Autopolyploidy is where both parents are of the same species, and an allopolyploid is of hybrid origin. This simple distinction will suffice here, but the range of possibilities are much more complex (Tate et al. 2005). Importantly, the majority of angiosperms are thought to be of either recent polyploid, or paleopolyploid, origin, or both of these (Masterson 1994, and see 4.9 to 4.12 below).

4.4 Plant TEs

4.4.1 Viability of TEs

It should be kept in mind that TEs can only be active, or be activated by various stresses, such as polyploidy, hybridisation, tissue culture etc. if they are viable (capable of transposition). There are very little data available on the viability of TEs in plants at present, such as there are for a few mammals (Tables A4-1 and 5-2) and a few other metazoans.

4.4.2. Overview

Wicker et al. (2007) provided a summary of the TEs in plants (Table 4-1). *Copia*-like retro-TEs are ubiquitous among plants (Voytas et al. 1992). They reported them in mosses (Bryophyta), horsetails (Sphenophyta), lycopods (Lycophyta), ferns

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

(Pterophyta), cycads (Cycadophyta), *Ginko* (Ginkophyta), *Gnetum* (Gnetophyta), conifers (Coniferophyta), in the photosynthetic protist *Volvox carteri*, but not in several other species of protists. They were also found in all 38 species of angiosperms examined, including both monocots and dicots. However, *Copia*-like retro TEs were not found in the several species of insects tested, nor in fish, frog, chicken, mouse, or humans. However, contrary to this, Wicker et al. (2007) list *Copia* as being found in metazoans and fungi, as well as in plants.

Of the eleven major groups of DNA-TEs in eukaryotes, all are found in invertebrates, nine are found in vertebrates, and only six are found in plants (Table 4-1) and more detail of TEs present in some sequenced plants is shown in Table 4-2. In humans and mice DNA-TEs make up about 1-5% of the TE content, but they are uncommon in most plants which have mainly LTR retro-TEs (Bennetzen 2000). Rice is an exception as DNA-TEs make up 85% of its TE content (Feschotte and Pritham 2007).

Unlike the majority of retro-TEs, many DNA-TEs show a bias for insertion into, or close to genes. This genic proximity of DNA-TE insertions gives them a good potential for generating allelic diversity in lineages. Also DNA-TE excision, which cannot occur with retro-TEs, can help to more rapidly generate allelic diversity (Feschotte and Pritham 2007).

Table 4-1: Superfamilies of TEs in plants (from Wicker et al. 2007).

**Chapter 4: The TE-Thrust Hypothesis and Plants:
Darwin’s “Abominable Mystery” and Other Puzzles**

Class	Subclass	Order	Superfamily
I retro-TEs		LTR*	<i>Copia, Gypsy</i>
		DIRS	<i>DIRS</i>
		PLE	<i>Penelope</i>
		LINE	<i>L1, I</i>
		SINE**	<i>tRNA</i>
II DNA-TEs	1	TIR	<i>Tc1-Mariner, hAT, Mutator, P, Pif, Harbinger, CACTA</i>
	2	Helitron	<i>Helitron</i>

* No retroviruses or ERVs are found in plants, except the *Envelope-Class* Retrovirus-like LTR retro-TEs (ECR-LTR) that have been reported in angiosperms (Vicent et al. 2001; Wright and Voytas 2002). *Gypsy* is ERV-like, but lacks an *env* gene, and *Copia* is similar, except that the *pol* genes are in a different order

**Wicker et al. also list Superfamily 7SL SINEs as being present in plants, but this appears to be an error, as they are only known in primates (the *Alu*), rodents (the *B1*), and tree shrews (Norihiro Okada, personal communication).

4.4.3 TEs in a Moss: A Remnant Early Embryophyte Lineage

The draft genome sequence of the moss *Physcomitrella patens*, the first bryophyte genome to be sequenced, revealed a strong presence of TEs. About one half (48%) of the *P. patens* 511 Mb 1C genome consists of LTR retro-TEs, and almost 5,000 of these are predicted to be full length. In the LTR retro-TEs 46% are *Gypsy*-like and 2% are *Copia*-like, and 14% of the total of these are inserted into other LTR retro-TEs of their kind, with only one full length element being inserted into a gene. About 900 solo-LTRs are also present. *P. patens* contains only one family of Helitron rolling circle DNA-TE with 19 members, and high

**Chapter 4: The TE-Thrust Hypothesis and Plants:
Darwin's "Abominable Mystery" and Other Puzzles**

sequence similarity (96%) which suggests activity within the last 3 Myr. Activity of LTR retro-TEs also had an activity peak approximately 1.5 Mya, preceded by invasion events approximately 3, 4 and 5.5 Mya (Rensing et al. 2008). It is suggested that multiple Helitron families evolved in all plant lineages, but that a rapid process of DNA removal has excised all members that have not been recently active (Rensing et al. 2008), an excision process that has been found in the genomes of other plants (Vitte and Bennetzen 2006).

Table 4-2. Genome Fraction (%) of TEs in Representative Angiosperm Species (all are diploids). A. Dicotyledons B. Monocotyledons

A						
Family	Rosaceae		Vitaceae	Brassicaceae	Fabaceae	
Species	<i>Malus x domestica</i> ¹	<i>Fragaria vesca</i> ²	<i>Vitis vinifera</i> ³	<i>Arabidopsis thaliana</i> ⁴	<i>Glycine max</i> ⁵	<i>Medicago truncatula</i> ⁶
Genome Size (Mbp)	742	240	487	125	1,115	375
Chromosome No. <i>n</i>	17	7	19	5	20	8
Type I: Retro-TEs						
LTR/ <i>Gypsy</i>	25.2	6.0	14.0	5.2	25.3	5.7
LTR/ <i>Copia</i>	5.5	4.6	4.8	1.4	10.7	4.1
LTR/Other	0.4	3.8	0.0	0.0	0.0	14.4
LINE	6.5	0.2	0.6	0.9	0.2	2.2
SINE	0.0	0.1	0.0	0.0	0.0	0.1
Total	37.6	14.7	19.4	7.5	36.2	26.5
Retro-TEs						
Type II: DNA-TEs						
<i>CACTA</i>	0.0	2.6	0.2	0.9	8.7	0.1
<i>Helitron</i>	0.0	0.1	0.0	5.6	0.5	0.2
<i>hAT</i>	0.3	0.6	0.8	0.3	0.0	0.2
<i>PIF/</i>	0.0	0.2	0.0	0.2	0.3	0.5
<i>Harbinger</i>						
<i>Tc1/</i>	0.0	0.0	0.0	0.4	0.0	0.0
<i>Mariner</i>						
<i>Mutator</i>	0.0	0.2	0.4	3.1	3.9	1.5
<i>Tourist</i>	0.6*	1.6*	0.0	0.0	0.3	0.2*
<i>Stowaway</i>			0.0	0.0	0.4	
Total	0.9	5.2	1.4	11.0	14.1	3.4
DNA-TEs						
Unclassified	3.9	0.9	0.7	0.0	0.0	0.6
TOTAL TEs	42.4	20.7	21.5	18.5	50.3	30.5

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

* all MITEs

B

Family	Poaceae			
Species	<i>Zea mays</i> ⁷	<i>Sorghum bicolor</i> ⁸	<i>Oryza sativa</i> ⁹	<i>Brachypodium distachyon</i> ¹⁰
Genome Size (Mbp)	2,300	730	389	271
Chromosome No. <i>n</i>	10	10	12	5
Type I: Retro-TEs				
LTR/ <i>Gypsy</i>	46.4	19.0	12.0	16.0
LTR/ <i>Copia</i>	23.7	5.2	2.5	4.9
LTR/Other	0.0	30.2	9.0	0.5
LINE	1.0	0.0	0.8	1.9
SINE	0.0	0.0	0.4	0.0
Total	75.6	54.5	25.8	23.3
Retro-TEs				
Type II: DNA-TEs				
<i>CACTA</i>	3.2	4.7	3.4	2.2
<i>Helitron</i>	2.2	0.8	0.3	0.2
<i>hAT</i>	1.1	0.0	0.5	0.2
<i>PIF/Harbinger</i>	0.0	0.0	0.0	0.4
<i>Tc1/Mariner</i>	0.0	0.0	0.0	0.1
<i>Mutator</i>	1.0	0.1	1.8	0.6
<i>Tourist</i>	1.0	0.9	1.5	0.2
<i>Stowaway</i>	0.1	0.2	1.7**	0.9
Total	8.6	7.5	13.7	4.8
DNA-TEs				
TOTAL TEs	84.2	62.0	39.5	28.1

** Large numbers of Stowaway MITEs (~178,000) are present in some cultivars of *Oryza sativa* (Lu et al. 2012) (Table 4-3)

^{1,3,4,5} Velasco et al. (2010), ²Shulaev et al. (2011), ⁶Young et al. (2011), ⁷Schnable et al. (2009), ^{8,9}Paterson et al. (2009), ¹⁰International Brachypodium Initiative (2010).

Table 4-3 MITE Superfamilies in rice (*Oryza sativa* cultivar Nipponbare) Data from Lu et al. 2012.

Superfamily	Family Number	Total Elements	Length of all elements (bp)
<i>CACTA</i>	6	3,859	9.47×10^5
<i>hAT</i>	81	15,299	3.64×10^6
<i>PIF/Harbinger</i>	88	59,407	1.21×10^7
<i>Tc1/Mariner</i>	47	50,207	9.04×10^6
<i>Mutator</i>	115	49,126	1.11×10^7
<i>Micron</i>	1	655	2.11×10^5
Total	338	178,553	3.70×10^7

4.4.4 *Envelope-Class* Retrovirus-like TEs

Class I *Envelope-Class* Retrovirus-like LTR retro-TEs (ECR-LTRs) are widespread, transcribed and spliced, and are insertionally polymorphic in angiosperms, but were not found in the tested ferns, cycads or conifers (Vicent et al. 2001). This finding that ECR-LTRs are confined to angiosperms may not hold, as putative ECR-LTRs have been found in the gymnosperm *Pinus pinaster*. This finding could indicate either that these sequences have been in plants for 360 Myr, long before the origin of the angiosperms, or alternatively that they have been acquired independently (Miguel et al. 2008).

The ECR-LTR *SIRE1* in the soybean *Glycine max* was found to have multiplied to up to 1,000 copies within the last 70,000 years. This *SIRE1* in *Glycine max* is uniquely not truncated or peppered with the usual nonsense or frameshift mutations found in most plant LTR retro-TEs (making them non-viable, or incapable of transposition), and the *env* gene is conserved with an open reading frame (ORF). Although the presence of this conserved ORF, which can be identified across diverse plant taxa, indicates that it has been selectively maintained, it is not known if the envelope protein has any functional role within the plant (Laten et al. 2003). Pearce (2007), finding envelope-lacking SIRE-1-related sequences in pea and broad bean, suggests that SIRE-1 was formed by the acquisition of the envelope gene by a conventional soybean Ty1-*copia* retro-TE.

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

Some invertebrates are known to have genomic ERVs, with a well known example being *Gypsy* in *Drosophila* (Bowen and McDonald 2001), and there are 20 families of LTR retro-TEs in *Drosophila*, with many families apparently recently active, as they are dimorphic (Nuzhdin 1999 cited by Neafsey 2004). Although the plant cell wall is a barrier to membrane-mediated transmission of ECR-LTRs, some rhabdoviruses and bunyaviruses have genes for envelope glycoproteins that enable these viruses to shuttle between invertebrates and plants (van-Regenmortel et al. 2000). These viruses bud off from the endomembrane system and accumulate in the cells until they are ingested by an invertebrate which can then carry them to another plant (Miguel et al. 2008). All ECR-LTRs in plants make use of invertebrate vectors in which the glycosylated envelope proteins enable host cell recognition and membrane fusion. However, these proteins have been shown to be dispensable within the plants themselves, so the *env* genes possibly may have no significant function, except for the interchange of these ECR-LTRs between plants and insects. (Laten et al. 2003 and the references therein). A continuing link between plants and invertebrates, and their co-evolution, could be mediated by these ECR-LTRs¹, and if indeed they are confined to angiosperms (as found by Vicent et al. 2001, but not by Miguel et al. 2008, who also found them in the gymnosperm *Pinus*), these phenomena could make a large contribution to an explanation for the extraordinary success of the angiosperms, compared to the gymnosperms, and perhaps for the origin of the angiosperms.

4.4.5 LINEs and SINEs in Plants

¹ The co-evolution of the plants and invertebrates would likely result in the co-evolution of their viruses too.

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

The LINEs and SINEs, which are generally so abundant in mammals (reviewed by Oliver and Greene 2009a; 2011), make up a very small percentage of the maize genome and are generally rare in plants (Kidwell 2002 and Tables 4-2 A and B). However, LINEs are present throughout the plant kingdom, although in much smaller numbers than LTR retro-TEs, and SINEs have been found in several angiosperms, and may be widespread in plants (Kumar and Bennetzen 1999).

4.4.6 TEs in the Exonic Regions of Rice Genomes

The following TEs were found in the exonic regions of rice genes: Class I: Ty1-*copia*, Ty3-gypsy, LINE, *p-SINE1*, and other retro-TEs. Class II TEs: *Ac/Ds*, CACTA, *En/Spm*, and other DNA-TEs. (Sakai et al. 2007). As these TEs were found in the exonic regions, it is likely that they could have modified the structure, and/or the expression, of these genes. This suggests that they could have facilitated the evolution of rice.

4.4.7 DNA-TEs in Plants

DNA-TEs became non-viable (incapable of transposition) in most mammals around 37 Mya (Pace and Feschotte 2007), with the notable exception of the vesper bats (Ray et al. 2008, and see 5.15.1), but are common in many plants. Although the bulk of TEs in most plants belong to the LTR retro-TE *Gypsy and Copia* Superfamilies which are clustered in intergenic regions (Feschotte et al. 2002), many DNA-TE Superfamilies are also present: *Tc1-Mariner*, *hAT*, *Mutator*, *P*, *PIF-Harbinger*, CACTA, and *Helitron* are also present (Wicker et al. 2007). *Helitrons*, for example, have been found in *Arabidopsis thaliana*, *Ipomoea tricolor*, *Oriza satvia*,

the fungus *Aspergillus nidulans* and are abundant in maize (Lal and Hannah 2005; Morgante et al. 2005) where they have been mostly inserted less than 250,000 years ago (Feschotte and Pritham 2009).

The TEs within the intergenic regions of plants are usually completely different, even between related taxa such as sorghum and maize, or between wheat and barley (Poaceae). These related pairs of taxa last shared a common ancestor less than 15 Mya. More than 80% of the intact LTR retro-TEs in all analysed angiosperms can be dated as insertions that occurred within the last five Myr, but there is also a very high rate of TE sequence removal (Bennetzen 2005).

4.4.8 *Helitron* and *Helitron*-type Rolling Circle DNA-TEs

Autonomous *Helitrons* (Class II, subclass 2), like many TEs, appear to be of ancient origin (Wicker et al. 2007). *Helitrons* are reviewed by Kapitonov and Jurka (2007). A large abundance of *Helitrons* can be found in the maize genome, and these have a peculiar predisposition to restructure genes and genomes. They can be large (>10kb) because of the capture of gene fragments from multiple locations, and in maize they have transduplicated and reshuffled a very large number of sequences. Although most *Helitrons* in maize carry only one or two gene fragments, some carry exons from up to nine different genes (Feschotte & Pritham 2009). *Helitrons* can cause very large changes in the maize genome. For example, a large majority of the sequence diversities which distinguish two well known inbred lines in maize are due to them, and they have a remarkable ability to incorporate host gene

sequences. One *Helitron* is known to contain pieces of 12 different genes, and some of the gene sequences in *Helitrons* are expressed. However, it is not known if *Helitrons* can capture whole genes, as only fragments, some spanning several exons have been detected (Lal and Hannah 2005).

Most *Helitrons* in maize are non-autonomous derivatives, and although *Helitrons* have been described mainly in plants, they also are found in animals, but apparently only in the very large family of the vesper bats among the mammals. Significantly, *Helibat Helitrons* constitute at least 3.4% of the genome of the bat species *Myotis lucifigus* (Pritham & Feschotte 2007; Ray et al 2008). *Helitrons* are also found in reptiles, fish, frogs, sea urchins, sea squirts, fruit flies, mosquitoes, nematodes and rotifers, starlet sea anemones, fungi, and protists (Kapitonov and Jurka 2007).

4.4 9 Pack Mule TEs

Class II *Mutator*-like TEs (*MULES*) are especially prevalent in higher plants. In maize, rice, and *Arabidopsis*, a few *MULES* were found to carry fragments of cellular genes. These chimaeric elements can be called *Pack-MULES* (Jiang et al. 2004; Hanada et al. 2009). Although found in many eukaryotes, *Pack-MULE* TEs mediate gene evolution especially in higher plants (Jiang et al. 2004), as do *CACTA* TEs (Sinzelle et al. 2009). In rice 3,000 *Pack-MULES* have captured >1,000 gene fragments from different chromosomal loci (Sinzelle et al. 2009). Although the TE *MULE* family has captured more than 1,000 gene fragments in the rice genome, in contrast to the *Helitron* gene fragment captures, there

is a suggestion that the function of the host gene may be destroyed during acquisition by *MULEs* (Lal and Hannah 2005).

4.4.10 MITEs and *MIRNA* genes in Plants

MITEs (Miniature Inverted-repeat Transposable Elements) are a heterogeneous group of non-autonomous DNA-TEs which are flanked by TIRs, and are of only a few dozen to a few hundred base pairs long, and are frequently found in or close to genes. They have been found in the chicken, as well as in plants, and in nematodes, insects and fish. In several species a few autonomous *Tc1-Mariner* DNA-TEs cause the origin and activation of MITEs, such as the many tens of thousands of *Stowaway* MITEs in rice. *PIF-Harbinger* DNA-TEs control the activation of *Tourist* MITEs in some other plants (Wicker et al. 2007). MITEs have good potential to become microRNA, or *MIRNA* genes because of their inverted repeats and short internal sequence. MicroRNAs (miRNAs) are a class of small RNAs found in plants, animals, and a diversity of other eukaryotes, and some DNA viruses. In plants the miRNAs are 20-22 nucleotides in size and are involved in post transcriptional gene silencing, and a minority of annotated *MIRNA* genes are conserved between plant families, but most are family or species specific (Cuperus et al. 2011).

4.4.11 Plants Silence their TEs by Cytosine Methylation

The TE component of plant genomes is high, up to 50-90% in some grasses. These are generally reversibly inactivated by epigenetic mechanisms (epigenetic silencing), including cytosine methylation (Kashkush and Khasdam 2007). DNA methylation patterns have been shown to change during the life of an

organism. Additionally, in some species methylation patterns can change dramatically during the formation of polyploids (Shaked et al. 2001; Madlung et al. 2002). Wide changes in genomic methylation have been reported in synthetic *Arabidopsis* allotetraploids. These included hypermethylation and demethylation, with demethylation being most frequent. In *Triticum* both F1 hybrids and allopolyploids displayed about 7% altered methylated sites (Shaked et al. 2001). By contrast, in one study, no significant methylation alteration was found in *Gossypium* allopolyploids. However, in other studies biased expression and epigenetically induced gene silencing have been demonstrated in both natural and synthetic allotetraploid *Gossypium* species (Adams 2007). Some of these alterations appear to have arisen early after polyploid formation, and been maintained in modern allopolyploid *Gossypium* species. Seemingly, various systems of allopolyploids may respond differently to hybridisation and genome duplication, and epigenetic changes can follow both hybridisation and polyploidisation. (Soltis et al. 2003).

4.5 Angiosperm Divergence and Radiation during the Cretaceous

Originating early in the Cretaceous (146-65 Mya) angiosperms dominated the world by the end of this period (Bond and Scott 2010). These authors hypothesise that it was the angiosperm tolerance of fire that enabled this rapid dominance. In addition to resprouting dicots, in fire prone areas the high flammability of the C₄ monocot grasses can produce a feedback process that

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

enhances fire activity, resulting in the maintenance of a grassland dominated landscape, such as a savanna (Pausas and Keeley 2009). Thus fire may have been one of the significant environmental factors, stimulating angiosperm adaptation and evolution, as TE activity is generally enhanced by stress in plants (Paun et al. 2010), and fire must dramatically stress plants.

Resprouting after fire is a widespread ability in all fire-prone environments in many angiosperm lineages (Pausas and Keeley 2009). In areas of high angiosperm diversity like the Cape of South Africa and the South West of Australia resprouters are common. They comprise 49-75% of the flora in southwestern Australia, but less than 50% in the Cape region (Bell 2001). These plant species can survive intermittent (5-25 year interval) fierce fires by resprouting, in addition to some seedling recruitment. Depending on the severity of the fire, plants resprout from buds located in the leaf axils of twigs, or sunken accessory (epicormic buds) on main stems, lignotuber buds, primary axillary buds on rhizomes, or adventitious buds on lateral roots. In savanna grasslands with fires at 1-5 year intervals, all of the limited number of trees and shrubs present are resprouters (Lamont et al. 2011). Most gymnosperms lack resprouting ability and cannot survive hot fires. However some gymnosperms can resprout, for example, the "living fossils" *Ginkgo biloba* and *Wollemia nobilis* (Pausas and Keeley 2009)

4.6 Somatic Evolution, by TE-Thrust, in Angiosperms

Somatic genomic modification may be a significant factor in plant adaptation and evolution. Indeed, in plants somatic mutations and

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

karyotypic changes have long been considered to be a significant source of genomic variation, both within and between individuals (Whitham and Slobodchikoff 1981). The absence of a sequestered germ line in plants has the result that shoot apical meristems, which have already generated immense numbers of somatic cells, must switch to the production of cell lines to produce sex organs, and the sex cells within them. Such conversion of the vegetative cell supply to reproduction at numerous meristems in an "aged" plant body must also carry with it the probability that the genomes of the founder cells of the multiple germ lines will differ (Whitham and Slobodchikoff 1981), as is observed in the well known "sports" of plants. This variation in the multiple germ lines could be particularly true if the TE consortium had been repeatedly activated by the stress of repetitive intermittent fires. Thus, the possibly large genomic differences in the multiple apical meristems of post fire resprouters, suggests somatic evolution. This is because some of these seemingly highly variable meristems must produce seed and result in occasional seedling recruitment. Such seedlings would be necessarily variable, and would be, as always, subject to natural selection. This then gives a plausible solution to the evolutionary paradox posed by Lamont and Wiens (2003), namely: 'how do these very long-lived (>300-500 years) resprouter plants, which rarely reproduce by seed, manage to evolve?'

4.6.1 Somatic Evolution and Epialleles in Angiosperms

Epigenetic information, that is the formation of epialleles which include heritable signals not encoded in the primary gene

sequence, can influence the functioning of cells and their response to the environment. The origins of epialleles are mostly due to cytosine methylation, histone modification, and small RNAs. Epialleles have been stably inherited across hundreds of generations in three angiosperm allopolyploid species of *Dactylorhiza* (Orchidaceae), due to frequent epigenetic meiotic persistence, and the late determination of reproductive cell lineages in higher plants, and epialleles are characteristically created during polyploid formation (Paun 2010). Epigenetic processes are fundamentally different from genetic coding changes as they can be directly influenced by the environment, potentially allowing the inheritance of acquired character states, which gives them a neoLamarckian flavour (Richards 2006). Stably inherited epialles could have added to the rapid evolution of the angiosperms, which readily form polyploids and/or hybridise, helping to give them an advantage over the gymnosperms.

4.7 Punctuated Equilibrium, Stasis, and "Fossil Species"

Stasis is the normal condition in the fossil record and rapid change occurs rarely (Gould and Eldredge 1977; Gould 2002). Both stasis and gradual change must be accounted for in a satisfactory theory of evolution as both of these occur. The rapid change from stasis or gradualism to a punctuation event occurs only rarely, and is hypothesised to often be triggered by a burst of TE activity in metazoans (Oliver and Greene 2009a, b; 2011). However, a punctuation event can also be caused by polyploidy, and this can be common in angiosperms whether a lineage is in stasis, gradually evolving, or evolving rapidly. In such lineages a

natural allopolyploid can almost instantaneously generate a new species, sometimes recurrently, and usually sympatrically. Polyploidy is common and individual polyploid species typically form multiple times (Tate et al. 2005). Allopolyploidy often initiates a wave of TE activity. Autopolyploidy is also common, and this similarly can generate new species, but these may be cryptic, at least initially.

4.7.1 Stasis in the Gymnosperms

The gymnosperms, which first appeared about 350 Mya (Gorelick and Olson 2011), comprise only about 0.3% of extant plant species (Crepet and Niklas 2009) and most groups seem to be in stasis, but some are still very successful in terms of biomass, forming extensive stands especially in the northern hemisphere, but also in the southern hemisphere. Gymnosperms have “fossil species” like *Ginkgo biloba*, the sole extant species in its genus, which has leaves similar in form and venation to those found in rocks deposited in the Mesozoic era (248-65 Mya) when ginkgo-like plants had a worldwide distribution (Foster & Gifford 1974). However, in contrast to the stasis of most gymnosperm groups there are “dynamic evolutionary processes”, in some, especially *Pinus*. *Pinus*-specific LTR retro-TEs have been identified (Burleigh et al. 2012). As, besides the angiosperms it is only in *Pinus pinaster* in the species rich *Pinus* that *Envelope*–Class Retrovirus-like LTR retro-TEs (ECR-LTRs) have been found (Miguel et al. 2008) it is proposed that this possibly explains the continued evolution of *Pinus* as such ECR-LTRs are here proposed have been significant facilitators in angiosperm evolution. This is supported by the evidence that the ECR-LTRs *Athila*-like

(*Ty3/Gypsy* group) and *SIRE1* (*Ty1/Copia* group) are currently active in most of, or all of, the genomes they inhabit (Marco and Marin 2005).

4.8 Causes of Increased TE-activity in Plants

There is evidence that hybridisation can increase TE activity in angiosperms. In the *Ty1/copia*-like LTR retro-TE and *Ty3/gypsy*-like LTR retro-TE superfamilies in sunflowers (*Helianthus*), massive TE derepression has occurred in three diploid hybrid species and these hybrid derivatives have genomes at least 50% larger than their diploid parents, partly explainable by a proliferation of (viable) *Ty3/gypsy*-like LTR retro-TEs. The extent of proliferation of LINEs was not investigated (Michalak 2010). Hybridisation has also been found to increase TE activity (ERVs in this case) in a marsupial hybrid (O'Neil et al. 1998; 2002), but not L1 LINE activity in a rhinoceros hybrid (Dobigny et al. 2006).

In their review Parisod et al. (2010) report that an indirect impact of TE-generated rearrangements on phenotypes has also been noted, and that the TEs may be targeted by substantial epigenetic alterations that could have an impact on gene expression and genome stability. For example, unequal or ectopic recombination due to TEs (passive TE-Thrust) has resulted in the recurrent loss of the *Hardness* locus in different subgenomes of various polyploid wheat species (Chantret et al. 2005). However, TE activation may be restricted to a few specific TE families and may not occur until the fourth generation in allopolyploids, suggesting that TE activation, in this case, may require meiosis during which homeologous genomes may interact (Petit et al. 2010).

4.8.1 Tissue Culture

Somaclonal variants in tissue culture are well known examples of somatic evolution which may be facilitated by TEs (Phillips et al. 1994; Jain 2001; Ngezahayo et al. 2009; Kumar and Benetzen 1999). Somaclonal variation in tissue culture is associated with point mutations, which are usually recessive, including chlorophyll deficiency, dwarfs, and necrotic leaves in maize (Phillips et al. 1994). Other variants include chromosomal rearrangements and recombination, and DNA methylation. High ploidy and high chromosome number explants yield more variability than those those of low ploidy and low chromosome number. The altered karyotypes include chromosomal rearrangements in either euploids or aneuploids (Jain 2001). In the case of TE caused somaclonal variation, its occurrence is also influenced by the particular families of TEs present in the explant genome. For example, an endogenous MITE in rice, *mPing*, is quiescent under normal conditions, but in tissue culture callus and regenerated shoots there can be an alteration both in the cytosine methylation and *mPing* transposition in certain rice genotypes (Ngezahayo et al. 2009). The *ToS17* (Ty1-copia group) in rice transposes in tissue culture and mostly inserts in or near coding regions (Kumar and Benetzen 1999). The *Tto1* (of the Ty1-copia group) in tobacco transposes during tissue culture, and in the location of viral, wounding, and pathogen attacks. However, only these two out of twenty one listed retro-TEs were found to be activated by tissue culture by Kumar and Benetzen (1999), which suggests that only a low proportion of the somaclonal variants in tissue culture are due to TE activity. However, the other nineteen listed

retro-TEs may simply have been non viable, which would account for their lack of increased activity.

In *Anigozanthos* ("kangaroo paws") the registered cultivar 'Lemon Whizz' was a somaclonal variant of the (*A. bicolor* x *A. humilis*) x *A. Flavidus*, diploid cultivar 'Bicentennial' (Australian Plant Varieties Journal 1991 vol. 4), indicating an example of a somaclonal variant of horticultural merit, or value. ('Bicentennial' was bred by K R Oliver).

4.9 The Origins of Natural Polyploids in Angiosperms

Natural polyploids can occur in one step processes, through somatic doubling, in the zygotic, embryonic, or meristematic cells of a plant. The polyploid tissues may result in polyploid seeds and progeny. Polyploids can also result, perhaps more commonly, through the production and combination of unreduced gametes.

In angiosperms there is a high mean frequency of unreduced gametes of ~0.6%, and this can rise to ~27.5% in hybrids. Such unreduced gametes may lead to triploids or tetraploids (reviewed by Leitch and Leitch 2012). Unreduced gametes can combine with unreduced gametes in the same non-hybrid plant (resulting in autopolyploids), or unreduced gametes of a different species (giving allopolyploids). Alternatively, in a two step process, a 'triploid bridge' may be involved if triploids are produced in a diploid population (reviewed by Soltis et al. 2003). However, if there were triploids in a population, they could also produce hexaploids, either through somatic doubling, or unreduced

gametes. Additionally, the combination of unreduced gametes from a hexaploid and a tetraploid, could produce pentaploids.

Polyploids are also very tolerant of aneuploidy, so the possibilities seem almost limitless. Such events could help to account for the extraordinary range of chromosome numbers in some genera. Macfarlane et al. (1987) give an example in the monocot genera *Conostylis* and *Anigozanthos* (Haemodoraceae). *Conostylis* has 45 species, where $n = 4, 5, 6, 7, 8, 14, 16, 21, 28$, whereas in the closely related *Anigozanthos* (Figure 4-1), all 11 species are $n=6$. In their review, Lönnig and Saedler (2002) list 25 species in 12 different families which have intraspecific differences in chromosome numbers. Seven of these species include one or more instances of $2n$ equalling an odd number, indicating that they are aneuploids. An example is *Nymphaea alba* with $2n = 48, 64, 84, 105, \text{ and } 112$. Each of these variants may be a cryptic species.

4.10 The Intolerance of Polyploidy in Gymnosperms

Among the gymnosperms *Pinus* has $n=12$. Induced polyploids in *Pinus* exhibit poor survival and growth and interspecific hybridisation does not increase the genome size of *Pinus* hybrid progeny above the levels of either parent (Williams et al. 2002, cited by Morse et al. 2009). This may be because there has been no burst of TE transposition to increase the genome size, as has been observed in angiosperm hybrid *Helianthus* (Michalak 2010). Of the few LTR retro-TEs identified in *Pinus*, all are also present in other genera, but a Gypsy LTR retro-TE apparently unique to *Picea* (spruces) has been found (Williams et al. 2002, cited by

Morse et al. 2009) which could be associated with the origin of the lineage, similar to the order, family, or genus restricted SINEs found in most mammalian lineages (Borodulina and Kramerov 2005), or the *Helitrons* found in the very large genus of the vesper bats, but not in other mammals (Pritham and Feschotte 2007; Ray et al. 2008).

4.11 Hybridisation, Polyploidy, Increased TE Activity, and Speciation or Adaptation in Allopolyploid Angiosperms.

In the wild, hybridisation and polyploidy are known to be prominent processes inducing diversification and speciation in plants (Stebbins 1950; Grant 1971; Abbot 1992; Masterson 1994; Rieseberg and Wendel 2004). Salmon et al. (2005) studied two wild polyploid hybrids in the genus *Spartina*, one of which was *Spartina x townsendii* (a natural cross, estimated to have occurred 150 years ago, between the American introduced species *S. alterniflora* and the European native species *S. maritima*). The F₁ hybrids were $2n = 62$, with the allopolyploid being $2n = 124$. This produced the highly invasive salt marsh allopolyploid species *S. anglica*. As the parental species are hexaploid, *S. anglica* is a dodecaploid. About 30% of the parental methylation patterns are altered in the allopolyploid and the F₁ hybrid, suggesting that hybridisation rather than genome doubling triggered most of the methylation changes observed in *S. anglica* (Salmon et al. 2005). *S. anglica* was able to rapidly invade habitats previously unoccupiable by its parent species (Lee 2003). This is not unusual as newly formed polyploids, especially allopolyploids, frequently exhibit range expansion (Ainouche et al. 2009). In the short term, polyploid genome evolution often results in rapid and biased

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

structural changes accompanied by activation of TEs and epigenetic changes that modulate gene expression.



Figure 4-1. An *Anigozanthos humilis* X *A. flavidus* allotetraploid hybrid ($n=12$), a cultivar derived from a genus where all species are $n=6$. (Plant breeding and photography by K R Oliver).

These may have important phenotypic consequences (Comai et al. 2000) and determine the adaptive success of newly formed allopolyploid species (Wendel & Doyle 2004). TEs certainly played a central role in the shock-induced genome dynamics during allopolyploid speciation in *S. anglica*, but apparently not by means of a transposition burst (Parisod et al. 2009). This suggests that an increased level of (viable) TE activity, in some cases, following

hybridisation or other stimuli, may sometimes be gradual rather than punctuated. This may have important implications for the TE-Thrust hypothesis, but it is not known, in the above case, whether the TEs in this example were viable or non-viable.

4.12 Instantaneous Sympatric Reproductive Isolation

It is notable also that tetraploids (for example) in the wild, are reproductively isolated from their diploid progenitors, and initially at least are in a very small population (perhaps as small as one plant, at their origin). Alleles and/or TE families can drift either to fixation or extinction in small reproductively isolated populations (Chapter 5). Any rare novel TE families of *de novo*, chimaera or syntheses origin, or gained by a rare HTT (horizontal transposon transfer) event in the tetraploids, would not readily be gained by the diploid progenitors as triploids are usually sterile. The tolerance of aneuploidy by polyploid plants may also create evolutionary opportunities over time. Polyploid genomes also permit extensive gene modification by TEs, as they contain duplicate copies of all genes and are well buffered from any possible deleterious gene modifications by TEs (Comai et al. 2000; Kashkush et al. 2003; Madlung et al. 2002). Autopolyploidy is much more common than was traditionally thought (Soltis et al. 2003). In wheat, which is hexaploid, there is much more TE activity in newly synthesised strains than in established varieties, and this TE activity appears to alter the expression of adjacent genes (Kashkush et al. 2003). TEs may sometimes be a cause of possibly advantageous gene silencing. An example is the loss of glutenin expression at the Glu-1 locus in hexaploid wheat due to

an 8 Kb insertion of a retro-TE in the coding region of this gene (Wendel 2000).

Both autopolyploidy and allopolyploidy can cause sympatric quantum speciation via instantaneous reproductive isolation (Grant 1971; Stebbins 1971), with major benefits of gene duplication and TE activity, suggesting that polyploidy and TE-Thrust may have worked together throughout the evolutionary history of the angiosperms, perhaps producing major transitions, innovations, and radiations. The angiosperms, contrasted with the cycads and gymnosperms, may be indicating another benefit of TE-Thrust, namely its activity after polyploidy, either with or without, prior or concurrent hybridisation.

4.13 Adaptive Potential due to TE-Thrust: TEs and the Domestication of Plants

4.13.1 Maize

The evolution of cultivated maize (*Zea mays* L. ssp. *mays*) from its wild progenitor teosinte (*Z. mays* ssp. *parviglumis*) was a puzzle and exemplifies one of the most striking and complex examples of morphological evolution in plants (Doebley et al. 1995). The differences between the wild and cultivated sub-species are extreme and Doebley et al. (1990) indicated that key differentiating traits were each under multigenic control. However, Wang et al. (1999) identified one gene, the *teosinte branched 1* (*tb1*) gene, which encodes a transcriptional regulator involved in branch growth repression in the female inflorescence, as largely controlling some of the great differences between the two subspecies, but they also suggested that maize domestication

required hundreds of years. A new study shows that a retro-TE insertion in the regulatory sequences of the *tb1* gene, already present at low frequencies in teosinte populations, was the main target of human selection. In maize, *tb1* expression is greater than in teosinte, which correlates with repressed branch outgrowth (Tsiantis 2011) or a switch from the “bushy” phenotype of teosinte to the unbranched female inflorescence of maize. A mutation of a second gene *Tgal* (*teosinte glume architecture*) results in the loss of the glume in maize, making it possible to process the seed. These data suggest that only a very few alleles may have been the target of selection during maize domestication, as is the case for rice (the *qSH1* and *sh4* genes) and other cereal domestications (Panaud 2009). In maize the most important of these, the *tb1*, was modified due to TE-Thrust (Tsiantis 2011). This suggests that human selection of maize from teosinte may have been easy and rapid, suggesting that rapid morphological changes may sometimes occur in the wild.

Similarly to maize, the domestication of rice involved only a few genes, and in general only a few out of the 30,000 to 40,000 genes in cereal crops have been involved in cereal domestications (Panaud 2009). We regard the presence of such low copy number TE modified genes as *tb1* in maize, or in any lineage, as examples of ‘adaptive potential’ due to TE-Thrust, where ‘adaptive potential’ and ‘evolutionary potential’ are convenient name for the extremes of a continuum which could be called ‘intra-genomic potential’. The adaptive potential due to TE-Thrust is hypothesised to be realised over decades or centuries,

while the evolutionary potential can be realised over thousands or millions of years (Chapter 5).

Adaptive potential is also called 'standing variation', meaning that it is not a 'new mutation', by Tsiantis (2011) and others. Adaptive potential due to TE insertions highlights the possibility of rapid morphological changes, as in teosinte to maize due to TE insertions and selection (natural, or human) which is in agreement with the TE-Thrust hypothesis. This illustrates a rapid, rather than gradual, capacity for adaptation, or even significant evolution in some cases. Many genes, of both dicotyledonous and monocotyledonous angiosperms which have been taken as wild type (the normal dominant gene), were found to have ancient, degenerate retro-TE sequence insertions in 5' or 3' flanking regions (White et al. 1994; Wessler et al. 1995; Kumar and Bennetzen 1999) This suggests that TE modifications to gene function, or gene expression, are common in the wild.

4.13.2 Soybean

Soybean, *Glycine max*, has a complex genome and is considered to be a paleopolyploid species (Shoemaker et al. 2006). This species is estimated to have undergone two major genome duplication events, about 15 Mya and 44 Mya (Schlueter et al 2004). Differential patterns of expression have often been detected between paralogous genes in soybean which indicate that neofunctionalisation or subfunctionalisation has occurred in these genes (Schlueter et al. 2006, 2007). An example involving gene duplication and a retro-TE insertion is the presence of two paralogs of the *phyA* gene which encodes phytochrome A,

designated *GmphyA1* and *GmphyA2*. The *A2* gene was modified by a TE transposition into its exon 1, causing function different to that of the *A1* paralogue. This alteration results in photoperiod insensitivity. The affective TE insertion was a *Ty1/copia*-like retro-Te (designated *SORE-1*; SOybean RetroElement-1), which remains transcriptionally active. The result was an early maturing trait, which is adaptive for colder environments (Kanazawa 2009). This gives a plausible demonstration of adaptive evolution due to TE-Thrust, combined with polyploidy.

4.14 Cycads, Angiosperms, Polyploidy, and TE-Thrust

Cycads originated about 275-300 Mya (Axsmith et al. 2003) and cycad diversity has always been low. No polyploids are known in cycads whereas 40-95% of angiosperms are polyploids, and gymnosperms with about 750-1260 species have 5-15% polyploidy. The cycads have changed very little since their origin, and have never been very diverse. Gorelick and Olson (2011) pose the question 'is lack of cycad diversity a result of a lack of polyploidy?' They answered mainly in the affirmative, but do allow that some other factors may have been influential, especially a lack of hybridisation, very small effective population sizes, and small numbers of large chromosomes which could minimise recombination. However, theirs is an interesting question, as there are only two or three extant cycad families containing about 150 species (Rai et al. 2003), although other estimates are much higher (Gorelick and Olson 2011). In contrast to this there are about 413 families of extant angiosperms containing 250,000 species or more (Scotland & Wortley 2003; Gorelick and Olson 2011). The lack of polyploidy and hybridisation could imply a lack

of TE activity in cycads, compared to angiosperms. Such a possible lack of TE activity suggests an additional explanation for their paucity of species, past and present, which is in line with the TE-Thrust hypothesis.

4.15 Devolution and Background Extinction in Seed Plants

We previously made the point that stasis is data and must be explainable in a satisfactory evolutionary theory, as it is in the TE-Thrust hypothesis. Background extinction, that is the extinctions other than those caused by the mass extinctions, are also data, and this must be explainable also. In their review Wiens and Slaton (2012) lament the lack of attention paid to the background extinction, and also the lack of an adequate vocabulary to discuss the issue, stating that language influences cognitive processes. They suggest 'devolution', which results in population decline and background extinction, as an antonym for 'evolution' which results in adaptation and speciation. Mass extinctions account for only ~5% of all extinctions while about 99% of species that ever existed are said to be extinct (Wiens and Slaton 2012). Plants may be more resistant to mass extinctions than animals (Wing 2004), but certainly many plants have succumbed to background extinction, and many more are specifically known to be devolving, in population decline, and headed for background extinction (Wiens and Slaton 2012). Data on TEs in plants is sparse, especially in devolving lineages, but examples from metazoans suggests (A4.5,6) that devolving lineages tend to have very few extant species, and few, if any, viable TEs.

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

Mobilome consortiums are born, and can be continually reinforced by acquisitions of viable TEs, due to horizontal transposon transfer, retrovirus endogenisation, de novo synthesis, or chimaera TE formation etc. However, if the rate of attrition due to accumulation of mutations is greater than the acquisition of viable TEs, the mobilome consortium may become completely non-viable over time, as has apparently happened in the naked mole rat (Table A4-1). The lineage, clade, or species lacking a viable mobilome consortium then may, as time passes, have little adaptive potential or evolutionary potential and become "relict populations", "fossil species", or succumb to devolution and background extinction. Most of the eukaryote species (~94%) that have ever existed have succumbed to devolution and background extinction (Wiens and Slaton 2012), or as Taylor (2004) succinctly puts it 'Just as the fate of all individuals is death, so that of all species is extinction'.

4.16 TE-Thrust and Convergent Evolution

Although TE-Thrust is hypothesised to facilitate much divergent evolution, it could also be a lesser, but still effective, contributor to parallel and/or convergent evolution². Abundant sources of exogenous DNA sequences, in retroviruses, for example, can be endogenised into the genomes of related or unrelated taxa, where they can be exapted for the same, or a similar, function, perhaps especially the *env* genes of ERVs, and the regulatory sequences in their LTRs. Horizontal transposon transfer (HTT) of DNA-TEs and retro-TEs can also result in unrelated taxa becoming endowed with the same, or similar, exaptable or regulatory

² Example of convergent evolution in mammals are also known (Emera et al. 2012)

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

sequences e.g. the transposase genes of DNA-TEs, or sequences of autonomous retro-TEs. Examples of convergent evolution in plants, which may or may not be related to TE-Thrust, are the daisy like flowers in Asteraceae, *Actinodium* (Myrtaceae), and *Actinotus* (Apiaceae). There is also floral mimicry among some Western Australian terrestrial orchids that do not produce nectar, and a range of unrelated genera, enabling the orchids to attract a similar range of pollinators. Some examples are: *Thelimytra speciosa* (Orchidaceae) and *Calectasia grandiflora* (Dasypogonaceae); *Diuris* (Orchidaceae) and a number of co-blooming Papilionaceae: *Daviesia*, *Pultanea* and *Isotropis* (Brown et al. 2008).

4.17 Summary and Conclusions

Although environmental and ecological factors have played a part in angiosperm dominance, these are not included in this review, which is mainly concerned with intra-genomic factors. Multiple facilitators of genomic evolution, such as ready tolerance of polyploidy, aneuploidy, and hybridisation have been available to angiosperms, but not to gymnosperms. These are both often accompanied by very significant increases in TE activity (TE-Thrust), resulting in angiosperm lineages having increased adaptive potential and evolutionary potential, compared to gymnosperm lineages. Most examples of spontaneous heritable epialleles are also found in angiosperms, usually following the stresses of polyploidy and/or hybridisation, which could be another advantage of the angiosperms over the gymnosperms. The EBN (endosperm balance number) is present in some angiosperms, but lacking in gymnosperms which lack double fertilisation, may also

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

have been a factor promoting angiosperm reproductive isolation and speciation.

In addition angiosperms, but perhaps not gymnosperms, appear to have had retrovirus-like viruses (envelope-class retrovirus-like TEs), the most mobile of all mobile DNA, able to enter their genomes, in an interchange with insects, giving them additional sources for genomic informational change, and for evolution and radiations.

Such multiple facilitators of genomic evolution may have enabled angiosperms to occupy diverse ecological niches, develop varied life-cycles and morphologies, and an increased ability to adapt to environmental factors. Such adaptations included resprouting after periodic hot fires in some regions, and/or co-evolution with the rapidly evolving mammalian browsers, grazers, or fruit eaters. In addition, the specificity of pollinating vectors in angiosperms, rather than the restriction to mainly wind pollination in gymnosperms, seemingly would have enabled and stimulated their co-evolution with a vast array of metazoan lineages, especially among the insects and birds.

An important possibility is that the much more ancient gymnosperm lineages have succumbed much more to devolution and background extinction than the much younger angiosperm lineages.

Chapter 4: The TE-Thrust Hypothesis and Plants: Darwin's "Abominable Mystery" and Other Puzzles

Much of the data presented here supports the occurrence of somatic evolution in plants (which generally does not occur in metazoans), as another component of the TE-Thrust hypothesis.

Although the TE-Thrust hypothesis is a new hypothesis which still needs much development and testing, it powerfully portrays the profound effects that waves of transposable element activity could produce in intermittently driving evolution in angiosperms, especially after hybridisation and/or polyploidy, and also the possible passive effects of homogenous consortia of inactive TEs.

The TE-Thrust hypothesis, together with other factors mentioned above, suggests an explanation for stasis, speedy adaptation, and rapid evolutionary transitions, and/or adaptive radiation events in angiosperms, suggesting at least a part explanation for Darwin's "abominable mystery". However, like many innovative hypotheses, the TE-Thrust hypothesis needs to be subjected to further investigation. If it is fully confirmed, it will, possibly in union with other known, proposed, or as yet unknown, facilitators of evolution, offer a new conceptual foundation for much of evolutionary theory. This could put an end to the lingering dominance of gradualism, as the sole or major mode of evolution. Working out the relationship between TEs and evolution, in terms of cause and effect, seems likely to be a fruitful area of research well into the foreseeable future. However, when this is fully achieved it could give birth to a new paradigm in evolutionary theory, which incorporates recognition of the essential role that mobile DNA has played in the evolution of the great diversity of life on earth.

Appendix to Chapter 4

TE-Thrust and Evolution, Devolution, and Background Extinction in the Metazoans

A4.1 Summary

TEs can be acquired by genomes by various means such as vertical transmission, HTT (horizontal transposon transfer), endogenisation of retroviruses, *de novo* syntheses or modifications etc. However, all of these TEs are subject to attrition by deleterious mutations. If the rate of attrition exceeds the rate of acquisition, or multiplication by transposition, then the number of viable (capable of active transposition) and functional (having sufficient homology for passive ectopic recombination) categories of TEs is depleted, and can be reduced to zero. This process, if all else is equal, is proposed to be causal to prolonged stasis, devolution, and/or “living fossils” and the ubiquitous background extinction. Data from some species of rapidly adapting/evolving rodents and some rodents in stasis, and the “living fossil” coelacanth lineage, are investigated and suggest support for the TE-Thrust hypothesis.

A4.2 Introduction

Devolution, evolution, and the background extinction were introduced in 4.15, in the Chapter devoted to plants. Background extinctions, that is, the extinctions other than those caused by the

well known mass extinctions, such as those at the end of the Permian period 245 Mya and at the Cretaceous-Tertiary (K-T) boundary 65 Mya, are also data and must be explainable in any satisfactory hypothesis regarding evolution. The relationship between the TE-Thrust hypothesis and these phenomena and some of the metazoans is briefly investigated here, both in some relatively young species and in an extremely ancient lineage. An assessment is then made as to whether or not the TE-Thrust hypothesis can provide an explanation for background extinction, a usually neglected aspect of evolution. Just as punctuated equilibrium is data that must be explainable in any satisfactory hypothesis regarding evolution, so the background extinction is also data that needs to be explainable.

The eventual decay into a non-viable state by TEs, via an accumulation of deleterious mutations, is noted (Kim et al. 2011). It is then proposed that if new acquisitions of TEs are lacking then all TE-Thrust is eventually likely to cease. If this happens then a lineage or taxon appears to be vulnerable to possible relictual or “living fossil” status which may result in it eventually succumbing to background extinction. This is because if TE-Thrust ceases then intra-genomic potential (the continuum of adaptive potential through to evolutionary potential) can be much reduced, as is shown here in some species of rodents, the naked mole rat, and possibly a ground squirrel, for example. The lineage or taxon, when TE-Thrust ceases, in the absence of other facilitators of evolution, and if all else is equal, may be liable to enter a period of stasis, and may have little ability to evolve, or to adapt to changing

conditions. It is then likely to be overcome by background extinction, which is the eventual fate of the overwhelming majority of species.

The continual background extinction, in combination with intermittent mass extinctions, is possibly a useful phenomenon for the diversification of life on earth, as it helps to allow ecological niches for the evolution of new and innovative lineages, via TE-Thrust and/or other facilitators of evolution. The result can be ‘evolutionary novelties’ (Wagner and Lynch 2008; 2010), such as the mammalian placenta in the metazoans, and the flowers of the angiosperms in the seed plants.

A4.3 Acquisition and Attrition of Viable TEs

The TE populations within a genome (the mobilome consortium) are born, and can be continually reinforced by the acquisitions of viable TEs, due to horizontal transposon transfer, retrovirus endogenisation, *de novo* syntheses, *de novo* modifications, or chimaera formation etc. However, if the rate of attrition due to accumulation of mutations is greater than the acquisition of viable TEs, the TE population may become completely non-viable (incapable of activity) over time, as has happened in the naked mole rat (Table A4-1). The lineage, clade, or species lacking a viable TE population will then, over evolutionary timescales, have little adaptive potential or evolutionary potential, as these are defined below (5.2), and could be liable to become relict populations, “fossil species”, or to succumb to devolution and background extinction.

Species of multi-cellular eukaryotes in the Phanerozoic (post preCambrian time) become extinct within 10 Myr of their time of origin, and some only survive less than a million years (Taylor 2004). Most of the eukaryote species (~94%) that have ever existed have succumbed to devolution and background extinction (Wiens and Slaton 2012), or as Taylor (2004) succinctly puts it 'Just as the fate of all individuals is death, so that of all species is extinction'.

A4.4 Evolution and Devolution in the Rodentia

Although it seems to be theoretically possible that TEs could become so active, and destructive, that they could cause the devolution and background extinction of a species, there appears to be no data indicating that this has happened. There is, however, data indicating that gradualism, or stasis, and possible devolution, with background extinction, are due to all of the genomic TEs in a species, being non-viable (incapable of activity), and therefore necessarily inactive. The almost worldwide mammalian Order Rodentia consists of nearly 2,300 species, and approximately 42% of mammals are rodents (Carleton and Musser 2005). At least two-thirds of all rodents (26% of mammals) belong to one family, the Muridae (rats, mice, voles, gerbils, hamsters and lemmings). The Old World Muridae subfamily Murinae is very speciose with well over 500 species (Michaux et al 2001). In the Murinae the genera *Rattus* (rats) and *Mus* (mice) have at least 50 species each. The evolution (adaptation and radiation) of these Murinae rodents appears to support both the active and the passive modes of TE-Thrust (Table 3-1 and 5.2). The highly adaptable Old World mouse (*Mus musculus*) and rat

(*Rattus norvegicus*) have an almost worldwide distribution. These species have about 40% largely homogenous genomic TEs, with few non-viable ancient L2 LINEs and numerous and mostly very highly active L1 LINEs. They also have about 7% SINEs, with few of these being the non-viable ancient MIR SINEs, and 92% being rodent specific, viable and effective SINEs. Although all of their <1% DNA-TEs are non-viable, they have about 10% ERVs/sLTRs (Mouse Genome Sequencing Consortium 2002; Rat Genome Sequencing Project Consortium 2004), many of which are very active and are closely related to mouse exogenous retroviruses (Maksakova 2006). In contrast to this, the mouse sized sole species in its genus, the naked mole rat (*Heterocephalus glaber*), although it has a genome size comparable to the mouse, has a genome content of only 25% TEs, but the whole of the mobilome consortium (TE content) is non viable, and therefore necessarily inactive (Table A4-1).

This indicates then, according the TE-Thrust hypothesis, and if all else is equal, the naked mole rat cannot evolve by means of active TE-Thrust, but at best can only evolve gradually by passive TE-Thrust (Table 3-1). However, it could have evolved in the distant past when much of its TE population was viable, that is, before all of its TEs were degraded by mutations. As there is only one species, and all of its TE population is non-viable, it would be predicted to have little adaptive potential or evolutionary potential by the TE-Thrust hypothesis (Box 5-1). The naked mole rat may well be a candidate for devolution and background extinction, as it may not be able to adapt to environmental or ecological change. In addition, the naked mole rats TE population is well below the

Appendix to Chapter 4: TE-Thrust and Evolution, Devolution, and Background Extinction in the Metazoans

normal 40% to 50% for mammals, which could indicate that its TEs have been non-viable for a very long time, and that many TEs have been excised by deletion events.

Table A4-1. Presence and Viability of Transposable Elements (TEs) in Mouse and Naked Mole Rat (Mouse Genome Sequencing Consortium 2002; Kim et al. 2011)

	Mouse	Naked Mole Rat
Genome Size (Gbp)	2.6	2.7
TE Content (% genome)	40.9	25
LINE	Many viable (LINE 1)	Non-viable
SINE(Lineage-specific)	Some viable (e.g. B1, B2)	Non-viable
SINE (Widespread)	Non-viable	Non-viable
LTR/ERV	Many viable	Non-viable
DNA-TE	Non-viable	Non-viable

Another recently discovered rodent species *Laonastes aenigmamus* (Diatomyidae: Rodentia) is a “living fossil”, with the divergence of the *Laonastes* genus estimated to have occurred 44 Mya., Until recently *Laonastes* was thought to been extinct for 11 Myr. *Laonastes* is the sole representative of an extinct rodent family Diatomyidae, and is distantly related to the Ctenodactylidae which include several fossil taxa but only five extant species (Huchon et al. 2007). Unfortunately no sequence data of the *Laonastes aenigmamus* genome is available. ‘If the genome of this living fossil species is sequenced it could constitute a good test for the likelihood, or otherwise, that the TE-Thrust is correct. If such data were available it would certainly help to test the implied prediction of the TE-Thrust hypothesis that this “living fossil” would have no viable TEs (unless they were acquired by HTT extremely

recently) and no, or few, old TEs with sufficient homology to cause ectopic recombination.'

I hypothesise that the many similar phenomena of devolution and background extinction phenomena evident in plants (Wiens and Slaton 2012) could have causes similar to these, as seen in the naked mole rat, resulting in a loss of adaptive potential and evolutionary potential. A possible explanation for devolution and background extinction, then, becomes another aspect of the TE-Thrust hypothesis.

However, there are other possible explanations of background extinctions such as outbreaks of lethal or debilitating infectious diseases, competition from emerging or newly encountered groups of organisms, and climatic changes etc. Nevertheless, any prior loss of adaptive potential due to the hypothesised TE-Thrust by the lineage under threat of extinction could exacerbate the potential of these other factors to cause background extinctions. In addition, species with a good adaptive potential due to the hypothesised TE-Thrust, such as the rat and the mouse, would appear to be far less vulnerable to these other threats than would a species such as the naked mole rat. In summary, a loss of adaptive potential due to TE-Thrust, combined with the adverse effects of the outbreaks of infectious diseases etc. could be a potent cause of background extinction

Wiens and Slaton (2012), do not mention TEs despite the high involvement of TEs in plant evolution, and give more traditional genetic explanations for devolution and background extinction,

such as a lack of genetic variability due to perpetually self-pollinating plants, or to the gaining of short term reproductive success, as in small short-lived mammals. However, the world-wide colonisation by the very adaptable short lived rat and mouse, both of which have high reproductive success, and very viable TE consortia, suggests that a high adaptive potential due to TE-Thrust could be a better explanation.

A4.5 Possible Devolution in the Ground Squirrel

The hibernating 13-lined ground squirrel *Spermophilus tridecemlineatus* (Scuridae, Rodentia) has a large population in a comparatively localised geographical distribution of about a quarter of the USA plus Canadian landmass (Beer and Morris 2005). TEs are much less abundant (26.3%) in this squirrel than in the mouse *Mus musculus* (39.2%) and in the rat *Rattus norvegicus* (41.5%) both of which are now distributed world-wide. Also, all of the studied common TE insertions in the genome of the ground squirrel (LTR, SINE, LINE) are very much older than those of the mouse and the rat, with only <0.5% of TEs showing <4% divergence from the consensus. In sharp contrast to this the TEs in the mouse have ~7% showing <4% divergence, and those in the rat have ~4% showing <4% divergence. In addition, an assessment of very old TEs of 15% to 19% divergence shows that all three species have about 7% or 8% of TEs in this range, and that the TEs are abundant in all of the studied categories (Platt and Ray 2012). This indicates that long ago all three species could have had equal benefit from the hypothesised TE-Thrust, and could have had roughly equal realisable intra-genomic

potential, and possibly the adaptability to spread, or to speciate, more widely.

However, the L1 LINE activity in the 13 lined ground squirrel has decreased drastically within the last 26 Myr, with the last L1 LINE insertions in *S. tridecemlineatus* dating to ~5.3 Mya, and with the last SINE insertions dating to ~4 Mya, indicating that all L1 LINES are now non-viable, or at least quiescent, and that the non-autonomous SINEs which are dependent on autonomous L1 LINES for transposition, even if they are viable, are now unable to transpose (Platt and Ray 2012). It appears then that in the ground squirrel, the normal, or perhaps rather rapid, attrition of its mobilome consortium has occurred without any compensatory acquisitions by any of the available means, such as the massive ERV acquisitions by some of the Sigmodontinae rodents (5.16.2), which make up the majority of neotropical rodents and about 22% of all mammalian species in South America (Reig 1986).

The TE-Thrust hypothesis then suggests that the ground squirrel is devolving towards relict status, or “living fossil” status, and towards eventual background extinction, unless there is a new acquisition of TEs. This may be a testable prediction of the TE-Thrust hypothesis, if one had a few million years to carry out the test, but such is not the case. However this ground squirrel does demonstrate the potential for attrition of the mobilome consortium over evolutionary timescales, and the need for new acquisitions, if TE-Thrust is to be maintained.

The above is notable as the highly adaptable mouse and rat, with a worldwide range, still have many young TE families of <1% divergence from the consensus. These total 23 TE families in the mouse with a total of 1,930 TE insertions, and 21 TE families in the rat with a total 5,755 TE insertions (Jurka et al. 2011). The mouse and the rat appear to have ‘realised’ a large ‘adaptive potential’ which could possibly account for their world-wide distribution. The *Mus* and *Rattus* genera have also ‘realised’ a large ‘evolutionary potential’ as they are species rich (~50 to 60 species each) and they both belong to the family Muridae (Rodentia) which makes up about 26% of all extant mammals.

A plausible interpretation of the above data is that the mouse and the rat are still adapting and evolving well, and they certainly appear to be, while the ground squirrel is possibly devolving, and is on the way, in some distant future, to becoming a “fossil species”, as other rodents have done, and eventually succumbing to background extinction, as other rodents have also done. All of this is in accord with the TE-Thrust hypothesis, and suggests that even in the Rodentia, which has evolved and diversified to make up 42% of extant mammalian species, there is good data which indicates that devolution and background extinction are ubiquitous. Such ubiquity of background extinction is well known to palaeontologists (Taylor 2004) so this part of the TE-Thrust hypothesis, regarding the acquisition and attrition of TEs and devolution and background extinction, concurs quite well with empirical data from an independent, and somewhat unrelated scientific discipline.

A4.6 The Archaic “Living Fossil Lineages” and “Prolonged Stasis”

The fleshy-finned fish lineages, the Coelacanth (two extant species) and the Lungfish (six extant species in three genera) are said to be an early departure from the progenitor lineage of all tetrapods. They are placed in the Sarcopterygii (fleshy-finned fishes and tetrapods) along with the Actinistia (coelacanths) and the Dipnoi (lungfishes) and the enormous number of species of tetrapods (Pough et al. 2005).

The lungfishes have genome sizes that are far too large for routine genomic analysis, while the *Latimeria menadoensis* genome is smaller than the human (3.1Gbp) or mouse (2.6Gbp) genome and could be sequenced, but this has not been done yet. Coelacanths, although abundant in the fossil record, were believed to have been extinct for 63 Myr, before a living specimen was identified in 1938. It appears that the coelacanth has little propensity for whole genome duplication (WGD) or frequent tandem gene duplications and appears to be little changed. It may therefore provide access to the state of the sarcopterygian genome just prior to the emergence of the tetrapods. However, WGD and the subsequent radiation of the teleost fishes have radically altered the teleost genome relative to the common ancestor of the coelacanths and the ray-finned fishes (Noonan et al. 2004).

The Indonesian Coelacanth, *L. menadoensis*, has changed very little in appearance from fossilised coelacanths of the Cretaceous period (146 to 65 Mya) and has a genome that is evolving only

slowly. Its Hox clusters have a high level of conservation and are only evolving gradually. In addition it has been shown to be evolving slowly with regard to its turnover of SINE retro-TEs. Whereas most retro-TEs exhibit rapid expansion and turnover during evolution, at least two SINE families that predate the coelacanth-tetrapod divergence have been retained in the coelacanth, but have been exapted for new functions in both coding and non-coding regions of the tetrapods (Amemiya et al. 2010). The coelacanth is an authentic relict of the Permian-Triassic (290-208 Mya) fauna and has a wealth of paradoxical characteristics including a miniscule brain surrounded by thick adipose tissue that fills the enormous skull cavity, and lingers on in small numbers in a specialised ecological niche in the deep ocean (Grassé 1977).

As the sole vestige of a 400 Myr old lineage that has also experienced relatively low rates of molecular change, the living coelacanths can provide key insights into the complement of TEs that were present in, and made contributions to, the evolution of the ancestral tetrapod lineage (Smith et al. 2012). In a partial sequence of an *L. menadoensis* genome Smith et al. (2012) found an estimated total of ~18% miscellaneous TEs, consisting of <4% SINEs, <10% LINEs (consisting of five superfamilies), ~1% LTR/ERV, <1% DNA-TEs (consisting of five superfamilies), and 1 to 4% of *LatiHarb1*, a seemingly recently active *Harbinger* DNA-TE. This is the first known instance of a *harbinger*-superfamily DNA-TE with contemporary activity in a vertebrate genome.

With regard to the *Harbinger* DNA-TEs, each ~8.7 kb *LatiHarb1* contains two coding regions, a transposase gene and a *MYB*-like gene of as yet unknown function. These *MYB*-like genes may play roles not directly linked to transposition. The vertebrate genes *harbi1* and *naif1* and also possibly *tsnare1* may trace their ancestry to the *harbinger* superfamily (Kapitinov and Jurka 2004; Sinzelle et al. 2008).

As the *Harbinger* DNA-TE is seemingly recently active, it is presumed that all of the other TEs are non-viable, and therefore necessarily inactive, although this was not a whole genome sequence, so we lack the full details.

In the genomes of mammals numerous genes and regulatory elements have originated from various TEs (Oliver and Greene 2009a; 2011). The coelacanth retains families of SINEs (Smith et al. 2012) although they are probably non-viable. Differing families of lineage specific SINEs have acquired functionality as both regulatory and coding sequences in mammalian lineages (Oliver and Greene 2009a; 2011).

It remains to be seen whether or not the *Harbinger* DNA-TE in the coelacanth really is a fossil, or whether it may have been a more recent example of HTT, and whether or not any of the heterogeneous collection of TE superfamilies in the coelacanth contain any currently viable TEs. The estimated TE content of the genome (~18%) is low by mammalian standards, which is usually around 40% to 50% in evolving lineages. Nevertheless, it is the viability and the homogeneity of the TEs, and the categories of

TEs making up the mobilome consortium, past and present, that are important factors in their capacity to activate TE-Thrust.

It seems that, if all else is equal, lineages may survive, and indeed even speciate a little, with but little phenotypic change over enormous periods of time if they only have a low number of heterogeneous and probably non-viable TEs. Genomes devoid of viable TEs, possibly leading to lineage stasis, may not always lead to severe devolution and ultimate lineage extinction, but lead instead to small populations of extant “living fossils”.

A4.7 Conclusions

From the very limited detailed data that are available at present, it is suggested that the TE-Thrust hypothesis is able to offer an explanation for devolution and background extinction and for “living fossil” lineages. As more data become available in the future, as they surely will, and other possible hypotheses are also considered, then this preliminary finding can be more fully investigated.

Chapter 5

Transposable Elements and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

5.1 Summary

In addition to the strong divergent evolution and significant and episodic evolutionary transitions and speciation we previously attributed to TE-Thrust, we have expanded the hypothesis to more fully account for the contribution of viruses to TE-Thrust and evolution. The concept of symbiosis and holobiontic genomes is acknowledged, with particular emphasis placed on the creativity potential of the union of retroviral genomes with vertebrate genomes. Further expansions of the TE-Thrust hypothesis are proposed regarding a fuller account of horizontal transfer of TEs, the life cycle of TEs, and also, in the case of mammals, the contributions of retroviruses to the functions of the placenta. The possibility of drift by TE families within isolated demes or disjunct populations is acknowledged, and in addition we suggest the possibility of HTT into such sub-populations. ‘Adaptive potential’ and ‘evolutionary potential’ are proposed as the extremes of a continuum of ‘intra-genomic potential’ due to TE-Thrust. Specific data are given, indicating ‘adaptive potential’ being realised with regard to insecticide resistance and other insect adaptations. In this regard there is agreement between TE-Thrust and the concept of adaptation by a change in allele frequencies. Evidence

on the realisation of 'evolutionary potential' is also presented, which is compatible with the known differential survivals, and radiations, of lineages. Collectively, these data further suggest the possibility, or likelihood, of punctuated episodes of speciation events and evolutionary transitions, coinciding with, and heavily underpinned by, intermittent bursts of activity by young TE families.

5.2 Introduction

Over the past two decades, much ground-breaking work has called attention to the importance of transposable elements (TEs) in evolution (Jurka 1998; Fedoroff 1999; Kidwell and Lisch 2001; Wessler 2001; Bowen and Jordon 2002; Ogiwara et al. 2002; Deninger et al. 2003; Oshima et al. 2003; Jurka 2004; Kazazian Jr. 2004; Wessler 2006; Brandt et al. 2005; Biémont and Vieira 2006; Volff 2006; Feschotte and Pritham 2007; Muotri et al. 2007; Piskurek and Okada 2007; Beauregard et al. 2008; Böhne et al. 2008; Sinzelle et al. 2008) and many others. Building on this body of work, we have proposed TEs as powerful facilitators of evolution (Oliver and Greene 2009a). More recently, with a further development and synthesis of these initial concepts, where we had only implied a hypothesis, we explicitly proposed the 'TE-Thrust hypothesis' (Oliver and Greene 2011). The basis of the TE-Thrust hypothesis is that TEs are powerful facilitators of evolution that can act to generate genetic novelties in both an active mode and a passive mode. Active mode: by transposition, including the exaptation of TE sequences as promoters, exons, or genes. Passive mode: when present in large homogeneous populations, TEs can cause ectopic DNA recombination. Fecund lineages,

those with many species (e.g. rodents and bats, which together make up 60% of mammals) are generally rich in viable (i.e. capable of activity) and active TE families, whereas non-fecund lineages (e.g. monotremes) have mainly non-viable (i.e. incapable of activity) and inactive TEs. Evolutionary transitions, e.g. the evolution of the higher primates, and evolutionary innovations such as the mammalian placenta, also appear to be facilitated by TEs (Oliver and Greene 2011).

An outline of the TE-Thrust Hypothesis is as follows: Many eukaryote lineages are able to tolerate some sacrifices in the present, that is, a genomic “load” or population, of mostly controlled, but possibly fitness reducing TEs. Such lineages may, thereby, fortuitously, gain a continuum of ‘intra-genomic potential’ whose extremities are conveniently described as ‘adaptive potential’ and ‘evolutionary potential.’ This intra-genomic potential may be realised in the present, and/or in the descendant lineage(s) of the future. Note that this does not imply any “aim” or “purpose” to evolution, or any ability of evolution to “see” into the future.

As environmental or ecological factors change, or the lineages adopt new habitats, these intra-genomic potentials can be realised. For example, adaptive potential can be realised to give small adaptive changes within a lineage, over short periods of time, such as the evolution of insecticide resistance, when insecticides become prevalent in the environment. Evolutionary potential can be realised over much longer periods of time, perhaps in adaptive radiations, as in some rodents or bats.

All of the hypothesized capabilities of TE-Thrust shown below are consistent with the data tabulated in Oliver and Greene (2011), but are expressed here in different ways. All of them refer only to the potential for adaptation or evolution due to the hypothesised TE-Thrust. As other facilitators of evolution will possibly also be active in addition to TE-Thrust, and as environmental and ecological factors can frequently change, all of these hypothesised capabilities of TE-Thrust need to be predicated by 'if all else is equal'. These modes of TE-Thrust are extremes of continuums, so intermediate modes must occur.

Mode 1. Evolutionary potential may be realised, in concert with, or following, significant intermittent bursts of TE activity, in heterogeneous and viable TE populations, whether large or small. This can underlie what we designate as 'Type I' punctuated equilibrium (stasis with punctuation events), due to intermittent active TE-Thrust.

Mode 2. Evolutionary potential may be realised, in concert with, or following, significant bursts of TE activity, in large homogenous and viable TE populations. This can result in what we designate as 'Type II' punctuated equilibrium (gradualism with punctuation events) due to both ongoing TE-Thrust (largely passive), and to intermittent active TE-Thrust. If the TE population is small, then only intermittent active TE-Thrust is likely to occur.

Mode 3. Non-viable heterogeneous TE populations, whether large or small, may result in evolutionary stasis, due to a lack of both

active and passive TE-Thrust.

Mode 4. If a non-viable TE population is both large and homogeneous, and not too degraded by mutations, then gradualism type evolution may occur, due largely to passive TE-Thrust. If the TE population is small, then little TE-Thrust is likely to occur.

5.3 TE-Thrust and Punctuated Equilibrium

Eldredge and Gould (1972) posed the concept of punctuated equilibrium from studies of the fossil record, as opposed to the then prevailing concept of phyletic gradualism. There is now independent support for punctuated equilibrium from studies of extant taxa (Cubo 2003; Matilla and Bokma 2008; Laurin et al. 2011), from co-evolution (Togu and Sota 2009), and in extant and ancient genomes of *Gossypium* species due to intermittent TE activity (Palmer et al. 2012). TE-Thrust provides an intra-genomic explanation for punctuated equilibrium (Oliver and Greene 2009a, 2009b; 2011) as has also been suggested by Zeh et al. (2009), via epigenetic changes, and/or endogenisation of retroviruses, in response to stress, and Parris (2009), via endogenisation of retroviruses and environmental change.

The actual processes of speciation events seem to be poorly understood but new species are said to emerge from rare single events, and freed from the gradual tug of natural selection (Venditti et al. 2010). Two components appear to be necessary: Reproductive isolation and intra-genomic variation. Of these, intra-genomic variation can be readily supplied by the hypothesised

TE-Thrust (Oliver and Greene 2011), and reproductive isolation can be provided by a variety of means, such as environmental changes, behavioural changes, physical factors that can divide a population into reproductively isolated sub-populations and genetic or genomic changes (Venditti et al. 2010). Karyotypic changes associated with TE-Thrust appear to be implicated in many cases of reproductive isolation, notably in both the Old World and the New World Muridae (5.16.2). Another example of karyotypic change and speciation may be the highly speciose genus *Mus* (Rodentia; Muridae, and Murinae) and its four extant subgenera, which has had an extremely high rate of karyotypic evolution with a 10 to 30 fold increase coincident with subgeneric cladogenesis (A2.10). Twenty nine chromosome rearrangements have been fixed during the diversification of this genus (Veyrunes et al. 2006).

Much TE activity (active TE-Thrust) is thought to occur in intermittent bursts that interrupt more quiescent periods of low activity (Bénil et al, 1999; Cantrell et al. 2005; Pritham and Feschotte 2007; De Boer et al. 2007; Ray et al. 2008; Parris 2009; Zeh et al. (2009); Thomas et al. 2011; Erickson et al. 2011). These punctuation events can occur especially after intermittent acquisitions of TEs. These new acquisitions of TEs can be due to:

- Intermittent HTT, or horizontal transposon transfer (Pace II et al. 2008; Shaack et al. 2010; Gilbert et al. 2012). This appears to be rare, and probably tends to occur more often in some DNA-TEs, LTR retro-TEs and the Bov-B LINE.

Chapter 5: TEs and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

- The *de novo* synthesis of chimaeric elements, e.g. the hominid specific SVA (Wang et al. 2005). This is probably rare.
- The *de novo* syntheses of various SINEs, the younger ones (<100 Myr) of which are lineage specific (Piskurek et al. 2003; von Sternberg and Shapiro 2005). This is probably relatively rare.
- Intermittent endogenisations of retroviruses (Bénit et al. 1999; Horie et al. 2010; Belyi et al. 2010). This may be common, especially in mammals, and is especially common in some rodents (Maksakova 2006).
- Hybridisation, especially in angiosperms (Michalak 2010). This appears to be common.
- Intermittent *de novo* modifications to successive families of TEs (e.g. L1 LINEs). This is common.

An example of an intermittent burst is the L1 LINE in ancestral primates, where among a large number of overlapping families, the L1PA6, L1PA7 and L1PA8 were amplified intensively around 47 Mya. This seemingly contributed to a very large Alu SINE, and retrocopy, amplification at this time (Oshima et al. 2003). TEs can result in the acceleration of the evolution of genes in a myriad of ways, providing a means for rapid species divergences in the affected lineages (Nekrutenko and Li 2001).

5.4 An Expansion of the TE-Thrust Hypothesis

Here the TE-Thrust hypothesis is further expanded from its original formulation. However, we acknowledge that in addition to TE-Thrust, other non-genomic facilitators of evolution may play a

part in radiations and evolution, such as dynamic external factors, including geological, environmental, and ecological changes. Such factors may result in fragmentation of populations into small local demes, or larger disjunct sub-populations, which can result in reproductive isolation with possible divergence into novel taxa (Wright 1931; Eldredge 1995; Jurka et al. 2011). In addition to alleles drifting to fixation or extinction in demes, TE families likely also do so (Jurka et al. 2011) and we are in agreement with this. Additionally, in TE-Thrust we hypothesise that novel TEs (as described above,) may very occasionally be introduced into some demes or disjunct populations, but not into others, ultimately causing evolutionary transitions or the evolution of new taxa. We see the 'Carrier Sub-Population hypothesis', (Jurka et al. 2011) as being complementary to the 'TE-Thrust hypothesis', and not contradictory to it, as it is about the fixation of TEs in populations and the details of mechanisms, or origins, of speciation, which were previously not included in the TE-Thrust hypothesis. In addition, the 'Carrier Sub-Population hypothesis' gains some support from the *Gossypium* specific *Gorge* retro-TEs (Palmer et al. 2012), as *Gorge* seems to have spread to fixation in a small progenitor population of the *Gossypium* genus. Indeed, both hypotheses are in agreement in strongly relating TEs to speciation and evolution, so should not be seen as rival hypotheses. However, as we will expand on later, we suggest that karyotypic changes due to ERV and other TE presence and activity are among the factors activating the reproductive isolation necessary for speciation. Nevertheless, we agree that geographic isolation into demes etc, and niche availability, and many other phenomena, (5.9) may also be factors.

We recognise that there are many other known genomic facilitators of evolution, besides TE-Thrust. A few apposite examples are: symbiosis (Ryan 2007, 2009); hybridisation (Ryan 2006; Larson et al. 2010); non-coding RNA (Heimberg et al. 2008; Mattick 2011); horizontal gene transfer (Richards and Dacks 2006); whole genome duplications (Hoffmann et al. 2011) and viral driven evolution (Villarreal 2005, 2009; Villarreal and Witzany 2010; Ryan 2007; Feschotte and Gilbert 2012), although we have also included some of this viral driven evolution in the TE-Thrust hypothesis from the beginning. Some facilitators of evolution may have greater importance in some clades than in others. For example, whole genome duplication (polyploidy) is apparently quite important in the evolution of angiosperms (Soltis et al. 2003). Ryan (2006) includes several of the examples above under the general descriptor “genomic creativity”.

5.5 Horizontal Transfer of TEs in TE-Thrust

Mobile DNA has been classified into Class I retro-TEs and Class II DNA-TEs which also include subclass 2 DNA-TEs (Helitrons and Mavericks), as have been described and reviewed elsewhere (Brindley et al. 2003; Wicker et al. 2007; Bohne et al. 2008; Kapitonov and Jurka 2008; Goodier and Kazazian Jr 2008; Hua-Van et al. 2011). DNA-TEs have long been known to be capable of horizontal transposon transfer (HTT) e.g. the *P* element DNA-TE in *Drosophila* (Anxolabehere et al. 1988; Daniels et al. 1990); the *Mariner* DNA-TE (Maruyama and Hart 1991; Robertson and Lampe 1995; Lampe et al. 2003), and DNA-TEs in the bat *Myotis lucifugus* (Ray et al. 2007; Pritham and Feschotte 2007).

However, HTT of retro-TEs, has been less well documented, except for some examples, including the patchily distributed Bov-B LINE, (Kordiš and Gubenšek 1998; Gogolovsky et al. 2008) and the Gypsy-like retro-TEs (Herédia et al. 2004).

HTT, although probably not common, is an important aspect of the TE-Thrust hypothesis that has so far only been given cursory attention (Oliver and Greene 2009a; 2009b; 2011). A review by Schaack et al. (2010) summarises over 200 known cases of HTT, twelve of which were between different phyla. About a half of these known HTTs involved retro-TEs, most of which were LTR retro-TEs. The remaining HTTs involved a variety of DNA-TEs. HTT is an important part of the life cycle of TEs as they generally accumulate mutations and eventually become non-viable in the genomes they occupy. This can downgrade the efficacy of TE-Thrust. However, they are sometimes enabled, via chance events, to periodically make fresh starts with fully functional elements in the genomes of other lineages. At least some TEs appear to be able to endure in the absence of HTT. For example, the LINE 1 (*L1*) retro-TE in mammals has persisted for 100 Myr with no known evidence of HTT (Khan et al. 2006; Furano et al. 2004), but has now become non-viable in a few mammalian lineages (Casavant et al. 2000; Cantrell et al. 2005, 2008; Erickson et al. 2011).

Viruses and bacteria appear to be likely vectors of HTT (Dupuy et al. 2011; Schaack et al. 2010; Piskurek and Okada 2007), but other possible vectors have been proposed, such as endoparasites and intracellular parasites (Schaack et al. 2010)

and others (Silva et al. 2004). Empirical data (Anxolabehere 1988; De Boer et al. 2007; Cantrell et al. 2005; Pritham and Feschotte 2007; Ray et al. 2008) and simulations (Le Rouzic and Capy 2005) both suggest that TE amplification occurs immediately after HTT of a viable TE copy and HTT has previously been proposed as a major force driving genomic variation and biological innovation (Schaack et al. 2010).

5.6 Holobionts and Holobiontic Genomes, and the Importance of the Highly Mobile Retroviruses

Exogenous retroviruses can become endogenised, and can be united with the host genome into a holobiontic genome in a new holobiont (Box 1) i.e. the partnership, or union, of symbionts (Ryan 2006; Gilbert et al. 2010). For example, the *ERVWE1* locus in the human genome comprises a conserved *env* gene together with the conserved 5' LTR of a retrovirus that contains regulatory elements. This locus additionally includes sections of human genetic sequences and these also play a role in regulation of the *env* gene, which codes for *Syncytin-1* (Ryan 2006). Syncytin-1 has a crucial function in trophoblast cell fusion in ape placental morphogenesis (Mi et al. 2000; Ryan 2006). This strongly suggests that selection has occurred at the level of the holobiontic genome in the human plus retrovirus holobiont (Ryan 2006).

Box 5-1

Glossary of Terms

Parasite and Symbiont: To most contemporary biologists a parasite is an often harmful organism in a partnership that benefits itself at the expense of

Chapter 5: TEs and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

the other partner, and a symbiont is an organism in a mutually beneficial partnership with another organism. However, **Symbiologists** define **Symbiosis** as: 'The living together of differently named (i.e. different species) organisms, including **parasitism**, **commensalism** and **mutualism**' (Ryan 2006; 2009) and this definition is used here. **TE-Thrust**: A hypothesised pushing force generated by TEs within genomes, that can facilitate adaptation, and punctuated or major evolution, within the corresponding lineages (Oliver and Greene 2011). **Virus**: Viruses are a part of biology because they possess genes, have group identity, replicate, evolve, and are adapted to particular hosts, biotic habitats and ecological niches. Most viruses are persistent and unapparent, i.e. not pathogenic (Villarreal 2005). **Viral Biogenesis**: Exogenous retroviruses, and some other exogenous RNA viruses, can act in mutualism when endogenised in other genomes, and their genomes are united with the host genome into a 'holobiontic genome'. **Holobiont**: The partnership, or union, of symbionts (Ryan 2007; Gilbert et al. 2010). **Mobilome**: A general term for the total content of the mobile DNA in any genome. **Mobilome Consortium** (Villarreal) implies that the presence or activity of each individual or category of TE, within the Mobilome, likely affects the mobilome as a whole e.g. SINE viability is coupled to LINE compatibility and viability. **Adaptive potential**: The potential of a lineage to adapt over decades or centuries. Such adaptation can be associated with one to several genes. **Evolutionary potential**: The potential of a lineage to evolve and radiate, possibly by punctuation events, over thousands or millions of years. Such evolution may be associated with major organisational and architectural genomic changes. Note: Adaptive potential and Evolutionary potential are not distinctly different, but are useful descriptors for the extremities of an **Intra-genomic potential** continuum.

Retroviruses appear to be the most mobile of all 'mobile DNA' as they can exist exogenously as infectious, or persisting viruses, as well as by becoming endogenised in host germ lines (Ryan 2006; Hughes and Coffin 2001, 2004). Exogenous retroviruses are distinct entities to those species whose genomes into which they

endogenise to become an ERV, and they have an extracellular or virion stage, with a protein capsid. ERVs then are a part of a holobiont organism. Other TEs in a genome are not considered to be a part of a holobiont organism, as they seemingly can only transfer from genome to genome, and can have no independent existence like that of an exogenous retrovirus species.

Endogenised retroviruses (ERVs) can multiply within a genome either by repeated endogenisations, or by retrotransposition within the genome (Wang et al. 2010). Over time, due to recombinations between their LTRs and deletions, ERVs often exist mostly as solo LTRs or sLTRs, (Coffin et al. 1997; Hughes and Coffin 2004). Many Class I elements which have LTRs (Long Terminal Repeats) are related to retroviruses, e.g. the *Copia*, *Gypsy*, and *Bel-Pao* superfamilies of LTR retro-TEs.

Retroviruses are present among all placental mammals (Bénit et al. 1999), are largely restricted to vertebrates, and are particularly abundant in mammals (Villarreal 2005). Retroviruses have been endogenised in mammalian germ lines many times during the evolution of mammals and nearly half a million have reached fixation in the human germ line (Feschotte and Gilbert 2012). Such ERVs have been a very important factor in mammalian evolution (Villarreal 2005), and are particularly associated with that mammalian innovation, the placenta (Oliver and Greene 2011). Endogenised retroviruses, and the role they play in evolution, have been extensively detailed elsewhere (Villarreal 1997, 2003; 2005; 2009; Ryan 2002; 2006; 2007).

5.7 Other RNA Viruses

Endogenous non-retroviral RNA virus elements, notably Bornaviruses, have also been found in mammalian genomes, including several primates and several rodents, and these viral sequences appear to have function (Horie et al. 2010; Belyi et al. 2010). Thus, viral-eukaryote holobiont organisms appear to be not uncommon, and these could have lead to significant evolutionary innovation. This enhances the explanatory power of the TE-Thrust hypothesis. In addition, surprisingly, it appears that almost all types of viruses can become endogenised, and these are known as endogenous viral elements, or EVEs (Katzourakis and Gifford 2010; Feschotte and Gilbert 2012).

5.8 Retroviruses and the Evolution of the Mammalian Placenta

The placenta represents a major evolutionary innovation that occurred over 160 Mya at the time of the divergence of the placental mammals. The circulatory and the metabolic benefits provided by this transient organ to the growing embryo and fetus have been well investigated, but less well understood is the origin of the placenta. The invasive syncytial plate, the precursor to the placenta, and the rapidly growing trophoblast, are developmentally unique to mammals (Harris 1991). Harris proposes that prior to the divergence of placental mammals, developing embryos became infected at an early intrauterine stage, with retroviruses, which gave rise to cellular proliferation and creation of the trophoblast. This may then have resulted in the formation of the highly invasive “tumour-like” vacuolated and microvillated syncytial plate and a primitive placenta (Harris 1991).

Interestingly, the placenta is atypically globally hypomethylated, allowing many ERVs and retro-TEs to be transcriptionally active within its tissue (Feschotte and Gilbert 2012). All placental species (mammals) have endogenised significant numbers of intact ERVs, and their expression in embryonic and reproductive tissue is common, with all seven intact HERVs (human endogenous retroviruses) being expressed in the placenta, among other tissues (Villarreal 2005). Although to date there is no proof that the fusogenic ERVs of pre-mammals resulted in the evolution of the mammalian placenta (Harris 1991; Dupressoir et al. 2009), it seems likely to be correct. Supporting evidence comes from the egg-laying platypus, which has a genome that is devoid of ERVs, although there are some thousands of ancient gypsy-class LTR retro-TEs (Warren et al. 2008). By contrast all examined placental mammal genomes do contain many ERVs (Mayer and Meese 2005; Villarreal 2005), with ERV/sLTRs constituting ~8% and ~10% of the human and mouse genomes, respectively (Mouse Genome Sequencing Consortium 2002). The LTRs of ERVs contain promoters, which can confer tissue-specific expression in the placenta, e.g. the *CYP19A1*, *IL2RB*, *NOS3* and *PTN* genes are solely expressed by an LTR promoter (Cohen et al. 2009). Although there are few known unique placenta-specific genes, numerous genes expressed in the human placenta are derived from retro-TEs and ERVs (Rawn and Cross 2008), e.g. the fusogenic, ERV derived, syncytin-1 and syncytin-2, with syncytin-2 also being immunosuppressive (Kämmerer et al. 2011). The efficient adaptive immune systems of mammals must fail to initiate an immune reaction to the antigens of their embryos and placentas, and mammals are very highly infected with the

generally immunosuppressive endogenous retroviruses (Villarreal 1997). Intriguingly, retroviruses are abundant around sperm heads and also coat the female placenta (Steele 2009). The advantage of the placenta could possibly explain why extant placental mammals number well over 5,000 species, while there are less than 300 extant species of marsupials (Pough et al. 2005).

5.9 Evolvability and the TE-Thrust Hypothesis

Mutation, including gene duplication and other DNA changes, is the driving force of evolution at both the genic and the phenotypic levels (Nei 2005; 2007). Crucially, Shapiro (2010) proposes that it is mobile DNA movement, rather than replication error that is the primary engine of protein evolution. Along the same lines, Hua Van et al. (2011) stress TEs as a major factor in evolution, while Beauregard (2008) proposes that “handy junk” can evolve into “necessary junk”. Wagner (Heard et al. 2010), in support of our original concepts (Oliver and Greene 2009a) states that, in general, ‘the kinds of genetic changes that are possible...depend on what kinds of TEs are present and active’ at any particular time, in the evolution of each lineage. Thus the potential for evolutionary innovations differs over time, contradicting the concept of gradualism in lineages.’ Caporale (2009) posits that ‘selection must act on the mechanisms that generate variation, much as it does on beaks and bones’. Earl and Deem (2004), with no mention of TEs, propose the evolution of mechanisms to facilitate evolution, and describe evolvability as a selectable trait. Further to this, Woods et al. (2011) found experimental evidence, in a study of bacteria, that long term evolvability may be important

for determining the long term success of a lineage, and that less fit lineages with greater evolvability may eventually out-compete lineages with greater fitness. All of the above are in good accord with the TE-Thrust hypothesis (Oliver and Greene 2011).

5.10 Reduced “Fitness” versus Enhanced “Adaptive Potential” “Evolutionary Potential” and “Lineage Selection”

Accumulation of TEs in the genome of *Drosophila melanogaster* has been found to be associated with a decrease in fitness (Pasyukova 2004). The reduced “fitness” in *Drosophila* may be an extreme case, because in *D. melanogaster* TEs cause over 50% of *de novo* mutations (Pasyukova 2004). In contrast to *D. melanogaster*, *de novo* disease-causing insertions in humans are relatively rare (Kazazian Jr. 1999; Deininger and Batzer 1999; Chen et al. 2005; Hedges and Batzer 2005). TE activity in the laboratory mouse falls between these two extremes (Maksakova et al. 2006; Kazazian Jr 1998; Mouse Genome Consortium 2002). There is, however, no conflict with the TE-Thrust hypothesis with this finding in *Drosophila*, as despite a fitness loss in some individuals in the present, there can be a fortuitous gain in adaptive potential and evolutionary potential to the lineage as a whole. TE_d-alleles (TE- deactivated or destroyed alleles), for example, usually lower the fitness of the lineage. However, TE_m-alleles (TE-modified alleles, which can be modified in either regulation or function, or duplicated), for example, increase the genetic diversity, and hence the adaptive potential, of the lineage. These TE_m-alleles allow the lineage to adapt to

environmental/ecological challenges in the present. Also, importantly, this adaptive potential may be latent in the present, and only be realised in the future, as environmental/ecological challenges change. This latent adaptive potential and/or evolutionary potential, increases the chances of the long-term survival of the lineage. In other words, TE-Thrust can result in latent adaptive potential (also called standing variation), which can be realised, if needed, in the future, and can result in the differential survival of lineages. This is the rationale for positing lineage selection in the TE-Thrust hypothesis (Oliver and Greene 2009a, 2009b, 2011).

5.11 Realisable ‘Adaptive Potential’ Due to TE-Thrust

TE-thrust is proposed to have facilitated adaptive change as we highlighted in the simian lineage (Oliver and Greene 2011). The ongoing ability of TEs to provide realisable adaptive potential is illustrated by TE-generated polymorphic traits identified in isolated populations of laboratory-bred mice (Table 5-1) and by structural variation in the human genome still being created by L1 activity (Ewing and Kazazian, 2010)

Due to their gaining resistance to recently developed insecticides, and their colonisation of new climatic regions, insects provide a good model to study very recent and ongoing realisation of adaptive potential due to TE-Thrust in action. The history of the use of insecticides is largely known and the adaptive evolution of resistance is rapid, and has been well studied. There have been multiple recent cases clearly demonstrating a functional link between TE-Thrust and this adaptive change (Chung et al.

2007; Darboux et al. 2007; González et al. 2009; 2010; Schmidt et al. 2010). A specific example of an adaptive benefit from TE activity is the development of insecticide resistance to synthetic insecticides such as DDT in this strain, since the widespread use of these insecticides commencing in the 1940s (Schmidt et al. 2010). The use of these insecticides allowed a study of the adaptive response to a single environmental component on a timescale that enabled multiple cumulative genetic changes to be observed. This was found to occur in four steps:

- Step 1. Increased insecticide resistance in the Hikone-R strain was initially derived from an insertion of a 491 bp LTR from an *Accord* retro-TE into the regulatory region of the *Cyp6g1* gene encoding a cytochrome P450 enzyme capable of metabolising multiple insecticides, especially DDT (Dabom et al. 2002; Schmidt et al. 2010). This TE insertion, which increases insecticide resistance in this and other strains, is not found in flies collected before 1940, but is now found at high frequency (32-100%) in contemporary *D. melanogaster* populations (Schmidt et al. 2010).
- Step 2. A duplication event yielding two copies of *Cyp6g1* in the Hikone-R strain of *Drosophila*. Possibly, the *Accord* TE insertion from step 1 and the gene duplication (step 2) occurred in the one complex event, requiring only one selective sweep to explain the observed rapid increase in insecticide resistance.
- Step 3. The insertion of a *HMS Beagle* TE into the previous insertion derived from the *Accord* LTR.

- Step 4. A partial *P*-element was inserted into the previous insertion derived from the *Accord* LTR, further increasing insecticide resistance. All flies that carry a *P*-element insertion also contain the *HMS Beagle* insertion.

These four steps have occurred within seventy years in the Hikone-R strain of *Drosophila melanogaster*, and the more derived the allele, the greater the resistance (Schmidt et al. 2010). Such allelic successions, whereby different adaptive alleles are substituted sequentially, have been demonstrated in several other studies of insecticide resistance (Schmidt et al. 2010).

An example, from another suborder of insects, of the adaptive potential of TEm-alleles is the resistance to a newly encountered natural insecticide, the microbial larvicide *Bacillus sphaericus*. This has as its major active constituent a binary toxin. Resistance in a field-evolved population of the West Nile virus vector, the mosquito *Culex pipiens*, was mediated by a TE insertion into the coding sequence of the midgut toxin receptor gene (*Cpm1*)

Table 5-1. Examples of Transposable Element (TE)-Generated Polymorphic Traits Identified in Inbred Mouse Strains. All are examples of active TE-Thrust.

TE-Generated Trait	Gene Affected*	Gene Function	TE	Mouse Strain	Type of Event	Effect	Tissue Expression
Behaviour, pain sensitivity and drug response	<i>Rp2</i> ⁸	GTPase activating protein	B1	DBA	Exonization	Novel isoform	Various
	<i>Comt</i> ^{7,9,12}	Catecholamine neurotransmitter degradation	B2	Various	Exonization	Novel isoform	Brain, various
Foetal survival ?	<i>Psg23</i> ¹	Pregnancy-specific glycoprotein	LTR	Various	Exonization	Novel isoform	Placenta
	<i>Wiz</i> ²	Transcriptional regulation	LTR	C57BL/6, C57BR/cdJ	Exonization	Novel isoform	Various
Opioid sensitivity	<i>Oprm1</i> ⁶	Opioid receptor	ERV	CXBK	Exonization	Novel isoform	Nervous system
Yellow fur /high body mass	<i>Agouti</i> ^{5,10}	Pigmentation /energy metabolism	ERV	Yellow obese	Regulatory	Major promoter	Various
	<i>Vipr2</i> ¹³	Vasoactive intestinal peptide receptor	L1	BALB/c	Regulatory	Positive regulation	Various
	<i>Alas1</i> ⁴	Non-erythroid heme metabolism	B2	DBA/2	Regulatory	Negative regulation	Various
	<i>Pcdha</i> ¹⁴	Neural circuit development	ERV	Various	Regulatory	Positive/ negative regulation	CNS
	<i>Ipp</i> ³	Cytoskeleton organisation?	LTR	Various	Regulatory	Alternative promoter	Placenta
Low C4 production	<i>C4</i> ¹⁶	Complement factor	B2	Various	Gene disruption	Low expression	Liver

Table 5-1 continued

TE-Generated Trait	Gene Affected*	Gene Function	TE	Mouse Strain	Type of Event	Effect	Tissue Expression
Persistence of α -fetoprotein and H19 expression	<i>Zhx2</i> ¹¹	Transcriptional repressor	ERV	BALB/cJ	Gene disruption	Low expression	Liver, various
White coat spotting	<i>Ednrb</i> ¹⁵	Endothelin receptor	?	SSL/LeJ	Gene disruption	Low expression	Various

References ¹Ball et al., 2004; ²Baust et al., 2002; ³Chang-Yeh et al., 1993; ⁴Chernova et al., 2008; ⁵Duhl et al., 1994; ⁶Han et al., 2006; ⁷Kember et al., 2010; ⁸King et al., 1986; ⁹Li et al., 2010a,b; ¹⁰Morgan et al., 1999; ¹¹Perincheri et al., 2005; ¹²Segall et al., 2010; ¹³Steel & Lutz, 2006; ¹⁴Sugino et al., 2004; ¹⁵Yamada et al., 2006; ¹⁶Zheng et al., 1992a,b

Table 5-2 Presence and Viability of Transposable Elements (TEs) in Different Mammalian Species

	Human¹	Mouse²	Naked Mole Rat³	Platypus⁴
Genome Size (Gbp)	3.1	2.6	2.7	2.3
TE Content (% genome)	45.5	40.9	25	44.6
LINE	Some viable (LINE1)	Some viable (LINE1)	Non-viable	Some possibly viable (mainly ancient LINE2)
SINE (Lineage-specific)	Some viable (Alu, SVA)	Many viable (e.g. B1, B2)	Non-viable	Rare/absent
SINE (Widespread)	Non-viable	Non-viable	Non-viable	Some possibly viable (mainly ancient MIR/Mon-1)
LTR/ERV	Some possibly viable	Many viable	Non-viable	Rare (LTR), absent (ERV)
DNA-TE	Non-viable	Non-viable	Non-viable	Rare

References for Table 5-2: ^{1,2} Mouse Genome Sequencing Consortium 2002. ³ Kim et al. 2011. ⁴ Warren et al. 2008.

Table 5-3. Specific Examples of Transposable Elements (TEs) Implicated in Rodent-Specific Traits. All are examples of active TE-Thrust except for *Arxes1/2* which is passive.

TE-Generated Trait	Gene Affected*	Gene Function	TE	Distribution	Type of Event	Effect	Tissue Expression
	<i>Mtfull</i> ¹¹	Unknown	LTR	>Mouse	Domestication	Novel gene	Ovary
Placental morphogenesis	<i>Syncytin-A</i> ⁶	Trophoblast cell fusion	ERV	Muridae	Domestication	Novel gene	Placenta
Placental morphogenesis	<i>Syncytin-B</i> ^{6,20}	Trophoblast cell fusion/ immunosuppression	ERV	Muridae	Domestication	Novel gene	Placenta
Virus resistance	<i>Fv1</i> ⁴	Blocker of retrovirus replication	ERV	Mus	Domestication	Novel gene	
	<i>Soro-1</i> ²¹	Unknown	ERV	Rat	Domestication	Novel gene	Heart, liver
	<i>Tyms</i> ⁹	Thymidylate synthetase	L1	>Mouse	Exonisation	Major isoform	Various
	<i>Pphln1</i> ¹³	Epithelial differentiation/nervous system development	SINE/LTR	>Mouse	Exonization	Novel isoforms	Fetal, various
Soluble LIFR	<i>Lifr</i> ²²	Cytokine receptor	B2	Mouse	Exonization	Novel isoforms	Various
	<i>H2-d</i> ¹⁵	Antigen presentation to the immune system	B2	Mouse	Exonization	Novel isoform	Various
	<i>H2-I</i> ¹⁵	Antigen presentation to the immune system	B2	Mouse	Exonization	Novel isoform	Various
	<i>Phkg1</i> ¹⁹	Glycogen catabolism	B2	>Mouse	Exonization	Novel isoform	Muscle, various

Table 5-3 continued

TE-Generated Trait	Gene Affected*	Gene Function	TE	Distribution	Type of Event	Effect	Tissue Expression
	<i>Tdpoz-T1</i> ¹²	Regulation of protein processing and ubiquitination?	L1/ERV/SINE 1/hAT	>Rat	Exonization	Novel isoforms	Testis, embryo
	<i>Tdpoz-T2</i> ¹²	Regulation of protein processing and ubiquitination?	L1/ERV	>Rat	Exonization	Novel isoforms	Testis, embryo
	<i>Pmse2</i> ³³	Proteasome activator	L1	>Mouse	Regulatory	Major promoter	Various
	<i>Ocm</i> ³	Calcium binding protein and growth factor	LTR	>Rat	Regulatory	Major promoter	Macrophage
	<i>Naip</i> ²⁷	Anti-apoptosis	LTR	>Muridae	Regulatory	Major/alter native promoter	Various
	<i>Mok-2</i> ¹	Transcription factor	B2	>Mouse	Regulatory	Negative regulation	Brain, testis
	<i>Igk</i> ²⁸	Immunoglobulin light chain	B1	>Mouse	Regulatory	Negative regulation	B cell
	<i>SINE/B1 small RNAs</i> ²⁴	Embryonic postranscriptional gene silencing?	B1	>Mouse	Regulatory	Negative regulation	Embryo
	<i>Ins1</i> ¹⁷	Insulin	LINE	>Rat	Regulatory	Negative regulation	Pancreas
	<i>EpoR</i> ³²	Erythropoietin receptor	?	>Mouse	Regulatory	Negative regulation	Erythroid
	<i>Gh</i> ¹⁸	Growth hormone	B2	>Mouse	Regulatory	Insulator element	Pituitary gland

Table 5-3 continued

TE-Generated Trait	Gene Affected*	Gene Function	TE	Distribution	Type of Event	Effect	Tissue Expression
	<i>Slp</i> ³⁰	Complement activity?	ERV	>Mouse	Regulatory	Androgen responsiveness	Liver, kidney
	<i>Lama3</i> ⁷	Cell attachment, migration and organization	B2	>Mouse	Regulatory	Alternative promoter	Various
	<i>Nkg2d</i> ¹⁶	NK and T cell activating receptor	B1	>Muridae	Regulatory	Alternative promoter	NK/T cells
Cell growth control?	<i>smyc/ms-myc</i> ³¹	Unknown	?	>Muridae	Retrotransposition	Novel gene	Embryo
	<i>N-myc</i> ⁸	Unknown	?	>Sciuridae	Retrotransposition	Novel gene	Brain
	<i>Zfa</i> ²	Unknown	?	>Mouse	Retrotransposition	Novel gene	Testis
Efficient energy utilisation?	<i>Ins1</i> ²⁹	Insulin	?	Old World Rats and Mice	Retrotransposition	Novel gene	Pancreas
	<i>Pabp2</i> ¹⁴	mRNA regulation	?	>Mouse	Retrotransposition	Novel gene	Testis
	<i>Amd2</i> ²⁵	Polyamine biosynthesis	?	>Mouse	Retrotransposition	Novel gene	Liver, various
	<i>G6pd2</i> ¹⁰	Pentose phosphate pathway enzyme	?	Mouse	Retrotransposition	Novel gene	Testis
	<i>Pem2</i> ²³	Transcription factor	?	>Rat	Retrotransposition	Novel gene	Epididymis
	<i>Phgpx</i> ⁵	Antioxidant defense, spermatogenesis	?	>Mouse	Retrotransposition	Novel gene	Various

Table 5-3 continued

TE-Generated Trait	Gene Affected*	Gene Function	TE	Distribution	Type of Event	Effect	Tissue Expression
	<i>Arxes1/2</i> ²⁶	Adipogenesis	?	>Rodent	Retrotransposition	Novel gene	Adipose tissue
	<i>Mrg(s)</i> ³⁴	Nociceptive neuron function	L1	>Mouse	Duplication	Novel genes	Sensory neurons

> = Maximum known distribution.

References for Table 5-3

¹Arranz *et al.*, 1994; ²Ashworth *et al.*, 1990; ³Banville & Boie, 1989; ⁴Benit *et al.*, 1997; ⁵Boschan *et al.*, 2002; ⁶Dupressoir *et al.* 2005; ⁷Ferrigno *et al.*, 2001; ⁸Fourel *et al.*, 1992; ⁹Harendza & Johnson, 1990; ¹⁰Hendriksen *et al.*, 1997; ¹¹Holt *et al.*, 2006; ¹²Huang *et al.*, 2009; ¹³Huh *et al.*, 2006; ¹⁴Kleene *et al.*, 1998; ¹⁵Kress *et al.*, 1984; ¹⁶Lai *et al.*, 2009; ¹⁷Laimins *et al.*, 1986; ¹⁸Lunyak *et al.*, 2007; ¹⁹Maichele *et al.*, 1993; ²⁰Manganey *et al.*, 2007; ²¹Martin *et al.*, 1995; ²²Michel *et al.*, 1997; ²³Nhim *et al.*, 1997; ²⁴Ohnishi *et al.*, 2011; ²⁵Persson *et al.*, 1995; ²⁶Prokesch *et al.*, 2011; ²⁷Romanish *et al.*, 2007; ²⁸Saksela & Baltimore, 1993; ²⁹Soares *et al.*, 1985; ³⁰Stavenhagen & Robins, 1988; ³¹Sugiyama *et al.*, 1999; ³²Youssoufian & Lodish, 1993; ³³Zaiss & Kloetzel, 1999; ³⁴Zylka *et al.*, 2003.

(Darboux et al. 2007). This TE insertion induced a new mRNA splicing event, by unmasking cryptic donor and acceptor sites located in this host *Cpm1* gene. The creation of a new intron results in the expression of an altered membrane protein that cannot interact with the toxin, giving an adaptation to environmental contact with this insecticide (Darboux et al. 2007).

The migration of *Drosophila melanogaster* out of sub-Saharan Africa and its adaptation to temperate climates in North America a few centuries ago and into Australia a century ago, represents another good example of latent adaptive potential due to TEs being realised in a recent real-world context. Various TEs, modifying a diverse set of genes, have apparently played a significant role in adaptation of these flies to temperate climates on both continents (Gonzalez et al. 2010). At least eight TEM alleles, which were present in low frequencies in the African population, but showed evidence of recent positive selection for adaptation to a temperate climate, were identified. Examples are:

- A solo-LTR inserted into a conserved region of the first intron of the *sra* gene, which critically affects female ovulation and courtship;
- A LINE-like TE inserted in the intergenic region between the *Jon65Aiv* and *Jon65Aiii* genes, both of which have been associated with odour-guided behaviour (Anholt and Mackay 2001);

- A LINE-like TE inserted into a circadian regulated gene *CG34353*; (González et al. 2010).

5.12 A Partial Unification of Empirically Derived TE-Thrust Data with more Theoretically Derived Syntheses

The latent adaptive potential of the alleles of the genes above, the *sra* gene, the *Jon65Aiv* and *Jon65Aiii* genes, and the *CG34353* gene, were realised in colonisation of new areas. These TEM-alleles are adaptive for the colonisation of temperate climates by *Drosophila melanogaster* and are present in low frequencies in the original sub-Saharan African population (González et al. 2010) where they were not adaptive, but were only potentially adaptive in a changed environment or ecosystem. Their presence in sub-Saharan African populations demonstrates latent adaptive potential, or standing variation, due to TE-Thrust. The realisation of this adaptive potential by rapid positive selection of these TEM-alleles, coinciding with the expansion of the flies into temperate areas, is a change in allele frequencies, as is proposed in modern evolutionary syntheses. Thus, in this respect at least, the TE-Thrust hypothesis and the Modern Synthesis are in agreement.

5.13 The Failure of Mutation Breeding

In a review, Lönnig (2005) described how, despite early enthusiasm and sustained effort, mutation breeding (in either plants or animals) has never been successful. The mutations caused by mutagens usually produce weaker or non-functional alleles of wild type genes. In TE-Thrust, however, the TEs usually

consist of functional coding or exaptable sequences, and often also of potent regulatory sequences, so that by insertion and in many other ways, e.g. exon shuffling in the active mode and ectopic recombination in the passive mode, they can make many beneficial changes, although they may sometimes do damage (Oliver and Greene 2009a; 2009b; 2011). TEs can alter the regulation or the structure of alleles, or duplicate them (Schmidt et al. 2010; Darboux et al. 2007; González et al. 2009; 2010) creating TEm-alleles. Therefore, although attempted breeding, adaptation or evolution, by using mutagens to generate alternative alleles almost always does not work (Lönnig 2005), adaptation or evolution by means of TE-Thrust generating TEm-alleles often does work.

5.14 Reduced “Fitness” versus Enhanced “Evolutionary Potential”

The question of whether or not the possible lowering of fitness in a lineage by TEs can result in enhanced evolutionary potential may be simplified into two competing hypotheses:

- The *Null Hypothesis*: TE-Thrust is not causal to adaptation, speciation, punctuation events, or evolution.
- The *Alternative Hypothesis*: TE-Thrust is causal to adaptation, speciation, punctuation events, and evolution.

5.15 Testing the *Alternative (TE-Thrust) Hypothesis*

5.15.1 The Vesper Bats and the *Alternative (TE-Thrust) Hypothesis*

Bats are notorious reservoirs of diverse viruses, and viruses are good candidates for the HTT of TEs. This may have facilitated the recurrent waves of HTT of DNA-TEs into bats (Ray et al. 2008). The radiation of the vesper bats (Vespertilionidae) appears to support the *Alternative Hypothesis* and the active mode of TE-Thrust. The vesper bats, which have an almost worldwide distribution (Nowak 1994), are a fecund lineage (407 species of the approximately 930 species of microbats or 8-9% of all extant mammal species), and include *Myotis*, the most speciose mammalian genus with about 103 species (Singleton 2007). Significantly, vesper bats have many viable and active DNA-TEs, which have been non-viable in most other mammals for 37 Myr (Pace and Feschotte 2007).

- The early radiation of the vesper bats is proposed to have been due to the HTT of *Helitron* DNA-TEs, called *Helibat*, into the vesper bat lineage about 30-36 Myra (Pritham and Feschotte 2007).
- Amplification of DNA-TEs is thought to follow HTT in a naive lineage, which can result in innovations in the genome (Pace et al. 2008).
- *Helibat* has amplified explosively up to at least 3.4% of the *Myotis lucifugus* genome (Ray et al. 2008).

- HTT of Helitrons, especially, can lead to diversification, and to dramatic shifts in the trajectory of genome evolution (Thomas et al. 2010).
- HTT of DNA-TEs can also lead to horizontal gene transfer (Thomas et al. 2010).
- Although Helitrons have not been detected in other mammals besides the vesper bats, they are abundant in plants, invertebrates, and zebrafish, and have been implicated in large-scale gene duplication and exon shuffling.
- There were other multiple waves of HTT of DNA-TEs in the bat lineage coinciding with a period of their rapid diversification 16-25 Mya (Teeling et al. 2005; Pritham and Feschotte 2007; Ray et al. 2008).
- A further burst of New World *Myotis* diversification 12-13 Mya was noted (Stadelmann et al. 2007), corresponding well with the period in which the most active transposition of a variety of DNA-TEs is estimated to have occurred (Ray et al. 2008).
- Such repeated waves of TE activity suggest a mechanism for generating the genetic diversity needed to result in the evolution of such great species richness as is observed in the vesper bats (Ray et al. 2008).
- Active L1 LINEs (Cantrell et al. 2008) and active VES SINES (Borodulina and Kramerov 1999) have also been found in vesper bats.

This mix of viable DNA-TEs and viable retro-TEs, unknown in other mammals, could have resulted in large architectural and organisational changes in their genomes and aided in the *Myotis* diversification, enabling adaptation to very diverse ecological niches within this lineage (Thomas et al. 2011; Pritham and Feschotte 2007). This suggests that much active TE-Thrust has operated during the very large radiation of the vesper bats during the last 36 Myr. A lack of data presently obscures any conclusions regarding any possible involvement of passive TE-Thrust. The predicted evolutionary outcome of such intermittently active populations of TEs is either gradualism or stasis with punctuation events, (Type I or II punctuated equilibrium). Current data suggest that this is correct for the *Vespertilionidae*.

The above data clearly suggest support for the *Alternative* (TE-Thrust) *Hypothesis*.

5.15.2 The Muridae Rodents and the *Alternative* (TE-Thrust) *Hypothesis*

The radiation of the Old World subfamily Murinae (Muridae; Rodentia) occurred about 20 Mya (Singleton et al 2007), and there are 122 genera and 529 species in the Murinae with *Mus* and *Rattus* separating about 12 Mya (Michaux et al. 2001). This radiation appears to support the *Alternative* (TE-Thrust) *Hypothesis*, and both the active and the passive modes of TE-Thrust. The rodents are the most fecund mammalian order comprising about 40% of extant mammalian species, with an

almost worldwide distribution. The Muridae family, which includes the true mice and rats, have been particularly successful and accounts for about two-thirds of all rodent species. Representatives of the subfamily Murinae (*Mus* and *Rattus*) possess large populations of relatively homogenous genomic TEs, with numerous viable and active retro-TEs, (Table 5-2) especially ERVs.

- The Old World mouse (*Mus*) and rat (*Rattus*), with some 50-60 species each in their respective genera have about 40% largely homogenous genomic TEs, with numerous viable and mostly highly active L1 LINEs and few non-viable ancient L2 LINEs, giving ~22% total LINEs (Table 5-2). They have about 7% SINEs, with most (92%) being lineage specific, viable and effective, although slightly diverse, with few being the non-viable ancient MIR SINEs. They also have less than 1% non-viable DNA-TEs (Mouse Genome Sequencing Consortium 2002; Rat Genome Sequencing Project Consortium 2004). The mouse has about 10% ERV/sLTRs many of which are very active and are closely related to mouse exogenous retroviruses (Maksakova et al. 2006).
- The fitness cost of their greatly enhanced evolutionary potential is much higher than in humans, as previously noted (Maksakova et al. 2006).

Although the generally small size of many rodents probably aided in their diversification, there has seemingly been much active TE-Thrust, as indicated by the many documented examples of rodent-

specific traits generated by TEs (Table 5-3). They are also quite well suited to passive TE-Thrust, as they have large homogenous populations of TEs, to facilitate TE-mediated duplications, inversions, deletions or karyotypic changes.

5.15.3 The Naked Mole Rat and the *Alternative (TE-Thrust) Hypothesis*

In sharp contrast to *Mus* and *Rattus*, which are both very rich in species and have abundant viable and active TEs (Mouse Genome Sequencing Consortium 2002; Rat Genome Sequencing Project Consortium 2004), the rodent genus *Heterocephalus*, has only one species (Buffenstein and Yahav 1991). In support of the *Alternative (TE-Thrust) Hypothesis*, sequencing of *H. glaber* (Kim et al. 2011), the very atypical, physiologically unique, eusocial, and long-lived naked mole rat, has shown that it possesses a non-viable, therefore necessarily inactive, and relatively small mobilome consortium (Table A4-1).

- The TEs of the naked mole rat, although they are homogenous and constitute 25% of the genome, are highly divergent, indicating they have been both non-viable and inactive for a very long time (Kim et al. 2011).
- As most mammals have 35-50% TEs, this suggests that a substantial portion of its TEs may have been lost altogether.

The data indicate that *H. glaber* has had little or no TE-Thrust, except in the remote past, and if all else is equal, it is in stasis or gradualism. (Note: Since viable and active TEs are known to

occasionally cause genetic diseases, these data suggest that there possibly could be less genetic disease and cancer in the individuals of species such as *H. glaber*).

5.15.4 The Platypus and the *Alternative* (TE-Thrust) *Hypothesis*

Although bats and rodents may owe some of their diversity of species to their small size, the monotremes are also rather small animals, so size would not appear to be a major factor in their lack of radiation, with some three species (Pough et al. 2005), including only one extant species of platypus. While a large fraction of the platypus genome consists of TEs, the fact that these are largely ancient and inactive (Table 5-2) appears to support the *Alternative* (TE-Thrust) *Hypothesis*.

- About 50% of the platypus genome is derived from TEs, but these consist of about 1.9 million severely truncated copies of the ancient L2 LINEs, only a very few of which are putatively viable, and 2.75 million copies of the ancient SINE MIR/Mon-1, which became “extinct” (non-viable) in marsupials and eutherians 60-100 Mya (Warren et al. 2008).
- DNA-TEs and LTR retro-TEs are quite rare, but there are thousands of copies of an ancient gypsy-class LTR retro-TE (Warren et al. 2008).
- There are apparently no ERV/sLTRs (Warren et al. 2008)
- There have seemingly never been any notable infiltrations by ERVs, or HTT of DNA-TEs. This is significant given the

aforementioned importance of retroviruses to the placenta, as well as given the critical role that DNA-TEs appear to have had in generating gene regulatory networks that underlie the ability of the uterine endometrium to accommodate pregnancy via embryonic implantation (Lynch et al. 2011). The platypus seems to never have had the L1 LINEs, or Bov-B LINEs, of most mammals, and has apparently never had lineage-specific SINEs, such as the Alu of simians, or the B1 of rodents.

- Platypus evolution has been extremely conservative, especially in tooth form and body size, for 120 myr (Flannery 1994).

Although the platypus has an abundance of a restricted but homogenous range of some ancient and seemingly mostly non-viable TEs, there appears to have been very little active TE-Thrust in the platypus genome in a long time. These data clearly suggest support for the *Alternative (TE-Thrust) Hypothesis* above. According to the TE-Thrust hypothesis, the platypus should support some passive TE-Thrust due to its large, but mostly non-viable, homogeneous TE consortium. The predicted evolutionary outcome of a large homogenous population of mostly non-viable TEs, is gradualism, as in the hypothesised mode 4 of TE-Thrust. This, from current data, appears likely to be correct for the platypus.

5.16 Recent Speciation and the *Alternative (TE-Thrust) Hypothesis*

5.16.1 Young TE Families are Associated with Recent Speciation.

Mammalian species with the highest numbers of young TE families (<1% divergence from the consensus sequences) such as the mouse (23 young TE families), rat (21), bat (15), Rhesus macaque (15) and human (12) represent the largest extant mammalian orders of Rodentia, Microchiroptera, and the Primates. In sharp contrast to this, very species poor extant lineages, such as alpaca, elephant, tenrec, armadillo and platypus do not harbour any young families of TEs (Jurka et al. 2011). Nevertheless, TE-Thrust predicts none ancient speciation events being attributed to older families of TEs when they were young, and this is supported by phylogenetic analysis (Jurka 2011). These data suggest significant support for the Alternative (TE-Thrust) Hypothesis.

5.16.2 Reproductive Isolation and Speciation and the *Alternative (TE-Thrust) Hypothesis*

Reproductive isolation is generally considered to be a prerequisite for speciation, and we have no disagreement with this. Jurka et al. (2011) attributed reproductive isolation to the division of a population into demes, and also associated speciation with the availability of occupiable niches, and we agree that these can be contributing factors in speciation. However, we present data below suggesting that young families of TEs, and also of karyotypic changes due to the presence and activity of young families of TEs (especially ERVs), are also important factors in reproductive isolation and speciation.

Chapter 5: TEs and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

- The order Rodentia, which comprises 40% of extant mammalian species originated >57 Mya (Oxford Reference Online: The Encyclopedia of Mammals). The family Muridae (Rodentia) which contains an extraordinary 26% of extant mammalian species evolved only 20 Mya (Singleton et al 2007).
- Karyotypic changes between the Old World mouse and rat, representing the very speciose *Mus* and *Rattus* genera (Muridae: Subfamily Murinae) have proceeded 10 times faster than between humans and cats (Stanyon et al. 1999). Both *Mus* and *Rattus* have a large number (50-60) of species.
- The Old World mouse and rat have 23 and 21 young families of TEs (<1% divergence from the consensus sequence) with total counts of inserted TEs in these young families of 1,930 and 5,755 respectively, many of which are ERVs or related sequences (Jurka et al. 2011), indicating much recent TE/ERV activity.
- Karyotypic changes are especially likely to result from ectopic recombination of ERV insertions as ERVs are very large in size and such insertions are both abundant and polymorphic (indicating recent insertions) in mice and rats (Feschotte and Gilbert 2012).
- The very large recent radiation of some New World rodents (Muridae: Subfamily Sigmodontinae) has been coincident

with extreme karyotypic variation between species (Grahm et al. 2005) and with extraordinarily numerous ERV (MysTR) endogenisations (Cantrell et al. 2005; Erickson et al. 2011).

- The above data clearly suggest that at least some types of karyotypic changes have been involved in reproductive isolation and rampant speciation in both the Old World, and the New World Muridae. Abundant ERV insertions appear to be a likely cause of the highly derived karyotypes in both the Murinae and the Sigmodontinae rodents.
- In sharp contrast to the rodents above, the sole extant species of the platypus represents a lineage that has been extremely conservative in its evolution during its 120 Myr history, even between Australian and South American (fossil) species (Flannery 1994). The extant platypus has no young TE families with <1% divergence from the consensus sequence (Jurka et al. 2011) so has apparently had no recent TE activity, and also has no ERVs. This suggests that, together, these factors amount to a lack intra-genomic evolutionary potential in the platypus, as posited in the TE-Thrust hypothesis.

The above data appear to strongly support the *Alternative* (TE-Thrust) *Hypothesis*.

5.16.3 The Green Anole Lizard, the Tuatara, and the *Alternative* (TE-Thrust) *Hypothesis*

The *Anolis* clade of lizards comprise some 400 species which have radiated extensively in the Neotropics. Genome sequencing of *Anolis carolinensis* (Alföldi et al. 2011) has shown that its genome possesses multiple young and highly active retro-TE and DNA-TE families.

- The genome of the green anole lizard, *A. carolinensis* contains about 30% active TEs, of which about 8% is comprised of a variety of LINEs (L1,L2, CR1, RTE, and R4) which seem to be recent insertions based on their sequence similarity (Alföldi et al. 2011; Novick 2009), with about another 5.3% being SINEs.
- DNA-TEs come in at least 68 families belonging to five superfamilies, hAT, Chapaev, Maverick, Tc/Mariner and Helitron (Novick 2010).

The green anole lizard has an extremely wide diversity of active TE families, with a low rate of accumulation, similar to the TE profile of teleostean fishes (Alföldi et al. 2011). Thus, active TE-Thrust appears to be strongly implicated as a significant factor in the major radiation of this lineage of lizards. A large heterogeneous consortium of intermittently active TEs is hypothesised to result in stasis with intermittent punctuation events (type I punctuated equilibrium), as in Mode 1 of the TE-Thrust Hypothesis.

The green anole lizard contrasts sharply with the two “lizard-like” “living fossil” species of the tuatara, which have a paucity of TEs estimated to be less than 3% of its genome (Wang et al. 2006).

So far as is known, these TEs appear to be non-viable (Kapitonov and Jurka 2006). This large difference between these clades strongly implicates the abundant viable active TEs of the green anole lizard in the evolution of the high diversity of taxa in this clade, and it also suggests that an almost complete lack of TE-Thrust in the tuatara is due to its apparent paucity of young, or viable, TE families. This is consistent with the evolutionary stasis apparent in this ancient remnant clade.

The above data relating to these reptilian clades, appear to clearly support the *Alternative (TE-Thrust) Hypothesis*

5.17 Summary of the Evidence for the *Alternative (TE-Thrust) Hypothesis*

It can, of course, be argued that this evidence in mammals (bats, rodents, and the platypus), reptiles (the green anole lizard and the tuatara), and the evolution of the mammalian placenta, is all only 'circumstantial evidence', and therefore does not demonstrate a causal link between TE-Thrust and enhanced evolutionary potential. This argument is seriously weakened by the abundance of young families of TEs in the largest extant mammalian orders of rodents, bats, and primates, and their absence in the elephant, alpaca, tenrec, armadillo and platypus (Jurka et al. 2011). In addition TE-thrust predicts more ancient speciation events being attributed to older families of TEs, when they were young, and this is supported by phylogenetic evidence (Jurka 2011). These data suggest significant support for, and are quite consistent with, the *Alternative (TE-Thrust) Hypothesis*.

Chapter 5: TEs and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

The argument of 'only circumstantial evidence' is further weakened by the wide range of known beneficial genomic modifications that are due to TEs in various lineages (Capy 1998; Jurka 1998; Brosius 1999; Shapiro 1999; Miller et al. 1999; Fedoroff 1999; Benetzen J L 2000; Kidwell and Lisch 2001; Nekrutenko and Li 2001; Bowen and Jordon 2002; Holmes 2002; Deininger et al. 2003; Kazazian Jr 2004; Kim 2004; Shapiro and von Sternberg 2005; Wessler 2006; Volf 2006; Muotri et al. 2007; Pritham and Feschotte 2007; Böhne et al. 2008; Goodier and Kazazian Jr. 2008; Pace et al. 2008; Rebollo et al. 2010; Walters et al. 2009; Oliver and Greene 2009a; 2009b; 2011; Schaack et al. 2010; Shapiro 2010; Thomas et al. 2011; Grechko 2011). Therefore, it seems that a causal link between recent TE activity, oftentimes resulting in reproductive isolation, and contemporaneous speciation events is indeed likely.

Some hard evidence can be provided in regard to adaptive potential and adaptive evolution in insecticide resistance by insects in the last 70 years, and adaptation to temperate climates in the last few centuries. However, a punctuation event is estimated to take between 15,000 and 40,000 years (Gould 2002). It appears then that, as yet, bursts of TE activity and punctuation events cannot be dated accurately enough to establish any definite correlation. However, some apparent correlations have been reported, suggesting that increased TE activity may indeed be basal to, or coincident with, punctuation events and evolutionary transitions, speciation, or large radiations. Some examples of these, in addition to those detailed above, are:

Chapter 5: TEs and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

- Oshima et al. (2003) found bursts of Alu SINE and retrocopies coincident with the radiation of the higher primates 40-50 Mya.
- DNA-TE activity coincided with speciation events in salmonoid fishes (de Boer et al. 2007).
- Bursts of transposition of *BS* element transposition have also shaped the genomes of at least two species of *Drosophila*, *D. mojavensis* and *D. recta* (Granzotto et al. 2011).
- Bursts of TE activity often follow polyploidisation events (Comai 2000), or hybridisation (Michalak 2010), in angiosperms, leading to speciation.

Some suggest that a role for TEs in speciation is speculative (Hua-Van et al. 2011), while others have given data which they readily acknowledge specifically suggests TE involvement in taxon radiations (de Boer et al 2007; Pritham and Feschotte 2007; Ray et al. 2008; Thomas et al. 2011). In our interpretation of the available data, we suggest that the evidence for a causal role in speciation, and evolutionary transitions, by the hypothesised TE-Thrust (Oliver and Greene 2011) and the *Alternative Hypothesis* is quite strong, as is indicated by the abundant data above. None of the data appears to support the *Null Hypothesis*. However, we acknowledge that some speciation events may be caused by other facilitators of evolution, a few apposite examples of which have been mentioned above.

5.18 Conclusions

Unfortunately, in recent times, the field of evolutionary biology appears to have paid much more attention to the outcomes of

Chapter 5: TEs and Viruses as Factors in Adaptation and Evolution: an Expansion and Strengthening of the TE-Thrust Hypothesis

genetic mutation in terms of the generation of variants within populations than the possible spectrum of mechanisms by which mutations emerge in the first place. While small-scale DNA base changes and deletions are important in evolution, TEs and viruses are uniquely placed, via TE-Thrust, to expeditiously cause complex or large-scale changes and thereby help explain macro-evolutionary change and the emergence of highly innovative adaptations.

Much still remains to be investigated, such as the relevance of TE-Thrust to other classes and phyla. Only a small number of lineages in the mammals and to a lesser extent, the insects and reptiles, and the plants, have been considered in regard to the TE-Thrust Hypothesis to date. As more and more genomes are being sequenced, it would be interesting indeed to investigate the link between TEs, exogenous and endogenous viruses, and enhanced adaptive potential, enhanced evolutionary potential, evolutionary transitions, and the occurrence of evolutionary speciation events and transitions, in the lineages of other taxa. It seems likely that in the great diversity of extant lineages, TE-Thrust and other facilitators of evolution will have had a greater or lesser impact on adaptation and evolution. There seems to be little doubt, however, that TEs and viruses have played a major and prominent role in the evolution of almost all of life on earth, and that TEs and viruses need to be recognised and included, as the TE-Thrust Hypothesis, in a much needed extension and expansion of evolutionary theory.

Chapter 6

General Discussion

6.1. The TE-Thrust Hypothesis

The TE-Thrust hypothesis has been presented as a working hypothesis. It offers an intra-genomic explanation, in concert with natural selection, for ‘adaptive potential’ and ‘evolutionary potential’, these being convenient labels for the extremes of a continuum of ‘intra-genomic potential’. This intra-genomic potential in a lineage can be realised, either in the present or in the future. The differences, large or small, in realisable intra-genomic potential in each lineage can result in the differential survival of lineages, either in the present or in the future.

The TE-Thrust hypothesis can offer explanations for such evolutionary phenomena as fecund (adaptive) radiations, evolutionary transitions, gradualism, punctuated equilibrium, stasis, devolution, “living fossils”, and background extinctions. However, as has been stressed throughout this thesis, the hypothesised TE-Thrust does not work alone, but works together with other known (and possibly some as yet unknown) facilitators of evolution, and I have frequently taken the contributions of these into account. Some examples are point mutations, whole genome duplications, symbiosis, and hybridisation.

None of this diminishes the role of natural selection, which operates on the gamete, the zygote, the embryo, the individual,

Chapter 6: General Discussion

and together with drift, on the deme, the population, and ultimately, the lineage. However, major empirical advances concerning the nature of the genome were not anticipated or predicted by traditional evolutionary theory. The ubiquitous and ancient presence of 'Mobile DNA' and its role in evolution was completely unknown when both 'neoDarwinism' and the 'Modern Synthesis' were formulated, proposed and accepted. Many other empirical advances were also not predicted or anticipated by theory, for example, the discovery of the long introns, reverse transcriptase, and the permeability of the Weisman barrier etc. Even in the 1990s it was estimated that human genome contained about 100,000 genes (or more) but the figure has now been found to be more like 20-25,000. The new empirical data, suggest that major revisions are needed in the field of evolutionary theory. Here I have proposed and developed the TE-Thrust hypothesis, as a contributory step along this pathway to the revision of evolutionary theory.

TE-Thrust should be seen, mainly as a constructive provider of variation, which despite occasional damage to individuals within a species, can ultimately, by chance, benefit the future lineage(s) of a species, or a subset of a species. TE-caused genomic modifications serve as a storehouse of intra-genomic potential, that can be realised in the present, or in the future, as adaptation or speciation.

Traditional evolutionary theory paid little attention to symbiosis. However, since the "Weisman barrier" has been shown to be permeable, symbiosis is being recognised, by some, as being

Chapter 6: General Discussion

very important, as exemplified by Frank Ryan (2002; 2006). Endogenised retroviruses as holobionts have been especially implicated in the evolution of the mammalian placenta. The concept and reality of holobionts and holobiontic genomes is discussed, with examples of retrovirus-vertebrate holobiontic genomes, in Chapter 5. Such holobiontic genomes are potentially much more creative, than either genome could be alone.

The 'TE-Thrust hypothesis' might have been better named as the 'Mobile DNA-Thrust hypothesis', or the 'Mobilome-Thrust hypothesis', to better recognise the part played by that most mobile of all mobile DNA, the viruses, especially the retroviruses and other RNA viruses. However, as endogenised retroviruses are regarded as TEs the problem is not great. However, as other DNA or RNA viruses besides the retroviruses have a more ambiguous position, and as other DNA, such as exons or genes can be mobile, both intra-genomic, and between genomes, calling the hypothesis the Mobile DNA-Thrust hypothesis would have made it more easily expandable. Nevertheless, it was published as the TE-Thrust hypothesis (Chapter 3) and that name has been retained here.

Environmental and ecological factors are very likely to have had a major impact on the trajectories of evolution in various lineages, and these, although acknowledged, have not been covered in this study which is mainly about intra-genomic changes due to the presence and activities of mobile DNA. Such intra-genomic changes can lead to adaptive phenotypic changes and/or reproductive isolation and speciation.

The TE-Thrust hypothesis is based purely on empirical data, and not on *a priori* theoretical considerations, as have been used in much of evolutionary theory.

This thesis examines the applicability of the TE-Thrust hypothesis mainly to mammals, angiosperms and insects. There are many other phyla, classes and clades to be examined, to enable assessment of the wider relevance of the TE-Thrust hypothesis. I have, in this thesis, argued that the TE-Thrust hypothesis is a valuable addition to evolutionary theory. However, we must be led by the data, rather than forcing any particular hypothesis upon it (Rose and Oakley 2007). Nevertheless, it seems that the TE-Thrust hypothesis could well become a significant part of some future extension of, or even a new paradigm of, evolutionary theory.

6.2 An Expansion of the Posits of the TE-Thrust Hypothesis

The data presented in Chapters 2 to 5 enable an expansion of the posits given in Chapter 1.

6.2.1 Posit (1): Transposable Elements (TEs) are ubiquitous and many are extremely ancient (Chapter 2: Table 1). While some are related to prokaryote Insertion Sequences, (e.g. *Helitrons* are related to bacterial *IS91* rolling circle TEs), some (e.g. ERVs and solo LTR retro-TEs) are related to retroviruses,

(Wicker et al. 2007). Others are synthesised *de novo* within a genome (e.g. the non-autonomous retro-TEs, the SINEs, and the non-autonomous DNA-TEs, the MITEs) or are chimaeras (e.g. the non-autonomous retro-TEs, the SVAs). Many TEs are known to have transferred from genome to genome by horizontal transposon transfer (HTT), often between completely unrelated lineages (Chapter 6). TEs are not merely “junk”, or “parasitic DNA”, as has been thought by some. Although occasionally harmful to individuals, TEs can be very beneficial to lineages, and are potentially powerful facilitators of evolution (Chapter 2.3 and 2.4).

6.2.2 Posit (2): TEs can cause genetic changes of great magnitude and variety within genomes, making genomes flexible and dynamic, so that they drive genomic evolution and the evolution of their lineage phenotypes. Because of TEs, genomes are not static, or nearly so, but tend to be dynamic and to change, sometimes rapidly. Causes of possible rapid genomic changes include, stress, whole genome duplications, hybridisation, HTT, or retrovirus endogenisations.

6.2.3 Posit (3): TEs can cause many genomic alterations that cannot be caused by any other “mutagens”. An example is exon shuffling by the autonomous retro-TEs, the *L1* LINEs (Table 2-2). *L1* LINEs have bi-directional promoters (sense and anti-sense) in their 5' UTR and can influence the expression of upstream genes. *L1* LINEs can also be the cause of exon shuffling, as they can carry with them

sequences from their 3' adjoining DNA. The *L1* LINEs in many mammals also mobilise the non-autonomous SINEs, the “younger” ones (< 100 Myrs) of which can be extremely important in facilitating evolution, e.g. the *Alu* SINE in primates. *L1* LINEs additionally create retrocopy genes by means of the reverse transcription of mRNA. These retrocopy genes, often in association with point mutations or other mutations, can sometimes become viable neogenes, such as *glud2*, for example. The DNA-TEs, *Helitrons* and *pack-mules* can carry gene fragments or exons to new locations, where they have the potential to modify gene function, or even to synthesise new genes. DNA-TE *transposase* genes can be exapted for new functions, as can the *env* genes of ERVs. For example ERV *env* genes have been exapted for the formation of the placental *Syncytin 1* and *Syncytin 2* genes in humans.

6.2.4 Posit (4): TEs can greatly modify gene regulation and gene regulation networks, and they can express genes in cell lines in which they were not previously expressed. For example, the LTRs of ERVs, or LTR retro-TEs, contain promoters which can influence the expression of nearby genes, as can the promoters of *L1* LINEs and *Alu* SINEs. An example is the switching of the expression of certain α -amylase genes from the pancreas to the salivary glands by an ERV acting as a tissue specific promoter in some Old World primates, including humans.

6.2.5 Posit (5): TE-Thrust can build, sculpt, and reformat genomes by both active and passive means. Active TE-Thrust is due to the active transposition of TEs, from either a heterogenous or a homogenous population of TEs. Passive TE-Thrust is due to ectopic recombination between homologous TE insertions, which can result in insertions, deletions, inversions, or translocations. Such passive TE-Thrust facilitated ectopic recombination is common only when there are large homogeneous populations of TEs. As an extension of this I have hypothesised four modes of TE-Thrust, in which the modes are extremes of parallel continuums. These modes offer explanations for stasis, gradualism, and punctuated equilibrium, and I have presented data which suggests that these hypothesised modes are likely to be correct (Chapters 3 and 5). The absence, or rarity, of viable TEs in some lineages may also explain, or help to explain, the devolution and relict or “living fossil” status, or background extinction of these lineages (Chapter 4, and the Appendix to Chapter 4). The origin of the adaptive immune system, the V(D)J recombination mechanism of jawed vertebrates is attributed to DNA-TEs, and *RAG1*, which gives the catalytic core for this reaction. (Feschotte and Pritham 2007), is very similar in sequence to DNA-TE *Transib* transposases which have been identified in the genomes of diverse invertebrates (Kapitonov and Jurka 2005).

6.2.6 Posit (6): TE-Thrust, via intermittent bursts of TE activity, sometimes results in macro-evolutionary punctuation events in lineages previously in stasis or gradualism. These often

result in a drive towards novelty, diversity, or complexity and thus a radiation of species. This is punctuated equilibrium, and is consistent with the greater part of the fossil record (Chapter 2), and is also consistent with Mode 1 and Mode 2 of the hypothesized major modes of TE-Thrust (Chapter 3: Table 1). Relatively recent (in the evolutionary timescale) punctuation events, coincident with, and seemingly caused by bursts of TE activity, have been recorded in (1) The Old World Murinae rodents, and (2) the South American Sigmodontinae rodents, coincident with a very large ERV invasion together with extreme karyotypic variation, and (3) the vesper bats, and (4) the higher primates, and (5) the salmonoid fish (Chapter 5). However, these cannot, as yet, be dated accurately enough to definitely establish a correlation, and there is always the possibility of other causal factors. Nevertheless, they are evidence that suggests cause and effect, but further research is needed to clarify the data in this area.

6.2.7 (Posit 7) An absence, or a paucity of viable TEs in a lineage results in stasis or in gradualism. Data from the naked mole rat (Chapters 5) and the platypus (Chapter 5), both of which are the only species in their respective genera suggest that this is correct. The naked mole rat has a smaller more heterogeneous mix of non- viable TEs, and fits well into Mode 3 (stasis) of the TE-Thrust hypothesis. The platypus has a larger population of a more homogeneous mix of TEs, with probably few viable TEs and fits well into Mode 4 (gradualism) of the TE-Thrust hypothesis (Chapter 3: Table

Chapter 6: General Discussion

1). Platypus evolution, and monotreme evolution generally, has been extremely conservative throughout their greater than 120 Myr history (Flannery 1994). These data further suggest that the hypothesised Modes of the TE-Thrust hypothesis may be correct.

6.2.8 Posit (8): Successful lineages do not destroy TEs, but have strong genomic control of transposition of TEs in the soma, where they are potentially damaging. However, there is less control of TE activity in the germ line and the early embryo in mammals, where their activity can generate potentially useful, neutral, and deleterious variations in the progeny. Useful variants then increase and deleterious variants decrease or are eliminated in future generations, by means of natural selection. Neutral, or only slightly deleterious allele or genome modifications, may be conserved by drift, and are a source of adaptive potential, and/or evolutionary potential, in the cases of environmental or ecological changes, or of lineage range expansion (Chapters 4 and 5).

6.2.9 Posit (9): Although sometimes harmful to some individuals, TEs can be very beneficial to lineages. Those lineages endowed with a suitable consortium of TEs are likely to survive and to radiate or proliferate, as such lineages have enhanced adaptive potential and enhanced evolutionary potential. Those lineages that lack a consortium of viable and active TEs, however, are liable to stasis and devolution in the long run, leading to “fossil species” and possible eventual extinction (Chapters 4 and 5).

A good example of the benefit of TEs to lineages is found in the order Rodentia. Although the mouse suffers 10 to 20 times the number of genetic diseases than humans, mouse-like rodents have diversified, adapted, and radiated enormously. The abundant TEs (especially ERVs) in the mouse-like rodents have been harmful to many individuals, but seem to have been enormously beneficial to the lineage, in terms of adaptability, evolution, and diversification of species.

The most active ERV families in mice have lost their *env* gene and infectious capacity and have morphed into retro-TEs with a high level of germ line activity (Feschotte and Gilbert 2012).

6.2.10 Posit (10): Clades or lineages deficient in viable TEs, and with heterogenous populations of non-viable TEs, tend to be non-fecund, can linger in prolonged stasis, and eventually may become “living fossils” or devolve and succumb to background extinction (Chapter 5). Conversely, clades or lineages well endowed with viable and active TEs, especially if they are homogenous, tend to be fecund, or species rich, and to speciate readily. It may be hard to find better contrasting examples of this than the mono-specific naked mole rat and the very speciose mouse-like rodents.

6.2.11 Posit (11): Exogenous retroviruses can infiltrate germ line genomes. Although sometimes harmful, ERVs are often beneficial as they contain promoters which can cause, or alter, gene regulation, and other potentially beneficial coding

sequences that can be exapted by genomes, e.g. *Syncitin-1* and *Syncitin-2*. The “Weismann barrier” once described as fundamental to neoDarwinism, and once used by some to deny the possibility of Ted Steele’s hypotheses (Steele et al.1998) has now been discredited. It has comparatively recently been recognised to have been penetrated on numerous occasions, especially by viruses, but also by horizontally transferred TEs (Chapter 5).

6.2.12 Posit (12): Recurring intermittent waves of TE acquisitions or activity (due to endogenisation of retroviruses (ERVs), horizontal transfer of TEs (HTT) *de novo* modifications to TEs, and/or TE response to stresses on host organisms, and/or variations in the effectiveness of epigenetic and other controls on TEs), can result in punctuated equilibrium type evolution, as observed in the fossil record. Although the recurring intermittent waves of activity are on record, in rodents and vesper bats etc. (Chapter 5), waves of TE activity can also occur in angiosperms, and are often associated with hybridisation and/or polyploidy, or even tissue culture and other such stresses (Chapter 4). The exact role of waves of TEs in punctuation events is still under investigation, and it probably depends a lot on which superfamilies and families the TEs originated from. Certainly the formation of a fertile allopolyploid can often be a punctuation event in angiosperm evolution, if the angiosperm is in stasis, evolving gradually, or even if it is evolving at a significant rate. This punctuation event is accompanied by waves of TE activity which are possibly due to multiple causes, as there is a whole genome

duplication event, a hybridisation event, and a relaxation of epigenetic controls all occurring concurrently. Heritable epialleles resulting from relaxation of epigenetic controls could also be a factor in angiosperm evolution (Chapter 4).

6.2.13 Posit (13): Endogenous retrovirus precursors may have been essential for much of the evolution of cellular biota, as they were an exogenous source of virus genes, such as reverse transcriptase, which are necessary for LTR retro-TE retro-transposition. Conversely, much of all virus evolution was probably facilitated by interactions with cellular biota. There has probably been a co-evolution of viral and cellular life, possibly dating from well before the Cambrian (Villarreal 2005; Feschotte and Gilbert 2012), but retroviruses may be a more recent innovation as their time of origin is unknown (Feschotte and Gilbert 2012). Retroviruses are confined to vertebrates, and some invertebrates (e.g. *Gypsy* in *Drosophila*), but possible retroviruses have recently been found in angiosperms, interacting both with and between, some insects and the angiosperms. Such interactions may have been involved in angiosperm evolution (Chapter 4).

6.2.14 Posit (14): Cellular defences against excessive TE activity have resulted in the capacity of genomes to generate epigenetic controls, such as methylation of CpG sequences. These can possibly modify the epigenome in response to environmental factors, in ways that may be heritable, such as heritable epi-alleles in plants (Chapter 4).

6.2.15 Posit (15): The consortium of differing categories (superfamilies or families) of TEs from endogenous or exogenous sources in a clade or lineage, and their interactions with cellular controls, usually changes over evolutionary time. It seems likely that such changes can have large effects, and directly influence the trajectory, tempo and mode of evolution. Endogenous changes to the consortium can involve *de novo* syntheses, e.g. SINEs, or *de novo* modifications to pre-resident TEs, e.g. new sub-families of LINEs or SINEs, or *de novo* syntheses of chimaeras such as the SVA elements in Hominids. Exogenous changes to the consortium can occur by the horizontal transfer (HTT) of TEs from other taxa, or the endogenisation of invasive exogenous retroviruses in the germ line of the lineage.

6.2.16 Posit (16): The presence of homogenous families of TEs within genomes makes them liable to karyotypic changes by ectopic recombination events. Such karyotypic changes may increase TE activity, so a synergistic system of TEs causing karyotypic changes, causing further TE activity, may be established. This could facilitate the evolution of, and possibly cause divergences, within the lineage.

6.2.17. Posit (17): Significant genomic changes within a lineage, such as duplications and deletions, inversions, translocations etc. can also result from ectopic recombination due to the presence of homogeneous families of TEs.

Chapter 6: General Discussion

6.2.18 Posit (18): Single sequence repeats (SSRs) are also an active accessory to TE-Thrust, often modifying gene regulation.

6.2.19 Posit (19): The *de novo* synthesis of new, or orphan, genes from TEs, and perhaps other non-coding DNA, which has recently been confirmed in fruit fly and humans (Toll-Riera et al. 2009) and may have provided a very beneficial source of new genes, throughout evolutionary history.

6.2.20 Posit (20): TEs can disrupt, and then resurrect genes, and also assist (by means of exon shuffling, and other phenomena) in the synthesis of new genes.

6.2.21 Posit (21): A taxon, in its overall range, is often composed of isolated or semi-isolated local populations (demes, or sometimes even disjunct sub-populations), which may not interbreed. If new TE infiltrations and/or surges of TE activity occur in one or more of these of these populations, or if TE families drift either to fixation or extinction in one or more of these, then such demes or disjunct sub-populations are likely to rapidly diverge from their ancestral genotype and phenotype. The same divergence between demes or sub-populations could occur even if all populations share the same TE consortium at the same level of activity. This is because identical TEs, transposing more or less randomly (active TE-Thrust), or causing more or less random ectopic recombination (passive TE-Thrust), would be likely to alter the genotype, and hence the phenotype, differently in each deme or sub-population. This could trigger an adaptive

radiation. Such a radiation may not always occur rapidly, as initially a novel change may only occur in one genome (excluding the possibility of multiple births), so even if very beneficial, it could take time to become fixed, probably by drift, in the deme, or sub-population. However, such a punctuation event could be relatively rapid, that is almost macro-evolutionary, compared to the near stasis usually inferred from much of the fossil record, and the gradualism implied by many contemporary models of evolution.

6.2.22 (Posit 22): Gene-centric variation and natural selection would normally push a taxon, or a subset of a taxon, higher up its metaphorical adaptive peak and confine it there. TE-Thrust could allow it to cross the metaphorical valley and adopt another peak (adaptation). Adaptive potential derived from TE-Thrust, gene-centric variation, further TE-Thrust, and natural selection, could then closely adapt the taxon to the adopted peak.

6.2.23 (Posit 23): TE-Thrust can sometimes result in almost macro-evolutionary punctuation events, or radiations, in lineages in stasis and these events can sometimes result in a drive towards complexity¹. A more gene-centric adaptive potential, or variation and natural selection process, results in microevolution, leading to fine-tuned adaptation. If only microevolution (adaptation) is occurring in a lineage then it remains more or less in stasis, and may devolve, until it

¹ 'The tendency for diversity and complexity to increase in evolutionary systems' is said to be 'Biology's First Law' (McShea and Brandon 2010). However, there is no denial of the background extinction of lineages in this 'tendency law'.

succumbs to background extinction, or becomes a “living fossil”.

6.2.24 Posit (24): TE-Thrust could be a major contributor to parallel, or to, convergent evolution. This is because there are abundant sources of exogenous DNA sequences, in e.g. retroviruses, that can be endogenised into the genomes of totally unrelated lineages or clades. Also TE families can transfer (HTT or horizontal transposon transfer) between totally unrelated lineages or clades. Some of these endogenised DNA sequences of exogenous origin, e.g. retrovirus *env* genes and LTRs, and many DNA-TE transposase genes, can become exapted or domesticated to become coding, or regulatory sequences etc., in totally unrelated lineages or clades, leading to convergent genomic evolution (Emera 2012) and possibly convergent morphological evolution as well.

6.2.25 (Posit 25): TE-Thrust can facilitate somatic evolution in plants, as these do not have a sequestered germ line throughout life. This can result in different meristems in the same plant having variations in their genotype and phenotype (Chapter 4).

6.2.26 (Posit 26): As there is evidence of *L1* LINE activity in neuronal progenitor cells in humans, TE-Thrust may thus also facilitate some somatic evolution in humans. This, possibly together with epigenetic effects, could help to explain individual differences and could possibly give an explanation for ‘discordant’ monozygotic twins. Such ‘discordance’ is

often said, perhaps incorrectly, to be due to 'nurture' rather than 'nature'.

6.2.27 (Posit 27): TE-Thrust appears to have been an important factor in the origin of such evolutionary novelties as the mammalian placenta (Chapter 6), and the jawed vertebrate immune system. As more data becomes available, it is likely that TE-Thrust will be recognised as an important factor in the origin of many other evolutionary novelties, as via TE contributions to the evolution of regulatory networks (Feschotte 2008; Feschotte and Gilbert 2012). Novelties are distinct from character transformations, such as the evolution of a bird wing, and are the evolution of new characters such as the carapace of the turtle, horns, flowers, and feathers, for example. These require the evolution of new gene regulatory networks (Wagner and Lynch 2010), but this is beyond the scope of this study.

The V(D)J site specific recombination reaction in the immune system of jawed vertebrates is a spectacular example of how TEs can generate complex and crucial functions in the host. There is strong support to indicate that key components of this originate from a formally active *Transib* DNA-TE (Sinzelle et al. 2009).

6.2.28 (Posit 28): The TE-thrust hypothesis offers an explanation for devolution and background extinction (Chapter 5). As ~99% of the species that have ever existed are extinct, and only 5% of all of these have been made extinct in the mass extinction events, the background extinction is very significant

data in evolutionary theory that needs an explanation in any adequate evolutionary theory. The loss of all viable TEs in a species or lineage, and the absence of any new acquisitions of TEs, would result in the corresponding loss of the hypothesised adaptive potential, and evolutionary potential due to TE-Thrust. With a lack of adaptive potential and evolutionary potential, such a species or lineage could then readily succumb to devolution and background extinction, or alternatively become a “living fossil”. The proposal of this explanation for devolution and background extinction does not imply exclusivity, and I acknowledge that there may be other valid explanations in some, or even many, cases.

6.2.29 (Posit 29): TEs may have made it, or helped to make it, a necessity to have two sexes in most eukaryotes, as asexual eukaryote lineages are said to fairly quickly become extinct (Arkhipova and Meselson 2004). The ubiquity of sexual reproduction could be because without it, there is no means of fixing new viable TEs within a population.

6. 3 Conclusions

The TE-Thrust hypothesis has been derived from, and is supported by, the study of peer reviewed published data on mammalian evolution, and to a lesser extent, angiosperm, insect, and reptilian evolution. As can be seen in the South American Sigmodontinae rodents, ERV/sLTRs can be a powerful factor in speciation due to TE-Thrust (see 5.16.2). For many lineages then, it appears that the TE-Trust hypothesis is well founded in the main part, although it still needs further development among many

Chapter 6: General Discussion

more diverse lineages than have been investigated to date. As more and more genomes are sequenced the data to determine its strengths and weaknesses will be more readily available.

6.3.1 Specific Falsifiable Predictions

This hypothesis is conceptual and is based squarely on (to date, somewhat sparse) empirical data, and is not quantitative, at least in its current form. However, some specific falsifiable predictions are (1) No mammalian lineage will be found that has only one, or a very few species, which has had abundant viable TE activity during the last few million years, with these TEs being still active now.

(2) No mammalian lineage will be found that has had an abundance of recent speciation, but that has not had very significant TE activity during these same recent millions of years.*

*As has been repeatedly stressed throughout the thesis, although TE-Thrust is hypothesised to be a powerful facilitator of evolution, it is not claimed to be the sole facilitator of evolution, as there are several other known facilitators of evolution, e.g. whole genome duplication. These other facilitators of evolution could either enhance or diminish these predicted outcomes and may need to be also taken into account.

6.3.2 Peer Acceptance

To date, this hypothesis has been mainly well received by most of our peers, with 48 mostly favorable citations to the paper which makes up Chapter 2. In addition the review which makes up Chapter 3 has been bannered as 'Highly Accessed' by the Mobile

Chapter 6: General Discussion

DNA Journal, which suggests that this review has also generated much interest.

The TE-Thrust hypothesis is presented for scrutiny and development by biologists in the future, before its likely acceptance as a possibly major component of the ongoing and essential further development of evolutionary theory.

Combined Bibliography

- Abbot RJ 1992 Plant invasions, interspecific hybridisation and the evolution of new plant taxa. *Trends in Ecology and Evolution* 7: 401-405.
- Ackerman H Udalova I Hull J & Kwiatkowski D 2002 Evolution of a polymorphic regulatory element in interferon-gamma through transposition and mutation. *Molecular Biology and Evolution* 19:884-890.
- Adams KL 2007 Evolution of duplicate gene expression in polyploidy and hybrid plants. *Journal of Heredity* 98: 136-141.
- Ainouche ML Fortune P Salmon A Grandbastion M-A Fukunaga K et al 2009 Hybridisation, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biological Invasion* 11: 1159-1173.
- Alderton D 1996 *Rodents Of the World*. A Blandford Book London.
- Alföldi J Di Palma F Grabherr M Williams C Kong L et al 2011 The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 447: 587-591. plus Supplementary Information.
- Allegucci C Thurston A Lucas E & Young L 2005 Epigenetics and the germline. *Reproduction* 129: 137–149.
- Amemiya CT Powers TP Prohaska SJ Grimwood J Schmutz J et al 2010 Complete HOX cluster characterisation of the coelacanth provides further evidence for slow evolution of its genome. *Proceedings of the National Academy of Sciences of the USA* 107: 3622-3527.
- Anholt RR & Mackay TF 2001 The genetic architecture of odour-guided behavior in *Drosophila melanogaster*. *Behaviour and Genetics* 31: 7-27.
- Anxolabéhère D Kidwell MG & Periquet G 1988 Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Molecular Biology and Evolution* 5: 252-269.
- Apoil PA Roubinet F Despiau S Mollicone R Oriol R & Blancher A 2000 Evolution of alpha 2-fucosyltransferase genes in primates: relation

Combined Bibliography

- between an intronic Alu-Y element and red cell expression of ABH antigens. *Molecular Biology and Evolution* 17:337-351.
- Arkhipova I & Meselon M 2004 Deleterious transposable elements and the extinction of asexuals. *BioEssays* 27: 76-85.
- Arranz V Kress M & Ernoult-Lange M 2004 The gene encoding the MOK-2 zinc-finger protein: characterization of its promoter and negative regulation by mouse Alu type-2 repetitive elements. *Gene* 149:293-298.
- Ashworth A Skene B Swift S & Lovell-Badge R 1990 Zfa is an expressed retroposon derived from an alternative transcript of the Zfx gene. *EMBO Journal* 9:1529-1534.
- Axsmith BJ Serbet R Krings M Taylor TN Taylor EI 2003 The enigmatic Paleozoic plants *Phasmatocycas* reconsidered. *American Journal of Botany* 90: 1585-1595.
- Babushok DV Ohshima K Ostertag EM Chen X Wang Y et al 2007 A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Research* 17:1129-1138.
- Bacolla A & Wells RD 2009 Non-B DNA conformations as determinants of mutagenesis and human disease. *Molecular Carcinogenesis* 48: 273-285.
- Bailey AD & Shen CK 1993 Sequential insertion of alu family repeats into specific genome sites of higher primates. *Proceedings of the National Academy of Science USA* 90:7205-7209.
- Bailey JA & Eichler EE 2006 Primate segmental duplications: crucibles of evolution, diversity and disease. *Nature Reviews Genetics* 7: 552– 564.
- Bailey JA Liu G & Eichler EE 2003 An Alu transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics* 73: 823–834.
- Ball M McLellan A Collins B Coadwell J Stewart F & Moore T 2004 An abundant placental transcript containing an IAP-LTR is allelic to mouse pregnancy-specific glycoprotein 23 (Psg23): cloning and genetic analysis. *Gene* 325:103-113.

Combined Bibliography

- Baillie K Barnett MW Upton KR Gerhardt DJ Richmond TA et al 2011 Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* doi:10.1038/nature10531.
- Bannert N & Kurth R 2004 Retroelements and the human genome: new perspectives on an old relation. *Proceedings of the National Academy of Science USA* 101: 14572–14579.
- Banville D & Boie Y 1989 Retroviral long terminal repeat is the promoter of the gene encoding the tumor-associated calcium-binding protein oncomodulin in the rat. *Journal of Molecular Biology* 207:481-490.
- Barrier M Robichaux RH & Purugganan MD 2001 Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Science USA* 98: 10208–10213.
- Batzer MA Deininger PL 2002 Alu repeats and human genomic diversity. *Nature Review Genetics* 3:370-379.
- Baust C Baillie GJ & Mager DL 2002 Insertional polymorphisms of ETn retrotransposons include a disruption of the *wiz* gene in C57BL/6 mice. *Mammalian Genome* 13:423-428.
- Beauregard A Curcio MJ & Belfort M 2008 The take and give between retrotransposable elements and their hosts. *Annual Review of Genetics* 42: 12.1-12.31.
- Beck CR Garcia-Perez JL Badge RM & Moran JV 2011 LINE-1 elements in structural variation and disease. *Annual Review of Genomics and Human Genetics* 12: 187-215.
- Beer J-A & Morris P 2005 *Encyclopedia of Mammals; An Essential Guide to Mammals of the World*. Grange Books The Grange Kent UK.
- Bejerano G Lowe CB Ahituv N King B Siepel A et al 2006 A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441: 87–90.
- Bekpen C Marques-Bonet T Alkan C Antonacci F Leogrande MB et al 2009 Death and resurrection of the human IRGM gene. *PLoS Genetics* 5:e1000403.

Combined Bibliography

- Belancio VP Hedges DJ & Deininger P 2008 Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Research* 18: 343–358.
- Bell DT 2001 Ecological response syndromes in the flora of southwestern Australia: fire resprouters versus reseeder. *Botanical Review* 67: 417-440.
- Belyi VA Levine A J & Shalka AM 2010 Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathogens* 6: e1001030.
- Bénit L De Parseval N Casella JF Callebaut I Cordonnier A & Heidmann T 1997 Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. *Journal of Virology* 71:5652-5657.
- Bénit L Lallemand JB Casella JF Philippe H & Heidmann T 1999 ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *Journal of Virology* 73:3301-3308.
- Bennetzen JL 2000 Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* 42: 251–269.
- Benntzen JL 2005 Transposable elements, gene creation and genome rearrangements in flowering plants. *Current Opinion in Genetics & Development* 15: 621-627.
- Beraldi R Pittoggi C Sciamanna I Mattei E & Spadafora C 2006 Expression of LINE-1 retroposons is essential for murine preimplantation development. *Molecular Reproduction and Development* 73: 279–287.
- Berger F 2008 Double fertilisation, from myths to reality. *Sexual Plant Reproduction* 21: 3-5.
- Betrán E & Long M 2002 Expansion of genome coding regions by acquisition of new genes. *Genetica* 115:65-80.
- Bhattacharyya MK Smith AM Ellis THN Hedley C & Martin C 1990 The wrinkled seed character of pea described by Mendel is caused by a

Combined Bibliography

- transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 60: 115-122.
- Bi S Gavrilova O Gong DW Mason MM & Reitman M 1997 Identification of a placental enhancer for the human leptin gene. *Journal of Biological Chemistry* 272:30583-30588.
- Bièche I Laurent A Laurendeau I Duret L Giovangrandi Y et al 2003 Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. *Biological Reproduction* 68:1422-1429.
- Biémont C & Vieira C 2006 Genetics: junk DNA as an evolutionary force. *Nature* 443: 521–524.
- Bird AP 1995 Gene number, noise reduction and biological complexity. *Trends in Genetics* 11: 94–100.
- Blaise S de Parseval N Bénit L & Heidmann T 2003 Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proceedings of the National Academy of Science USA* 100:13013-13018.
- Bogerd HP Wiegand HL Hulme AE Garcia-Perez JL O'Shea KS et al 2006 Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proceedings of the National Academy of Science USA* 103: 8780–8785.
- Böhne A Brunet F Galiana-Arnoux D Schultheis C & Volff JN 2008 Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research* 16: 203–215.
- Bond WJ & Scott AC 2010 Fire and the spread of flowering plants in the Cretaceous. *New Phytologist* 188: 1137-1150.
- Borodulina OR & Kramerov DA 1999 Wide distribution of short interspersed elements among eukaryote genomes. *FEBS Letters* 457: 409-413.
- Borodulina OR & Kramerov DA 2001 Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mammalian Genome* 12: 779-786.
- Borodulina OR & Kramerov DA 2005 PCR-based approach to SINE isolation: simple and complex SINEs. *Gene* 349: 197-205.

Combined Bibliography

- Boschan C Borchert A Ufer C Thiele BJ & Kuhn H 2002 Discovery of a functional retrotransposon of the murine phospholipid hydroperoxide glutathione peroxidase: chromosomal localization and tissue-specific expression pattern. *Genomics* 79:387-394.
- Bourque G Leong B Vega VB Chen X Lee YL et al 2008 Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research* 18: 1752–1762.
- Bowen NJ & Jordon IK 2002 Transposable elements and the evolution of eukaryotic complexity. *Current Issues in Molecular Biology* 4: 65–76.
- Bowen NJ & Jordan IK 2007 Exaptation of protein coding sequences from transposable elements. *Genome Dynamics* 3: 147–162.
- Bowen NJ & McDonald JF 2001 *Drosophila* euchromatic LTR retrotransposons are much younger than in the host species in which they reside. *Genome Research* 11: 1527-1540.
- Bowler PJ 2003 *Evolution: The History of an Idea*. University of California California London England.
- Brandt J Schrauth S Veith AM Froschauer A Haneke T et al 2005 Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. *Gene* 345: 101–111.
- Brettell RI & Dennis ES 1991 Reactivation of a silent Ac following tissue culture is associated with heritable alterations in its methylation pattern. *Molecular Genomics and Genetics* 229: 365-372.
- Brindley PJ Laha T McManus DP & Loukas A 2003 Mobile genetic elements colonizing the genomes of metazoan parasites. *Trends in Parasitology* 19: 79-87.
- Brini AT Lee GM & Kinet JP 1993 Involvement of Alu sequences in the cell-specific regulation of transcription of the gamma chain of Fc and T cell receptors. *Journal of Biological Chemistry* 268:1355-1361.
- Britten R 2006 Transposable elements have contributed to thousands of human proteins. *Proceedings of the National Academy of Science USA* 103: 1798–1803.

Combined Bibliography

- Britten RJ 2010 Transposable element insertions have strongly affected human evolution. *Proceedings of the National Academy of Science USA* 107:19945-19948.
- Britten RJ & Davidson EH 1971 Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *The quarterly Review of Biology* 46: 111-138.
- Brosius J 1991 Retroposons - seeds of evolution. *Science* 251:753.
- Brosius J 1999 Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 107: 209-238.
- Brouha B Meischl C Ostertag E de Boer M Zhang Y et al 2002 Evidence consistent with human L1 retrotransposition in maternal meiosis I. *American Journal of Human Genetics* 71: 327–336.
- Brouha B Schustak J Badge RM Lutz-Prigge S Farley AH et al 2003 Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Science USA* 100: 5280–5285.
- Brown A Dundas P Dixon K & Hopper S 2008 *Orchids of Western Australia*. UWA Press Perth.
- Buchon N & Vaury C 2006 RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity* 96: 195–202.
- Buffenstein R & Yahav S 1991 Is the naked mole-rat *Heterocephalus glaber* an endothermic yet poikilothermic mammal? *Journal of Thermal Biology* 16: 227-232.
- Burki F & Kaessmann H 2004 Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter. *Nature Genetics* 36:1061–1063.
- Burleigh JG Barbazuk WB Davis JM Morse AM & Soltis PS 2012 Exploring diversification and genome size evolution in extant gymnosperms through phylogenetic synthesis. *Journal of Botany*. Article ID 292857. doi:10.1155/2012/292857.
- Bush GL Case SM Wilson AC & Patton JL 1977 Rapid speciation and chromosomal evolution in mammals. *Proceedings of the National Academy of Science USA* 74: 3942-3946.

Combined Bibliography

- Bussel JD & James SH 1997 Rocks as museums of evolutionary processes. Granite Outcrops Symposium: Journal of the Royal Society of Western Australia 80: 221-229.
- Calver M Lyubery A McComb J & Bamford M 2009 Environmental Biology. Cambridge University Press Melbourne.
- Cam HP Noma K Ebina H Levin HL & Grewal SI 2008 Host genome surveillance for retrotransposons by transposon-derived proteins. Nature 451: 431–436.
- Cantrell MA Ederer MM Erickson IK Swier VJ Baker RJ & Wichman HA 2005 MysTR: an endogenous retrovirus family in mammals that is undergoing recent amplifications to unprecedented copy numbers. Journal of Virology 79:14698-14707.
- Cantrell MA Scott LA Brown CJ Martinez AR & Wichman HA 2008 Loss of LINE 1 activity in the megabats, Genetics 178: 393-404.
- Caporale LH 2006 An Overview of the Implicit Genome. In: The Implicit Genome. Caporale L ed Oxford University Press New York.
- Caporale LH 2009 Putting together the pieces: evolutionary mechanisms at work within genomes. BioEssays 31: 700-702.
- Capy P 1998 A plastic genome. Nature 396: 522-523.
- Capy P Langin T Higuet D Maurer P & Bazin C 1997 Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica 100: 63–72.
- Caras IW Davitz MA Rhee L Weddell G Martin DW Jr & Nussenzweig V 1987 Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. Nature 325:545-549.
- Carleton MD & Musser GG 2005 Order Rodentia *In*: Wilso DE Reader DM eds. Mammal Species of the World: A Taxonomic and Geographic Reference. John Hopkins University Press Baltimore.
- Carroll S B 2005 Evolution at two levels: on genes and form. PLoS Biology 3: e245.
- Casavant NC Scott LA Cantrell MA Wiggins LE Baker RJ & Wichman HA 2000 The end of the LINE?: Lack of recent L1 activity in a group of South American rodents. Genetics 154: 1809-1817.

Combined Bibliography

- Chang-Yeh A Mold DE Brilliant MH & Huang RC 1993 The mouse intracisternal A particle-promoted placental gene retrotransposition is mouse-strain-specific. *Proceedings of the National Academy of Science USA* 90:292-296.
- Chantret N Salse J Sabot F Rahman S Bellec A et al. 2005 Molecular basis of evolutionary events that shaped the *Hardness* locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *The Plant Cell* 17: 1033-1045.
- Chen JM Stenson PD Cooper DN & Férec C 2005 A systematic analysis of LINE-1 endonuclease dependent retro-transpositional events causing human genetic disease. *Human Genetics* 117: 411–427.
- Chen S & Li X 2007 Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. *BMC Evolutionary Biology* 7: 46.
- Chernova T Higginson FM Davies R & Smith AG 2008 B2 SINE retrotransposon causes polymorphic expression of mouse 5-aminolevulinic acid synthase 1 gene. *Biochemical and Biophysical Research Communications* 377:515-520.
- Chu WM Ballard R Carpick BW Williams BR & Schmid CW 1998 Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Molecular and Cell Biology* 18: 58–68.
- Chung H Bogwitz MR McCart C Andrianopoulos A, French-Constant RH Bterham P & Dabon PJ 2007 *Cis*-Regulatory elements in the Accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* 175: 1071-1077.
- Coffin JM Hughes SH Varmus HE editors *Retroviruses* 1997 Cold Spring harbor Laboratory Press: Cold Spring Harbor New York.
- Cohen CJ Lock WM & Mager DL 2009 Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* 448:105-114.
- Comai L 2000 Genetic and epigenetic interactions in allopolyploid plants. *Plant Molecular Biology* 43: 387-399.

Combined Bibliography

- Comfort N 2001a From controlling elements to transposons: Barbara McClintock and the Nobel Prize. *TRENDS in Genetics* 17: 475-478.
- Comfort N 2001b Barbara McClintock. *Encyclopedia of Life Sciences*. John Wiley & Sons Ltd.
- Cordaux R & Batzer MA 2009 The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* 10:691-703.
- Cordaux R Udit S Batzer M & Feschotte C 2006 Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proceedings of the National Academy of Science USA* 103: 8101–8106.
- Costas J 2003 Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L. *Journal of Molecular Evolution* 56: 181–186.
- Coufal NG Garcia-Perez JL Peng GE Yeo GW Mu Y et al 2009 L1 retrotransposition in human neural progenitor cells. *Nature* 460:1127-1131.
- Crepet WL & Niklas KJ 2009 Darwin's second "Abominable Mystery": why are there so many angiosperm species? *American Journal of Botany* 96: 366-381.
- Cserti CM & Dzik WH 2007 The ABO blood group system and *Plasmodium falciparum* malaria. *Blood* 110:2250-2258.
- Cubo J 2003 Evidence for speciation change in the evolution of ratites (Aves: Palaeognathae). *Biological Journal of the Linnean Society* 80: 99-106.
- Cuperus JT Fahlgren N & Carrington JC 2011 Evolution and functional diversification of *MIRNA* genes. *The Plant Cell* 23: 431-442.
- Dabom PJ Yen JL Bogwitz MR Le Goff G Feil E et al 2002 A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* 297: 2253-2256.
- Damert A Löwer J & Löwer R 2004 Leptin receptor isoform 219.1: an example of protein evolution by LINE-1-mediated human-specific retrotransposition of a coding SVA element. *Molecular Biology and Evolution* 21:647-651.

Combined Bibliography

- Daniels SB Peterson KR Strausbaugh LD Kidwell MG & Chovnick A 1990 Evidence for the horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics* 124: 339-345
- Darboux I Charles J-F Pauchet Y Warot S & Pauron D 2007 Transposon-mediated resistance to *Bacillus sphaericus* in a field-evolved population of *Culex pipiens* (Diptera: Culicidae). *Cellular Microbiology* 9: 2022-2029.
- Davies TJ Barraclough TG Chase MW Soltis PS Soltis DE & Savolainen V 2004 Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Science USA* 101: 1904-1909. www.pnas.org/cgi/doi/10.1073/pnas.0308127100
- DeBerardinis RJ Goodier JL Ostertag EM & Kazazian HH Jr 1998 Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nature Genetics* 20: 288–290.
- De Boer JG Yazawa R Davidson WS & Koop BF 2007 Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8: 422. doi:10.1186/1471-2164-8-422 .
- Deininger PL & Batzer MA 1999 Alu repeats and human disease. *Molecular Genetics and Metabolism* 67: 183–193.
- Deininger PL Moran JV Batzer MA & Kazazian HH Jr 2003 Mobile elements and mammalian genome evolution. *Current Opinion in Genetics and Development* 13: 651–658.
- De Mendoza A Escobedo DE Dávila IM & Saldaña H 2004 Expansion and divergence of the GH locus between spider monkey and chimpanzee. *Gene* 336:185-193.
- Dennis ES & Brettrell RI 1990 DNA methylation of maize transposable elements is correlated with activity. *Philosophical Transactions of the Royal Society of London B Biological Science* 326: 217-219.
- Dewannieux M Esnault C & Heidmann T 2003 LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics* 35:41-48.

Combined Bibliography

- Dimitri P & Junakovic N 1999 Revising the selfish DNA hypothesis, new evidence on accumulation of transposable elements in heterochromatin. *TIG* 15: 123-124.
- Ding & Voinnet 2007 Antiviral immunity directed by small RNAs. *Cell* 130: 413-426.
- Dobigny G Waters PD & Robinson T J 2006 Absence of hypomethylation and LINE-1 amplification in a white x black rhinoceros hybrid. *Genetica* 127: 81-86.
- Doebly J Stec A Wendel J & Edwards M 1990 Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. *Proceedings of the National Academy of Science USA* 87: 9888-9892.
- Doebly J Stec A & Gustus C 1995 teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141: 333-346.
- Donlin MJ Lisch D & Freeling M 1995 Tissue-specific accumulation of MURB, a protein encoded by MuDR, the autonomous regulator of the Mutator transposable element family. *Plant Cell* 7: 1989-2000.
- Doolittle WF & Sapienza C 1980 Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603.
- Dover GA 1982 Molecular Drive: A cohesive model of species evolution. *Nature* 299: 111-117.
- Duhl DM Vrieling H Miller KA Wolff GL & Barsh GS 1994 Neomorphic agouti mutations in obese yellow mice. *Nature Genetics* 8:59-65.
- Dulai KS von Dornum M Mollon JD Hunt DM 1999 The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Research* 9:629-638.
- Dunn CA Medstrand P & Mager DL 2003 An endogenous retroviral long terminal repeat is the dominant promoter for human beta 1,3-galactosyltransferase 5 in the colon. *Proceedings of the National Academy of Science USA* 100:12841-12846.

Combined Bibliography

- Dunn CA Romanish MT Gutierrez LE van de Lagemaat LN & Mager DL 2006 Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* 366:335-342.
- Dupressoir A & Heidmann T 1996 Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice. *Molecular and Cell Biology* 16: 4495–4503.
- Dupressoir A Marceau G Vernochet C Bénit L Kanellopoulos C et al 2005 Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proceedings of the National Academy of Science USA* 102:725-730.
- Dupressoir A Vernochet C Bawa O Harper F Pierron G et al 2009 Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proceedings of the National Academy of Science USA* 106: 12127-12132.
- Dupuy C Periquet G Serbielle C Bézier A Louis F & Drezden J-M 2011 Transfer of a chromosomal *Maverick* to endogenous bracovirus in a parasitoid wasp. *Genetica* 139: 489-496.
- Earl DJ & Deem MW 2004 Evolvability is a selectable trait. *Proceedings of the National Academy of Science USA* 101: 11531-11536.
- Ebihara M Ohba H Ohno SI & Yoshikawa T 2002 Genomic organization and promoter analysis of the human nicotinic acetylcholine receptor alpha6 subunit (CHNRA6) gene: Alu and other elements direct transcriptional repression. *Gene* 298:101-108.
- Eickbush TH 2002 Repair by retrotransposition. *Nature Genetics* 31: 126–27.
- Eickbush TH & Furano AV 2002 Fruit flies and humans respond differently to retrotransposons. *Current Opinions in Genetics and Development* 12: 669– 674.
- Eickbush TH & Jamburuthugoda VK 2008 The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research* 134: 221-234. doi:10.1016/j.virusres.2007.12.010.
- Eldredge N 1986 *Time Frames: The Rethinking of Darwinian Evolution and the Theory of Punctuated Equilibria*. William Heinemann Ltd London.

Combined Bibliography

- Eldredge N 1995 Reinventing Darwin: The Great Debate at the High Table of Evolutionary Theory. John Wiley New York.
- Eldredge N & Gould SJ 1972 Punctuated equilibria: an alternative to phyletic gradualism. In: Models in Paleobiology. TJM Schopf (ed) Freeman Cooper San Francisco pp 82-115.
- Emera D Casola C Lynch VJ Wildman DE Agnew D & Wagner G P 2012 Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. *Molecular Biology and Evolution* 29: 239-247.
- Erickson IK Cantrell MA Scott LA & Wichman HA 2011 Retrofitting the genome: L1 extinction follows retroviral expansion in a group of Muroid rodents. *Journal of Virology* doi:10.1128/JVI.05180-11.
- Erwin DH 2001 Lessons from the past: Biotic recoveries from mass extinctions. *Proceedings of the National Academy of Science USA* 98: 5399–5403.
- Evgen'ev MB Zelentsova H Poluectova H Lyozin GT Velei- kodvorskaja V et al 2000 Mobile elements and chromosomal evolution in the virilis group of *Drosophila*. *Proceedings of the National Academy of Science USA* 97: 11337– 11342.
- Evsikov AV de Vries WN Peaston AE Radford EE Fancher KS et al 2004 Systems biology of the 2-cell mouse embryo. *Cytogenetics and Genome Research* 105: 240–250.
- Ewing AD & Kazazian HH Jr 2010 High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Research* 20: 1262-1270.
- Fantaccione S Woodrow P & Pontecorvo G 2008 Identification of a family of SINEs and LINEs in the *Pipistrellus kuhli* genome: A new structural and functional symbiotic relationship. *Genomics* 91: 178-185.
- Farcas R Schneider E Frauenknecht K Kondova I Bontrop R et al 2009 Differences in DNA methylation patterns and expression of the CCRK gene in human and nonhuman primate cortices. *Molecular Biology and Evolution* 26:1379-1389.

Combined Bibliography

- Fedoroff NV 1999 Transposable elements as a molecular evolutionary force. *Annals of the New York Academy of Science* 870: 251–264.
- Fedoroff N 2000 Transposons and gene evolution in plants. *Proceedings of the National Academy of Science* 97: 7002-7007.
- Ferrigno O Virolle T Djabari Z Ortonne JP White RJ & Aberdam D 2001 Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nature Genetics* 28:77-81.
- Feschotte C 2008 Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* 9: 397–405.
- Feschotte C & Gilbert C 2012 Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics* 13: 283-296.
- Feschotte C & Pritham EJ 2007 DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* 41: 331–368.
- Feschotte C & Pritham EJ 2009 A cornucopia of Helitrons shapes the maize genome. *Proceedings of the National Academy of Science USA* 106: 19747-19748.
- Feschotte C Jiang N & Wessler SR 2002 Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3: 329-341.
- Fillingham JS Thing TA Vythilingum N Keuroghlian A Bruno D et al 2004 A non-long terminal repeat retrotransposon family is restricted to the germ line micronucleus of the ciliated protozoan *Tetrahymena thermophila*. *Eukaryotic Cell* 3: 157–169.
- Fitch DH Bailey WJ Tagle DA Goodman M Sieu L & Slightom JL 1991 Duplication of the gammaglobin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proceedings of the National Academy of Science USA* 88:7396-7400.
- Flannery TF 1994 *The Future Eaters*. Reed Books Chatswood NSW.
- Fornasari D Battaglioli E Flora A Terzano S & Clementi F 1997 Structural and functional characterization of the human alpha3 nicotinic subunit gene promoter. *Molecular Pharmacology* 51:250-261.

Combined Bibliography

- Fortey RA 1985 Gradualism and punctuated equilibrium as competing and complementary theories. *Special Papers in Palaeontology* 33: 17–28. Cited by Gould 2002.
- Foster AS & Gifford EM 1974 *Comparative Morphology of Vascular Plants* (Second edition). Freeman San Fransisco.
- Fourel G Transy C Tennant BC & Buendia MA 1992 Expression of the woodchuck N-myc2 retroposon in brain and in liver tumors is driven by a cryptic N-myc promoter. *Molecular and Cellular Biology* 12:5336-5344.
- Fried C Prohaska SJ & Stadler PF 2004 Exclusion of repetitive DNA elements from gnathostome Hox clusters. *Journal of Experimental Zoology B Molecular and Developmental Evolution* 302: 165–173.
- Friedman WE 2009 The meaning of Darwin's "abominable mystery" *American Journal of Botany* 96: 5-21.
- Friedman WE & Carmichael JS 1996 Double fertilization in Gnetales: Implications for understanding reproductive diversification among seed plants. *International Journal of Plant Science* 157(6 Suppl) S77-S94.
- Furano AV Duvernell DD & Boissinot S 2004 L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends in Genetics* 20: 9-14.
- Gager CS 1910 *Translators notes on Intracellular Pangenesis*. (University of Missouri).
- Galagan JE & Selker EU 2004 RIP: the evolutionary cost of genome defense. *Trends in Genetics* 20: 417–423.
- Galagan JE Calvo SE Borkovich KA Selker EU Read ND et al 2003 The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859–868.
- Gallardo MH Kausel G Jiménez A Bacquet C González C et al. 2004 Whole genome duplications in South American desert rodents (Octodontidae). *Biological Journal of the Linnean Society* 82: 443-451.
- Gallardo MH González CA & Cebrián I 2006 Molecular cytogenetics and alloteraploidy in the red vizcacha rat, *Tympanoctomys barrerae* (Rodentia, Octodontidae). *Genomics* 88: 214-221.

Combined Bibliography

- Georgiev GP 1984 Mobile genetic elements in animal cells and their biological significance. *European Journal of Biochemistry* 145:203-220.
- Gerasimova TI Matjunina LV Mizrokhi LJ & Georgiev GP 1985 Successive transposition explosions in *Drosophila melanogaster* and reverse transpositions of mobile dispersed genetic elements. *EMBO Journal* 4: 3773–3779.
- Gerlo S Davis JR Mager DL & Kooijman R 2006 Prolactin in man: a tale of two promoters. *BioEssays* 28:1051-1055.
- Gibbs RA Weinstock GM Metzker M L Muzny DM Sodergren EJ et al 2004 Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493–521.
- Gibbs RA Rogers J Katze MG Bumgarner R Weinstock GM et al 2007 Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222–234.
- Gilbert C Hernandez SS Flores-Benabib J Smith EN & Feschotte C 2012 Rampant horizontal transfer of *SPIN* transposons in squamate reptiles. *Molecular Biology and Evolution* 29: 503-515.
- Gilbert N & Labuda D 1999 CORE-SINES: Eukaryotic short interspersed retroposing elements with common sequence motifs. *Proceedings of the National Academy of Science USA* 96:2869-2874.
- Gilbert SF McDonald E Boyle N Buttino N Gyi L et al 2010 Symbiosis as a source of selectable epigenetic variation: taking the heat for the big guy. *Philosophical Transactions of the Royal Society B* 365: 671-678. doi: 1098/rstb.2009.0245.
- Ginzburg LR Bingham PM & Yoo S 1984 On the theory of speciation induced by transposable elements. *Genetics* 107: 331–341.
- Gogolevsky KP Vassetzky NS & Kramerov DA 2008 Bov-B-mobilised SINES in vertebrate genomes. *Gene* 407: 75-85.
- Goldschmidt R 1940 *The Material Basis of Evolution*. Yale University Press New Haven. *Nature* 284: 604–607.
- Gombart AF Saito T & Koeffler HP 2009 Exaptation of an ancient Alu short interspersed element provides a highly conserved vitamin D-mediated

Combined Bibliography

- innate immune response in humans and primates. *BMC Genomics* 10:321.
- González J Macpherson M & Petrov DA 2009 A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Molecular Biology and Evolution* 26: 1949-1961.
- González J Karasov TL Messer PW & Petrov DA 2010 Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genetics* 6(4): e1000905. doi:10.1371/journal.pgen.1000905
- Goodier JL & Kazazian Jr HH 2008 Retrotransposons revisited: The restraint and rehabilitation of parasites. *Cell* 135 23–35.
- Goodier JL Ostertag EM & Kazazian HH Jr 2000 Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Human Molecular Genetics* 9:653-657.
- Goodier JL Ostertag EM Du K & Kazazian HH Jr 2001 A novel active L1 retrotransposon subfamily in the mouse. *Genome Research* 11: 1677–1685.
- Gorelick R & Olson K 2011 Is the lack of cycad (Cycadales) diversity a result of a lack of polyploidy? *Botanical Journal of the Linnean Society* 165: 156-157.
- Gould SJ 2002 *The Structure of Evolutionary Theory*. The Belknap Press of Harvard University Press Cambridge Massachusetts.
- Gould SJ & Eldredge N 1977 Punctuated equilibria; the tempo and mode of evolution reconsidered. *Paleobiology* 3: 115–151.
- Grahn RA Rinehart TA Cantrell MA & Wichman HA 2005 Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. *Cytogenetics and Genome Research* 110: 407-415.
- Grant V 1971 *Plant Speciation*. Columbia University Press New York.
- Granzotto A Lopes FR Vieira C & Carareto CMA 2011 Vertical inheritance and bursts of transposition have shaped the evolution of the *BS* non-LTR retrotransposon in *Drosophila*. *Molecular Genetics and Genomics* DOI 10.1007/s00438-011-0629-9.

Combined Bibliography

- Grassé P-P 1977 Evolution of Living Organisms. Academic Press, New York.
- Grechko VV 2011 Repeated DNA sequences as an engine of biological diversification. *Molecular Biology* 45: 704-727.
- Grover D Majumder PP Rao BC Brahmachari SK & Mukerji M 2003 Nonrandom distribution of alu elements in genes of various functional categories: insight from analysis of human chromosomes 21 and 22. *Molecular Biology and Evolution* 20: 1420–1424.
- Gupta S Gallavotti A Stryker GA Schmidt RJ & Lal SK 2005 A novel class of *Helitron*-related transposable elements in maize contain portions of multiple pseudogenes. *Plant Molecular Biology* 57: 115-127.
- Hagan CR Sfield RF & Rudin CM 2003 Human Alu element retrotransposition induced by genotoxic stress. *Nature Genetics* 35: 219–220.
- Hambor JE Mennone J Coon ME Hanke JH & Kavathas P 1993 Identification and characterization of an Alu-containing, T-cell-specific enhancer located in the last intron of the human CD8 alpha gene. *Molecular and Cell Biology* 13:7056-7070.
- Han K Sen SK Wang J Callinan PA Lee J et al 2005 Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Research* 33:4040-4052.
- Han K Lee J Meyer TJ Wang J Sen SK et al 2007 Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS Genetics* 3: 1939–1949.
- Han W Kasai S Hata H Takahashi T Takamatsu Y Yamamoto H I, et al 2006 Intracisternal A-particle element in the 3' noncoding region of the mu-opioid receptor gene in CXBK mice: a new genetic mechanism underlying differences in opioid sensitivity. *Pharmacogenetics and Genomics* 16:451-460.
- Hanada K Vallrgo V Nobuta K Slotkin RK Lisch D et al 2009 The functional role of Pack-MULEs in rice inferred from purifying selection and expression profile. *The Plant Cell* 21: 25-38.
- Harendza CJ & Johnson LF 1990 Polyadenylation signal of the mouse thymidylate synthase gene was created by insertion of an L1 repetitive

Combined Bibliography

- element downstream of the open reading frame. Proceedings of the National Academy of Science USA 87:2531-2535.
- Haring E Hagemann S Pinsker W 2000 Ancient and recent horizontal invasions of Drosophilids by P elements. Journal of Molecular Evolution 51:577-586.
- Harris JR 1991 The evolution of placental mammals FEBS 10508 295: 1.2.3. 3-4.
- Häsler J & Strub K 2006 Alu RNP and Alu RNA regulate translation initiation in vitro. Nucleic Acids Research 34: 2374–2385.
- Hata K & Sakaki Y 1997 Identification of critical CpG sites for repression of L1 transcription by DNA methylation. Gene 189: 227–234.
- Hayakawa T Satta Y Gagneux P Varki A & Takahata N 2001 Alu-mediated inactivation of the human CMP- N-acetyl-neuraminic acid hydroxylase gene. Proceedings of the National Academy of Science USA 98:11399-11404.
- Heard E Tishkoff S Todd JA Vidal M Wagner GP et al 2010 Ten years of genetics and genomics: what have we achieved and where are we heading? Nature Reviews Genetics doi:10.1038/nrg2878 .
- Hedges DJ & Batzer MA 2005 From the margins of the genome: mobile elements shape primate evolution. BioEssays 27: 785–794.
- Hedges DJ Callinan PA Cordaux R Xing J Barnes E & Batzer MA 2004 Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Research 14:1068-1075.
- Heidmann O Vernochet C Dupressoir A & Heidmann T 2009 Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new “syncytin” in a third order of mammals. Retrovirology 6:107.
- Heimberg AM Sempere LF Moy VN Donoghue PCJ & Peterson KJ 2008 MicroRNAs and the advent of vertebrate morphological complexity. Proceedings of the National Academy of Science of the USA 105: 2946-2950.
- Hendriksen PJ Hoogerbrugge JW Baarends WM de Boer P Vreeburg JT et al 1997 Testis-specific expression of a functional retroposon encoding

Combined Bibliography

- glucose-6-phosphate dehydrogenase in the mouse. *Genomics* 41:350-359.
- Herédia F Loreto ELS & Valente VLS 2004 Complex evolution of *gypsy* in Drosopholoid species. *Molecular Biology and Evolution* 21: 1831-1942.
- Hess JF Fox M Schmid C & Shen CK 1983 Molecular evolution of the human adult alpha-globin-like gene region: insertion and deletion of Alu family repeats and non-Alu DNA sequences. *Proceedings of the National Academy of Science USA* 80:5970-5974.
- Hewitt SM Fraizer GC & Saunders GF 1995 Transcriptional silencer of the Wilms' tumor gene WT1 contains an Alu repeat. *Journal of Biological Chemistry* 270:17908-17912.
- Hickey DA 1982 ~~Sex~~ DNA: a sexually -transmitted nuclear parasite. *Genetics* 101: 519–531.
- Hill AS Foot NJ Chaplin TL & Young BD 2000 The most frequent constitutional translocation in humans, the t(11;22)(q23;q11) is due to a highly specific alu -mediated recombination. *Human Molecular Genetics* 9: 1525–1532.
- Hirochika H Sugimoto K Otsuki Y Tsugawa H & Kanda M 1996 Retrotransposons of rice involved in mutations induced by tissue culture. *Proceedings of the National Academy of Science USA* 93: 7783-7788.
- Hoffmann FG Opazo JC & Storz J 2011 Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates. *Molecular Biology and Evolution* doi:10.1093/molbev/msr207
- Holmes I 2002 Transcendent elements: whole-genome screens and open evolutionary questions. *Genome Research* 12: 1152-1155.
- Holt JE Roman SD Aitken RJ & McLaughlin EA 2006 Identification and characterization of a novel Mt-retrotransposon highly represented in the female mouse germline. *Genomics* 87:490-409.

Combined Bibliography

- Horie M Honda T Suzuki Y Kobayashi Y Daito T et al 2010 Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* 463: 84-88. doi:1038/nature08695.
- Howard BH & Sakamoto KD 1990 Alu interspersed repeats: selfish DNA or a functional gene family? *The New Biologist* 2: 759-770.
- Howe HF & Smallwood 1982 Ecology of seed dispersal. *Annual Review of Ecology and Systematics* 13: 201-228.
- Hsu DW Lin MJ Lee TL Wen SC Chen X & Shen CK 1999 Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues. *Proceedings of the National Academy of Science USA* 96:9751-9756.
- Huang CJ Lin WY Chang CM & Choo KB 2009 Transcription of the rat testis-specific Rtdpoz-T1 and -T2 retrogenes during embryo development: co-transcription and frequent exonisation of transposable element sequences. *BMC Molecular Biology* 10:74.
- Hua-Van A Le Rouzic A Boutin TS Filée J & Capy P 2011 The struggle for life of the genome's selfish architects. *Biology Direct* 6: 19 <http://www.biology-direct.com/content/6/1/19>
- Huchon D Chevret P Jordan U Kilpatrick CW Ranwez V et al. 2007 Multiple molecular evidences for a living mammalian fossil. *Proceedings of the National Academy of Sciences of the USA* 104: 7495-7499.
- Huda A Mariño-Ramirez L & King Jordan I 2010 Epigenetic histone modifications of human transposable elements: genome defence versus exaptation. *Mobile DNA* 1:2.
- Hughes JF & Coffin JM 2001 Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nature Genetics* 29: 487-489.
- Hughes JF and Coffin JM 2004 Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proceedings of the National Academy of Sciences USA* 101: 1668-1672.
- Huh JW Kim TH Yi JM Park ES Kim WY et al 2006 Molecular evolution of the periphilin gene in relation to human endogenous retrovirus m element. *Journal of Molecular Evolution* 62:730-737.

Combined Bibliography

- Huh JW Ha H Kim DS & Kim HS 2008a Placenta-restricted expression of LTR- derived NOS3. *Placenta* 29:602-608.
- Huh JW Kim DS Ha HS Lee JR Kim YJ et al 2008b Cooperative exonization of MaLR and AluJo elements contributed an alternative promoter and novel splice variants of RNF19. *Gene* 424:63-70.
- Huh JW Kim YH Lee SR Kim H Kim DS et al 2009 Gain of new exons and promoters by lineage-specific transposable elements- integration and conservation event on CHRM3 gene. *Molecules and Cells* 28:111-117.
- International Chicken Genome Sequencing Consortium 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.
- International Brachypodium Initiative 2010 Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463:763-8.
- Jain SM 2001 Tissue culture-derived variation in crop improvement. *Euphytica* 118: 153-156.
- Jiang N Bao Z Zhang X Eddy SR & Wessler SR 2004 Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431: 569-573.
- Jin H Selfe J Whitehouse C Morris JR Solomon E & Roberts RG 2004 Structural evolution of the BRCA1 genomic region in primates. *Genomics* 84:1071-1082.
- Johnson ME Cheng Z Morrison VA Scherer S Ventura M et al 2006a Recurrent duplication-driven transposition of DNA during hominoid evolution. *Proceedings of the National Academy of Science USA* 103: 17626–17631.
- Johnson RM Pritchko T Gumucio D Wildman DE Uddin M & Goodman M 2006b Phylogenetic comparisons suggest that distance from the locus control region guides developmental expression of primate beta-type globin genes. *Proceedings of the National Academy of Science USA* 103:3186-3191.
- Johnston SL den Nijs TPM Peloquin SJ & Hanneman RE Jr 1980 The significance of genic balance to endosperm development in interspecific crosses. *Theoretical & Applied Genetics* 57: 5–9.

Combined Bibliography

- Jordan IK Rogozin IB Glazko GV & Koonin EV 2003 Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics* 19: 68–72.
- Jurka J 1989a Subfamily structure and evolution of the human L1 family of repetitive sequences. *Journal of Molecular Evolution* 29: 496-503.
- Jurka J 1989b Novel families of interspersed repetitive elements from the human genome. *Nucleic Acid research* 18:137-141.
- Jurka J 1998 Repeats in genomic DNA: mining and meaning. *Current Opinion in Structural Biology* 8: 333-337.
- Jurka J 2004 Evolutionary impact of human Alu repetitive elements. *Current Opinions in Genetics and Development* 14: 603–608.
- Jurka J 2008 Conserved eukaryotic transposable elements and the evolution of gene regulation. *Cellular and Molecular Life Sciences* 65: 201-204.
- Jurka J Kapitonov VV Klonowski P Walichiewicz J & Smit AF 1996/7 Identification of new medium reiteration frequency repeats in the genomes of Primates, Rodentia and Lagomorpha. *Genetics* 3: 235-247.
- Jurka J & Kapitonov 1999 Sectorial mutagenesis by transposable elements. *Genetica* 107: 239-248.
- Jurka J Bao W & Kojima KK 2011 Families of transposable elements, population structure and the origin of species. *Biology Direct* 6:44 doi:10.1186/1745-6150-6-44.
- Kramerov DA & Vassetsky NS 2011 Origin and evolution of SINEs in eukaryote genomes. *Heredity* 107: 487-495.
- Kämmerer U Germeyer A Stengel S Kapp M & Denner J 2011 Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta. *Journal of Reproduction and Immunology* doi:10.1016/j.jri.2011.06.102
- Kanazawa A Liu B Kong F Arase S & Abe J 2009 Adaptive evolution involving gene duplication and insertion of a novel *Ty1/copia*-like retrotransposon in soybean. *Journal of Molecular Evolution* 69: 164-175. DOI 10.1007/s00239-009-9161-1.

Combined Bibliography

- Kapitonov VV & Jurka J 2004 *Harbinger* transposons and an ancient HARBI1 gene derived from a transposase. *DNA and Cell Biology* 23: 311-324.
- Kapitonov VV & Jurka J 2005 RAG1 Core and V(D)J recombination signal sequences were derived from transib transposons. *PLoS Biology* 3:e181. [PubMed: 15898832]
- Kapitonov VV & Jurka J 2006 Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Science USA* 103: 4540–4545.
- Kapitonov VV & Jurka J 2007 *Helitrons* on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics* 23: 521-529.
- Kapitonov VV & Jurka J 2008 A universal *class* of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics* 9: 411–412.
- Kashkush K Feldman M Levy AA 2003 Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 33: 102-106.
- Kashi Y & King DG 2006 Has simple sequence repeat mutability been selected to facilitate evolution? *Israel Journal of Ecology & Evolution* 52: 331–334.
- Kashkush K and Khasdam V 2007 Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics* 177: 1975-1985.
- Kassahn KS Dang VT Wilkins SJ Perkins AC & Ragan MA 2009 Evolution of gene function and regulatory control after whole-genome duplication: Comparative analysis in vertebrates. *Genome Research* 19: 1404-1418.
- Katzourakis A & Gifford RJ 2010 Endogenous viral elements in animal genomes. *PLoS Genetics* 6(11): e1001191.doi:10.1371/journal.pgen.1001191.
- Kauffman SA & Johnsen S 1991 Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches. *Journal of Theoretical Biology* 149: 467–505.

Combined Bibliography

- Kazazian Jr HH 1998 Mobile elements and disease. *Current Opinion in Genetics and Development* 8: 343-350.
- Kazazian HH Jr 1999 An estimated frequency of endogenous insertional mutations in humans. *Nature Genetics* 22: 130.
- Kazazian HH Jr 2004 Mobile elements: drivers of genome evolution. *Science* 303: 1626–1632.
- Keary P 1996 *The New Penguin Dictionary of Biology*. Penguin Books London.
- Kehrer-Sawatzki H Sandig C Chuzhanova N Goidts V Szamalek JM et al 2005 Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*). *Human Mutation* 25:45-55.
- Kejnovsky E Leitch IJ & Leitch AR 2009 Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Ecology and Evolution* doi:10.1016/j.tree.2009.04.010
- Keller A 2007 *Drosophila melanogaster's* history as a human commensal. *Current Biology* 17: R77-R81.
- Kember RL Fernandes C Tunbridge EM Liu L Payá-Cano JL et al 2010 A B2 SINE insertion in the *Comt1* gene (*Comt1(B2i)*) results in an overexpressing, behavior modifying allele present in classical inbred mouse strains. *Genes Brain Behaviour* 9:925-932.
- Khan H Smit A & Boissinot S 2006 Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research* 16: 78–87.
- Kidwell MG 2002 Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49–63.
- Kidwell MG & Lisch D 1997 Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Science of the USA* 94: 7704-7711.
- Kidwell MG & Lisch DR 2000 Transposable elements and host genome evolution. *Trends in Ecology and Evolution* 15: 95-99.
- Kidwell MG & Lisch DR 2001 Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 55: 1–24.

Combined Bibliography

- Kim EB Fang X Fushan AA Huang Z Lobanov AV et al. 2011 Genome sequence reveals insights into physiology and longevity of the naked mole rat. *Nature* 479:223-227.
- Kim TM Hong SJ & Rhyu MG 2004 Periodic explosive expansion of human retroelements associated with the evolution of the hominoid primate. *Journal of Korean Medical Science* 19: 177–185.
- Kimura RH Choudary PV Stone KK & Schmid CW 2001 Stress induction of Bm1 RNA in silkworm larvae: SINEs, an unusual class of stress genes. *Cell Stress Chaperones* 6: 263–272.
- King D Snider LD & Lingrel JB 1986 Polymorphism in an androgen-regulated mouse gene is the result of the insertion of a B1 repetitive element into the transcription unit. *Molecular and Cell Biology* 6:209-217.
- King DC Trifonov EN & Yechezkei K 2006 Turning knobs in the genome: evolution of simple sequence repeats by indirect selection. In: *The Implicit Genome*. L Caporale (ed). Oxford University Press New York.
- Kjeldbjerg AL Villesen P Aagaard L & Pedersen FS 2008 Gene conversion and purifying selection of a placenta-specific ERV-V envelope gene during simian evolution. *BMC Evolutionary Biology* 8:266.
- Kleene KC Mulligan E Steiger D Donohue K & Mastrangelo MA 1998 The mouse gene encoding the testis-specific isoform of Poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes. *Journal of Molecular Evolution* 47: 275–281.
- Komiyama H Aoki A Tanaka S Maekawa H Kato Y et al 2010 Alu-derived cis-element regulates tumorigenesis-dependent gastric expression of GASDERMIN B (GSDMB). *Genes Genetics and Systematics* 85:75-83.
- Kordiš D & Gubenšek F 1998. Unusual horizontal transfer for a long interspersed nuclear element between distant vertebrate classes. *Proceedings of the National Academy of Sciences USA* 95: 10704-10709.
- Kramerov DA & Vassetsky NS 2011 Origin and evolution of SINEs in eukaryote genomes. *Heredity* 107: 487-495.

Combined Bibliography

- Kress M Barra Y Seidman JG Khoury G Jay G 1984 Functional insertion of an Alu type 2 (B2 SINE) repetitive sequence in murine class I genes. *Science* 226:974-977.
- Krull M Petrusma M Makalowski W Brosius J & Schmitz J 2007 Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Research* 17:1139-1145.
- Kumar A & Bennetzen JL 1999 Plant retrotransposons. *Annual Review of Genetics* 33:479-532.
- Kunarso G Chia NY Jeyakani J Hwang C Lu X et al 2010 Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics* 42:631-634.
- Kutschera U & Niklas KJ 2004 The modern theory of biological evolution: an expanded synthesis. *Naturwissenschaften* 91: 255–276.
- Labuda D & Striker G 1989 Sequence conservation in Alu evolution. *Nucleic Acids Research* 17:2477-2491.
- Lacroix MC Guibourdenche J Frendo JL Muller F & Evain-Brion D 2002 Human placental growth hormone-a review. *Placenta* 23:S87-94.
- Lahn BT & Page DC 1999 Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature Genetics* 21:429-433.
- Lai CB Zhang Y Rogers SL & Mager DL 2009 Creation of the two isoforms of rodent NKG2D was driven by a B1 retrotransposon insertion. *Nucleic Acids Research* 37:3032-3043.
- Lai F Chen CX Carter KC & Nishikura K 1997 Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Molecular and Cell Biology* 17:2413-2424.
- Lal SK & Hannah LC 2005 Massive changes of the maize genome are caused by *Helitrons*. *Heredity* 95: 421-422.
- Laimins L Holmgren-König M & Khoury G 1986 Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus *Proceedings of the National Academy of Science USA* 83:3151-3155.

Combined Bibliography

- Lamont BB & Wiens D 2003 Are seed set and speciation rates always low among species that resprout after fire, and why? *Evolutionary Ecology* 17: 277-292.
- Lamont BB Enright NJ & He T 2011 Fitness and evolution of resprouters in relation to fire. *Plant Ecology* DOI 10.1007/s11258-011-9982-3.
- Lampe DJ Witherspoon DJ Soto-Adames FN & Robertson HM 2003 Recent horizontal transfer of mellifera subfamily *Mariner* transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. *Molecular Biology and Evolution* 20: 554-562.
- Lander ES Linton LM Birren B Nusbaum C Zody MC et al 2001 Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- Landry JR Medstrand P & Mager DL 2001 Repetitive elements in the 5' untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. *Genomics* 76:110-116.
- Landry JR Rouhi A Medstrand P & Mager DL 2002 The Opitz syndrome gene *Mid1* is transcribed from a human endogenous retroviral promoter. *Molecular Biology and Evolution* 19:1934-1942.
- Lang D Zimmer AD Rensing SA & Reski R 2008 Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends in Plant Science* 13: 542-549.
- Laperriere D Wang TT White JH & Mader S 2007 Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genomics* 8: 23.
- Larsen PA Marchan-Rivadeneira MR & Baker RJ 2010 Natural hybridisation generates mammalian lineage with species characteristics. [www.pnas.org/cgi/doi/10,1073/pnas.1000133107](http://www.pnas.org/cgi/doi/10.1073/pnas.1000133107)
- Larsson E Andersson AC & Nilsson BO 1994 Expression of an endogenous retrovirus (ERV3 HERV-R) in human reproductive and embryonic tissues - evidence for a function for envelope gene products. *Upsala Journal of Medical Sciences* 99:113-120.

Combined Bibliography

- Laten HM Havecker ER Farmer LM & Voytas DF 2003 *SIRE1*, an endogenous retrovirus family from *Glycine max*, is highly homogeneous and evolutionary young. *Molecular Biology and Evolution* 20: 1222-1230.
- Lau NC Seto AG Kim J Kuramochi-Miyagawa S Nakano T et al. 2006 Characterisation of the piRNA complex from rat testes. *Science* 313: 363-367.
- Laurin M Gussekloo SWS Marjanovic D Legendre L & Cubo J 2011 Testing gradual and speciation models of evolution in extant taxa: the example of ratites. *Journal of Evolutionary Biology*. doi: 10.1111/j.1420-9101.2011.02422.x
- Leblanc P Desset S Giorgi F Taddei AR Fausto AM et al 2000 Life cycle of an endogenous retrovirus, ZAM, in *Drosophila melanogaster*. *Journal of Virology* 74: 10658–10669.
- Lee JR Huh JW Kim DS Ha HS Ahn K et al 2009 Lineage specific evolutionary events on SFTP gene: Alu recombination-mediated deletion (ARMD), exonization, and alternative splicing events. *Gene* 435:29-35.
- Lee RW 2003 Physiological adaptations of the invasive cordgrass *Spartina anglica* to reducing sediments: rhizome metabolic gas fluxes and enhance O₂ and H₂S transport. *Marine Biology* 143: 9-15.
- Lee Y Ise T Ha D Saint Fleur A Hahn Y et al 2006 Evolution and expression of chimeric POTE-actin genes in the human genome. *Proceedings of the National Academy of Science USA* 103:17885-17890.
- Le Goff W Guerin M Chapman MJ & Thillet J 2003 A CYP7A promoter binding factor site and Alu repeat in the distal promoter region are implicated in regulation of human CETP gene expression. *Journal of Lipid Research* 44:902-910.
- Leitch AR & Leitch IJ 2012 Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist* doi: 10.1111/j.1469-8137.2012.04105.x
- Le Rouzic A & Capy P 2005 The first steps of transposable element invasion: Parasitic strategy vs. genetic drift. *Genetics* 169: 1033-1043.

Combined Bibliography

- Levanon EY Eisenberg E Yelin R Nemzer S Hallegger M et al 2004 Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nature Biotechnology* 22:1001-1005.
- Li CY Zhang Y Wang Z Zhang Cao C et al 2010 A human-specific de novo protein-coding gene associated with human brain functions. *PLoS Computational Biology* 6:e1000734.
- Li TH & Schmid CW 2001 Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* 276: 135–141.
- Li Z Mulligan MK Wang X Miles MF & Lu L Williams RW 2010b A transposon in *Comt* generates mRNA variants and causes widespread expression and behavioral differences among mice. *PLoS One* 5:e12181.
- Lindblad-Toh K Wade CM Mikkelsen TS Karlsson EK Jaffe DB et al 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438: 803–819.
- Liu B Kanazawa A Matsumura H Takahashi R Harada K & Abe J 2008 Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180: 995-1007.
- Liu G Zhao S Bailey JA Sahinalp SC Alkan C et al 2003 Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Research* 13:358-368.
- Liu WM Chu W M Choudary PV & Schmid CW 1995 Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Research* 23: 1758–1765.
- Lönnig W-E 2005 Mutation breeding, evolution, and the law of recurrent variation. *Recent Research Developments in Genetics and Breeding* 2: 45-70.
- Lönnig W-E and Saedler H 2002 Chromosomal rearrangements and transposable elements. *Annual Review of Genetics* 36: 389-410.
- Lu C Chen JJ Zhang Y Qun H Su W & Kuang H 2012 Miniature Inverted-Repeat Transposable Elements (MITEs) have been accumulated

Combined Bibliography

- through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Molecular Biology and Evolution* 29:1005-1017.
- Lunyak VV Prefontaine GG Núñez E Cramer T Ju BG et al 2007 Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 317:248-251.
- Lynch VJ Leclerc RD May G & Wagner GP 2011 Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics* doi:10.1038/ng.917
- Lyon MF 2000 LINE-1 elements and X chromosome inactivation: a function for “junk” DNA? *Proceedings of the National Academy of Science USA* 97: 6248–6249.
- Macfarlane TD Hopper SD Purdie RW George AS & Patrick SJ 1987 Haemodoraceae; *Flora of Australia* Volume 45 Australian Government Publishing Service Canberra.
- Madlung A Masuelli R Watson B Reynolds S Davison S 2002 Remodelling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* 129: 733-746.
- Maichele AJ Farwell NJ & Chamberlain JS 1993 A B2 repeat insertion generates alternate structures of the mouse muscle gamma-phosphorylase kinase gene. *Genomics* 16:139-49.
- Maksakova IA Romanish MT Gagnier L Dunn CA van de Lagemaat LN & Mager DL 2006 Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genetics* 2: e2.
- Mallon AM Wilming L Weekes J Gilbert JG Ashurst J et al 2004 Organization and evolution of a gene-rich region of the mouse genome: a 12.7-Mb region deleted in the Del(13)Svea36H mouse. *Genome Research* 14: 1888–1901.
- Mangeny M Renard M Schlecht-Louf G Bouallaga I Heidmann O et al 2007. Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proceedings of the National Academy of Science USA* 104:20534-20539.

Combined Bibliography

- Manning G & Scheeff E 2010 How the vertebrates were made: selective pruning of a double-duplicated genome. *BMC Biology* 8:144.
- Marco A & Marin I 2005 Retrovirus-like elements in plants. *Recent Research and Development in Plant Science*. 3: ISBN: 81-7736-245-3 .
- Margulis L 1991 in *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis* (Margulis L & Fester R eds) p.2 MIT Press.
- Margulis L Chapman MJ 1998 Endosymbioses: cyclical and permanent in evolution. *Trends in Microbiology* 6:342-346.
- Marques AC Dupanloup I Vinckenbosch N Reymond A & Kaessmann H 2005 Emergence of young human genes after a burst of retroposition in primates. *PLoS Biology* 3: e357.
- Martin XJ Schwartz K & Boheler KR 1995 Characterization of a novel transcript of retroviral origin expressed in rat heart and liver. *Journal of Molecular and Cellular Cardiology* 27:589-597.
- Maruyama K & Hart DL 1991 Evidence for interspecific transfer of the transposable element mariner between *Drosophila* and *Zaprionus*. *Journal of Molecular Biology* 33: 514-524.
- Masterson J 1994 Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264: 421-424.
- Mätlik K Redik K & Speek M 2006 L1 antisense promoter drives tissue-specific transcription of human genes. *Journal of Biomedicine and Biotechnology* 2006:71753.
- Mattick JS 2011 The central role of RNA in human development and cognition. *FEBS Letters* 585: 1600-1616.
- Mattick JS & Mehler MF 2008 RNA editing, DNA recoding and the evolution of human cognition. *Trends in Neuroscience* 31:227-233.
- Mattila TM & Bokma F 2008 Extant mammal body masses suggest punctuated equilibrium. *Proceedings of the Royal Society* 2008 275: 2195-2199. doi:10.1098/rspb.2008.0354
- Matzke MA Mette MF Aufsatz W Jakowitsch J & Matzke AJ 1999 Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. *Genetica* 107: 271–287.

Combined Bibliography

- Mayer J & Meese E 2005 Human endogenous retroviruses in the primate lineage and their influence on host genomes. *Cytogenetics and Genome Research* 110:448-456.
- McClintock B 1950 The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Science USA* 36: 344–355.
- McClintock B 1953 Induction of instability at selected loci in maize. *Genetics* 38: 579-599.
- McClintock B 1956 Controlling elements and the gene. *Cold Spring Harbour Symposium Quantitative Biology* 21: 197–216.
- McClintock B 1984 The significance of responses of the genome to challenge. *Science* 226: 792–801.
- McClintock B 1987 *The Discovery and Characterisation of Transposable Elements*. Garland New York.
- McDonald JF 1995 Transposable elements: possible catalysts of organismic evolution. *Trends in Ecology and Evolution* 10: 123-126.
- McFadden J & Knowles G 1997 Escape from evolutionary stasis by transposon-mediated deleterious mutations. *Journal of Theoretical Biology* 186: 441–447.
- McHaffie GS & Ralston SH 1995 Origin of a negative calcium response element in an ALU-repeat: implications for regulation of gene expression by extracellular calcium. *Bone* 17:11-14.
- McLysaght A Hokamp K Wolfe KH 2002 Extensive genomic duplication during early chordate evolution. *Nature Genetics* 31:200-204.
- McShea & Brandon 2010 *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems*. The University of Chicago Press: Chicago.
- Medstrand P Landry JR & Mager D 2001: Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *Journal of Biological Chemistry* 276:1896-1903.
- Mi S Lee X Li X Veldman GM Finnerty H et al JM 2000 Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785-789.

Combined Bibliography

- Michalak P 2010 An eruption of mobile elements in genomes of hybrid sunflowers. *Heredity* 104: 329-330.
- Michaux J Reyes A & Catzeflis 2001 Evolutionary history of the most speciose mammals: molecular phylogeny of muroid rodents. *Molecular Biology and Evolution* 18: 2017-2031.
- Michel D Chatelain G Mauduit C Benahmed M & Brun G 1997 Recent evolutionary acquisition of alternative pre-mRNA splicing and 3' processing regulations induced by intronic B2 SINE insertion. *Nucleic Acids Research* 25:3228-3234.
- Miguel C Simões M Oliveira MM & Rocheta M 2008 Envelope-like retrotransposons in the plant kingdom: evidence of their presence in gymnosperms (*Pinus pinaster*). *Journal of Molecular Evolution* 67: 517-525. DOI 10.1007/s00239-008-9168-3.
- Mikkelsen TS Hillier LW Eichler EE Zody MC Jaffe DB et al 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69–87.
- Mikkelsen TS Wakefield MJ Aken B Amemiya CT Chang JL et al 2007 Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447: 167–177.
- Miller WJ McDonald JF Nouaud D & Anxolabéhère D 1999 Molecular domestication – more than a sporadic episode in evolution. *Genetica* 107: 197-207.
- Mills RE Bennett EA Iskow RC Luttig CT Tsui C et al 2006 Recently mobilized transposons in the human and chimpanzee genomes. *American Journal of Human Genetics* 78: 671–679.
- Mills RE Bennett EA Iskow RC & Devine SE 2007 Which transposable elements are active in the human genome? *Trends in Genetics* 23: 183–191.
- Mola G Vela E Fernández-Figueras MT Isamat M & Muñoz-Mármol AM 2007 Exonization of Alu-generated splice variants in the survivin gene of human and non-human primates. *Journal of Molecular Biology* 366:1055-1063.

Combined Bibliography

- Monk M 1995 Epigenetic programming of differential gene expression in development and evolution. *Developmental Genetics* 17:188-197.
- Moran JV DeBerardinis RJ & Kazazian HH Jr 1999 Exon shuffling by L1 retrotransposition. *Science* 283: 1530–1534.
- Morgan HD Sutherland HG Martin DI & Whitelaw E 1999 Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics* 23:314-318.
- Morgan HD Santos F Green K Dean W & Reik W 2005 Epigenetic reprogramming in mammals. *Human Molecular Genetics* 14: R47–R58.
- Morgante M Brunner S Pea G Fengler K & Zuccolo A et al 2005 Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genetics* 17: 997-1002. doi:10.1038/ng1615.
- Morse AM Peterson DG Islam-Faridi MN Smith KE Magbanua Z et al 2009 Evolution of genome size and complexity in *Pinus*. *PLoS ONE* 4: e4332. doi:10.1371/journal.pone.0004332
- Mouse Genome Sequencing Consortium 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.
- Muotri AR Marchetto MC Coufal NG & Gage FH 2007 The necessary junk: new functions for transposable elements. *Human Molecular Genetics* 16: R159–R167.
- Nakayama K & Ishida T 2006 Alu-mediated 100-kb deletion in the primate genome: the loss of the agouti signalling protein gene in the lesser apes. *Genome Research* 16:485-490.
- Nam K & Ellegren H 2012 Recombination drives vertebrate genome contraction, *PLoS Genetics* 8: e1002680 1-9.
- Neafsey DE Blumenstiel JP & Hartel DL 2004 Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. *Molecular Biology and Evolution* 21: 2310-2318.
- Nei M 2005 Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution* 22: 2318-2342.

Combined Bibliography

- Nei M 2007 The new mutation theory of phenotypic evolution. Proceedings of the National Academy of Sciences USA 104: 12235-12242.
- Nekrutenko A & Li WH 2001 Transposable elements are found in a large number of human protein-coding genes. Trends in Genetics 17: 619–621.
- Ngezahayo F Xu C Wang H Jiang L Pang Y and Liu B 2009 Tissue culture induced transpositional activity of *mPing* is correlated with cytosine methylation in rice. BMC Plant Biology 9:91 doi:10.1186/1471-2229-9-91.
- Nhim RP Lindsey JS & Wilkinson MF 1997 A processed homeobox gene expressed in a stage-, tissue- and region-specific manner in epididymis. Gene 185:27127-6.
- Nigumann P Redik K Mätlik K & Speek M 2002 Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79:628-634.
- Nishihara H Smit AF & Okada N 2006 Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Research 16: 864–874.
- Nnuzhdin SV 1999 Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica 107: 128-137.
- Noonan JP Grimwood J Danke J Schmutz J Dickson M et al 2004 Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. Genome Research 14: 2397-2405.
- Norris J Fan D Aleman C Marks JR Futreal PA et al 1995 Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. Journal of Biological Chemistry 270:22777-22782.
- Novick PA Basta H Floumanhaft M McClure MA & Boissinot S 2009 The evolutionary dynamics of autonomous Non-LTR Retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. Molecular Biology and Evolution 26: 1811-1822

Combined Bibliography

- Nowak RM 1994 Walkers Bats of the World. The John Hopkins University Press Baltimore.
- O'Donnell KA & Burns K H 2010 Mobilising diversity: transposable element insertions in genetic variation and disease. *Mobile DNA* 1:21.
- Ogiwara I Miya M Oshima K & Okada N 2002 V-SINEs: A new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Research* 12: 316-324.
- Ohno S 1970 Evolution by Gene Duplication. New York Springer-Verlag.
- Okamura K & Lai EC 2008 Endogenous small interfering RNAs in animals. *Nat. Rev. Molecular Cell Biology* 9: 673-678.
- Olivares M López MC Garcia-Pérez JL Briones P Pulgar M & Thomas MC 2003 The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. *Biochimica et Biophysica Acta* 1626: 25–32.
- Oliver KR & Greene WK 2009a Transposable elements: powerful facilitators of evolution. *BioEssays* 31: 703–714.
- Oliver KR & Greene WK 2009b The Genomic Drive Hypothesis and punctuated evolutionary taxonations, or radiations. *Journal of the Royal Society of Western Australia* 92: 447-451.
- Oliver KR & Greene WR 2011 Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. *Mobile DNA* 2:8.
- Olson M V 1999 Molecular evolution '99 When Less is More: Gene Loss as an Engine of Evolutionary Change. *American Journal of Human Genetics* 64:18-23.
- O'Neil RJW O'Neil MJ and Graves JAM 1998 Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393: 68-72.
- O'Neil RJW O'Neil MJ & Graves JAM 2002 Corrigendum Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 420: 106.
- Ohnishi Y Totoki Y Toyoda A Watanabe T Yamamoto Y et al 2012 Active role of small non-coding RNAs derived from SINE/B1 retrotransposon

Combined Bibliography

- during early mouse development. *Molecular Biology Reporter* 39:903-909.
- Orgel LE & Crick FH 1980 Selfish DNA: the ultimate parasite. *Nature* 284: 604–607.
- Orr HA Masly JP & Presgraves DC 2004 Speciation genes. *Current Opinion in Genetics and Development* 4: 675–679.
- Oshima K Hattori M Yada T Gojobori T Sakaki Y & Okada N 2003 Whole-genome screening indicates a possible burst of formation of processes pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biology* 4: R74 (<http://genomebiology.com/2003/4/11/R74>).
- Ostertag EM Goodier JL Zhang Y & Kazazian HH Jr 2003 SVA elements are nonautonomous retrotransposons that cause disease in humans. *American Journal of Human Genetics* 73:1444-1451.
- Oxford Reference Online: The Encyclopedia of Mammals <http://0=www.oxfordreference.com.prospero.murdoch.edu.au>. 23/3/2010.
- Pace JK II & Feschotte C 2007 The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Research* 17: 422–432.
- Pace JK II Gilbert C Clark MS & Feschotte C 2008 Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proceedings of the National Academy of Sciences of the USA* 105: 17023–17028.
- Palmer SA Clapham AJ Rose P Freitas FO Owen BD et al 2012 Archaeogenomic evidence of punctuated genome evolution in *Gossypium*. *MBE Advance Access* published February 14, 2012.
- Panaud O 2009 The molecular bases of cereal domestication and the history of rice. *Comptes Rendus Biologies* 332: 267-272.
- Pardue M L & DeBaryshe P G 2003 Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annual Revue of Genetics* 37: 485-511.

Combined Bibliography

- Parisod C Salmon A Zerjal T Tenailon M Grandbastien M-A 2009 Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist* 184: 1003-1015.
- Parisod C Alix K Just J Petit M Sarilar V et al 2010 Impact of transposable elements on the organisation and function of allopolyploid genomes. *New Phytologist* 186: 37-45.
- Parris GE 2009 A hypothetical master development program for multi cellular organisms: ontogeny and phylogeny. *Bioscience Hypotheses* 2: 3–12.
- Parrott AM & Mathews MB 2009 snaR genes: recent descendants of Alu involved in the evolution of chorionic gonadotropins. *Cold Spring Harbor Symposium in Quantitative Biology* 74:363-373.
- Pasyukova EG Nuzhdin SV Morozova TV & Mackay TFC 2004 Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. *Journal of Heredity* 95: 284-290.
- Paterson AH Bowers JE Bruggmann R Dubchak I Grimwood J et al 2009 The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551-6.
- Paun O Bateman RM Fay MF Hedrén M Civeyrel L & Chase MW 2010 Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactyloctenium*: Orchidaceae). *Molecular Biology and Evolution* 27:2465-2473 doi:10.1093/molbev/ mmsq150
- Pausas J G & Keeley 2009 A burning story: the role of fire in the history of life. *BioScience* 59: 593-601.
- Pearce SR 2007 SIRE-1, a putative plant retrovirus is closely related to a legume TY1-Copia retrotransposon family. *Cellular and Molecular Biology Letters* 12: 120-126. DOI: 10.2478/s11658-006-0053-z
- Peaston AE Evsikov AV Graber JH de Vries WN Holbrook AE et al 2004 Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Developmental Cell* 7: 597–606.
- Perincheri S Dingle RW Peterson ML & Spear BT 2005 Hereditary persistence of alpha-fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in the *Zhx2* gene.

Combined Bibliography

- Proceedings of the National Academy of Science USA 2005 102:396-401.
- Perry GH Dominy NJ Claw KG Lee AS Fiegler H et al 2007 Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* 2007 39:1256-1260.
- Persson K Holm I & Heby O 1995 Cloning and sequencing of an intronless mouse S-adenosylmethionine decarboxylase gene coding for a functional enzyme strongly expressed in the liver. *Journal of Biological Chemistry* 270:5642-5648.
- Petit M Guidat C Daniel J Montoriol E Lim KY et al 2010 Mobilisation of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist* 186: 135-147.
- Phillips RL Kaeppler SM & Olhoft P 1994 Genetic instability of plant tissue cultures: breakdown of normal controls. *Proceedings of the National Academy of Science USA* 91: 5222-5226.
- Picard G 1976 Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of I factor. *Genetics* 83: 107– 123.
- Piedrafita FJ Molander RB Vansant G Orlova EA Pfahl M & Reynolds WF 1996 An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. *Journal of Biological Chemistry* 271:14412-14420.
- Piedrafita FJ Molander RB Vansant G Orlova EA Pfahl M & Reynolds WF 1999 An Alu element in the myeloperoxidase promoter contains a composite tropoelastin gene during primate evolution. *Journal of Molecular Evolution* 49:664-671. doi:10.1186/1759-8753-2-8.
- Pigliuchi M & Kaplan J 2006 *Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology*. The University of Chicago Press Ltd London.
- Piriyapongsa J & Jordan I K 2008 Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 14: 814-821.
- Piriyapongsa J Mariño-Ramírez L & Jordan I K 2007a Origin and evolution of human microRNAs from transposable elements. *Genetics* 176: 1323– 1337.

Combined Bibliography

- Piriyapongsa J Polavarapu N Borodovsky M & McDonald J 2007b Exonization of the LTR transposable elements in human genome. *BMC Genomics* 8: 291
- Piskurek O & Okada N 2007 Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proceedings of the National Academy of Sciences USA* 104: 12046-12051.
- Piskurek O Nikaido M Boeadi MB & Okada N 2003 Unique mammalian tRNA derived repetitive elements in dermopterans: the t-SINE family and its transposition through multiple sources. *Molecular Biology and Evolution* 20: 1659-1668.
- Platt RN & Ray DA 2012 A non-LTR retroelement extinction in *Spermophilus tridecemlineatus*. *Gene* doi:10.1016/j.gene. 2012.03.051
- Polak P & Domany E 2006 Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7:133.
- Pollard KS Salama SR Lambert N Lambot MA Coppens S Pedersen JS et al 2006 An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167-172.
- Pontius JU Mullikin JC Smith DR Agencourt Sequencing Team Lindblad-Toh K. et al 2007 Initial sequence and comparative analysis of the cat genome. *Genome Research* 17: 1675–1689.
- Pough FH Janis CM & Heiser JB 2005 In: *Vetebtrate Life*. Benjamin Cummings San Francisco.
- Poulter RTM Goodwin TJD & Butler MI 2003 Vertebrate helitrons and other novel *Helitrons*. *Gene* 313: 201-212.
- Pritham EJ & Feschotte C 2007 Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proceedings of the National Academy of Science USA* 104:1895–1900.
- Prokesch A Bogner-Strauss JG Hackl H Rieder D Neuhold C et al 2011 Arxes: retrotransposed genes required for adipogenesis. *Nucleic Acids Research* 39:3224-3239.

Combined Bibliography

- Prudhomme S Bonnaud B & Mallet F 2005 Endogenous retroviruses and animal reproduction. *Cytogenetics and Genome Research* 110:353-364.
- Putnam NH Butts T Ferrier DE Furlong RF Hellsten U et al 2008 The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453: 1064–1071.
- Quentin Y 1992 Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Research* 20:487-493.
- Rai HS O'Brien HE Reeves PA Olmsread RG & Graham SW 2003 Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* 29: 350-359.
- Rankin DJ Bichsel M & Wagner A 2010 Mobile DNA can drive lineage extinction in prokaryotic populations. *Journal of Evolutionary Biology* 23: 2422-2431. doi:10.1111/j.1420-9101.2010.02106.x
- Rat Genome Sequencing Project Consortium 2004 Genome sequence of the brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493-521.
- Raven PH Evert RF & Eichhorn SE 1992 *Biology of Plants* 5th edition Worth Publishers New York.
- Rawn SM & Cross JC 2008 The evolution, regulation, and function of placenta-specific genes. *Annual Review of Cell and Developmental Biology* 24: 159-181.
- Ray DA Pagan HJ Thompson ML & Stevens RD 2007 Bats with hATs: evidence for recent DNA transposon activity in genus *Myotis*. *Molecular Biology & Evolution* 24: 632–639.
- Ray DA Feschotte C Pagan HJ Smith JD Pritham EJ et al 2008 Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Research* 18: 717–728.
- Rebollo R Horard B Hubert B & Vieira C 2010 Jumping genes and epigenetics: towards new species. *Gene* 454: 1-7.

Combined Bibliography

- Reig OA 1986 Diversity patterns and differentiation in high Andean rodents. In: High Altitude Tropical Biogeography (Vuilleumier F & Monasterio M eds) Oxford University Press London. 404-440.
- Reik W 2007 Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447: 425–432.
- Reinton N Haugen TB Orstavik S Skålhegg BS Hansson V et al 1998 The gene encoding the C gamma catalytic subunit of cAMP- dependent protein kinase is a transcribed retroposon. *Genomics* 49:290-297.
- Rensing S A Lang D Zimmer AD Terry A Salamov A et al 2008 The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64-69.
- Richards EJ 2006 Inherited epigenetic variation-revisiting soft inheritance. *Nature Reviews Genetics* 7: 395-402.
- Richards TA & Dacks JB 2006 Evolution of filamentous plant pathogens: Gene exchange across eukaryote kingdoms. *Current Biology* 16: 1857-1864.
- Ridley M 1999 *Genome: The autobiography of a species in 23 chapters*. Fourth Estate Limited London.
- Ridley M 2004 *Evolution* (3rd ed) Blackwell Publishing Oxford UK.
- Rieseberg LH 2001 Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* 16: 351–358.
- Rieseberg LH & Wendel JF 2004 Plant speciation – rise of the poor cousins. *New Phytologist* 161: 3-8.
- Rio DC 1990 Molecular mechanisms regulating *Drosophila* P element transposition. *Annual Review of Genetics* 24: 543–578.
- Robertson HM & Lampe DJ 1995 Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. *Molecular Biology and Evolution* 12: 850-862.
- Rodriguez IR Mazuruk K Schoen TJ & Chader GJ 1994 Structural analysis of the human hydroxyindole-O-methyltransferase gene. Presence of two distinct promoters. *Journal of Biological Chemistry* 269:31969-31977.

Combined Bibliography

- Rodriguez J Vives L Jordà M Morales C Muñoz M et al 2008 Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. *Nucleic Acids Research* 2008. 36: 770–784.
- Romanish MT Lock WM van de Lagemaat LN Dunn CA & Mager DL 2007 Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. *PLoS Genetics* 3: e10.
- Romanish MT Nakamura H Lai CB Wang Y & Mager DL 2009 A novel protein isoform of the multicopy human NAIP gene derives from intragenic Alu SINE promoters. *PLoS One* 4:e5761.
- Rose MR & Oakley 2007 The new biology: beyond the Modern Synthesis. *Biology Direct* 2:30 doi:10.1186/1745-6150-2-30
- Rudin CM & Thompson CB 2001 Transcriptional activation of short interspersed elements by DNA-damaging agents. *Genes Chromosomes Cancer* 30: 64–71.
- Ryan F 2002 Darwin's Blind Spot - Evolution Beyond Natural Selection. Houghton Mifflin Company, New York.
- Ryan FP 2006 Genomic creativity and natural selection: a modern synthesis. *Biological Journal of the Linnean Society* 88: 655-672.
- Ryan FP 2007 Viruses as symbionts. *Symbiosis* 44: 11-21.
- Ryan FP 2009 An alternative approach to medical genetics based on modern evolutionary biology. Part 1: mutation and symbiogenesis. *Journal of the Royal Society of Medicine* 102: 272-277.
- Sakai H Tanaka T & Itoh T 2007 Birth and death of genes promoted by transposable elements in *Oryza sativa*. *Gene* 392: 59-63. Plus supplementary materials.
- Saksela K & Baltimore D 1993 Negative regulation of immunoglobulin kappa light-chain gene transcription by a short sequence homologous to the murine B1 repetitive element. *Molecular and Cell Biology* 13:3698-3705.
- Salmon A Ainouche ML Wendel JF 2005 Genetic and epigenetic consequences of recent hybridisation and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* 14: 1163-1175.

Combined Bibliography

- Schaack TJ & Pritham E 2010 Pervasive horizontal transfer of rolling circle transposons among animals. *Genome Biology and Evolution* 2: 656-664.
- Schaack S Gilbert C & Feschotte C 2010 Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology and Evolution* 25: 537-546.
- Schatz DG 2004 Antigen receptor genes and the evolution of a recombinase. *Seminars in Immunology* 16: 245–256.
- Schlueter JA Dixon P Grainger C Grant D Clark L et al. 2004 Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47: 868-878.
- Schlueter JA Scheffler BE & Shoemaker RC 2006 Sequence conservation of homeologous bacterial artificial chromosomes and transcription of homoeologous genes in soybean (*Glycine max* L. Merr). *Genetics* 174: 1017-1028.
- Schlueter JA Vasylenko-Saunders IF Deshpande S Yi J Siegfried M et al 2007 The FAD2 gene family of soybean: insights into the structural and functional divergence of a paleopolyploid genome. *Crop Science* 47: S-14-S-26.
- Schmid CW 1998 Does SINE evolution preclude Alu function? *Nucleic Acids Research* 26:4541-4550.
- Schmid CW 1991 Human Alu subfamilies and their methylation revealed by blot hybridization. *Nucleic Acids Research* 19: 5613–5617.
- Schmidt CW & Jelinek WR 1982 The Alu family of dispersed repetitive sequences. *Science* 216: 1065-1070.
- Schmidt JM Good RT Appleton B Sherrard J Raymant GC et al 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* Locus. *PLoS Genetics* 6: e1000998. doi:10.1371/journal.pgen.1000998 .
- Schnable PS Ware D Fulton RS Stein JC Wei F et al 2009 The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112-5.

Combined Bibliography

- Schneider D Duperchy E Depeyrot J Coursange E Lenski RE et al 2002 Genomic comparisons among *Escherichia coli* strains B, K-12, and O157. H7 using IS elements as molecular markers. *BMC Microbiol* 2: 18.
- Schulte AM Lai S Kurtz A Czubayko F Riegel AT & Wellstein A 1996 Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proceedings of the National Academy of Science USA* 93:14759-14764.
- Schulz WA Steinhoff C & Florl AR 2006 Methylation of endogenous human retroelements in health and disease. *Current Topics in Microbiology and Immunology* 310: 211–250.
- Schwartz A Chan D C Brown LG Alagappan R Pettay D et al 1998 Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Human Molecular Genetics* 7: 1–11.
- Scofield MA Xiong W Haas MJ Zeng Y & Cox GS 2000 Sequence analysis of the human glycoprotein hormone alpha-subunit gene 5'-flanking DNA and identification of a potential regulatory element as an Alu repetitive sequence. *Biochimica et Biophysica Acta* 1493:302-318.
- Scotland RW & Wortley AH 2003 How many species of seed plants are there? *Taxon* 52: 101-104.
- Segall SK Nackley AG Diatchenko L Lariviere WR Lu X et al 2010 Comt1 genotype and expression predicts anxiety and nociceptive sensitivity in inbred strains of mice. *Genes Brain Behaviour* 8:933-946.
- Sela N Mersch B Hotz-Wagenblatt A & Ast G 2010 Characteristics of transposable element exonization within human and mouse. *PLoS One* 5:e10907.
- Sen SK Han K Wang J Lee J Wang H et al 2006 Human genomic deletions mediated by recombination between Alu elements. *American Journal of Human Genetics* 79: 41–53.
- Shaked H Kashkush K Ozkan H Feldman M & Levy AA 2001 Sequence elimination and cytosine methylation are rapid and reproducible

Combined Bibliography

- responses of the genome to wide hybridisation and allopolyploidy in wheat. *The Plant Cell* 13: 1749-1759.
- Shapiro JA 1999 Transposable elements as the key to a 21st century view of evolution. *Genetica* 107: 171-179.
- Shapiro JA 2010 Mobile DNA and evolution in the 21st century. *Mobile DNA* 1:4 .
- Shapiro JA & von Sternberg R 2005 Why repetitive DNA is essential to genome function. *Biological Reviews* 80: 1-24.
- Sharon G Segal D Ringo JM Hefetz A Zilber-Rosenberg I & Rosenberg E 2010 Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Science USA* 107: 20051-20056.
- Shen S Lin L Cai JJ Jiang P Kenkel EJ et al 2011 Widespread establishment and regulatory impact of Alu exons in human genes. *Proceedings of the National Academy of Science USA* 108:2837-2842.
- Shephard EA Chandan P Stevanovic-Walker M Edwards M & Phillips IR 2007 Alternative promoters and repetitive DNA elements define the species- dependent tissue-specific expression of the FMO1 genes of human and mouse. *Biochemistry Journal* 406:491-499.
- Shoemaker RC Schlueter JA & Doyle JF 2006 Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinion in Plant Biology* 9: 104-109.
- Shulaev V Sargent DJ Crowhurst RN Mockler TC Folkerts et al 2011 The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43:109-16.
- Silva JC Loreto EL & Clark JB 2004 Factors that affect the horizontal transfer of transposable elements. *Current Issues in Molecular Biology* 6: 57-72.
- Simmons GM 1992 Horizontal transfer of hobo transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. *Molecular Biology and Evolution* 9: 1050-1060.
- Simmons NB 2005 Evolution. An Eocene big bang for bats. *Science* 307: 527–528.

Combined Bibliography

- Simons C Pheasant M Makunin IV & Mattick JS 2006 Transposon-free regions in mammalian genomes. *Genome Research* 16: 164–172.
- Sin HS Huh JW Kim DS Kang DW Min DS et al 2006 Transcriptional control of the HERV-H LTR element of the GSDML gene in human tissues and cancer cells. *Archives of Virology* 151:1985-1994.
- Singer SS Männel DN Hehlhans T Brosius J & Schmitz J 2004 From “junk” to gene: curriculum vitae of a primate receptor isoform gene. *Journal of Molecular Biology* 341:883-886.
- Singleton G Dickman CR Stoddart DM & Stockley P 2007 *The Encyclopaedia of Mammals*. Ed. David McDonald. Oxford University Press. Oxford Reference Online 2010.
- Sinzelle L Kapitonov VV Grzela DP Jursch T Jurka J et al 2008 Transposition of a reconstructed *Harbinger* element in human cells and functional homology with two transposon-derived cellular genes. *Proceedings of the National Academy of Science USA* 105: 4715-4720.
- Sinzelle L Izsvák Z & Ivics Z 2009 Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cellular and Molecular Life Sciences* 66: 1073-1093.
- Smalheiser NR & Torvik VI 2006 Alu elements within human mRNAs are probable microRNA targets. *Trends in Genetics* 22: 532–536.
- Smit AF 1993 Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Research* 21: 1863–1872.
- Smith JJ Sumiyama K Amemiya CT 2012 A living fossil in the genome of a living fossil: Harbinger transposons in the coelacanth genome. *Molecular Biology and Evolution* 29: 985-993.
- Soares MB Schon E Henderson A Karathanasis SK Cate R et al 1985 RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Molecular and Cellular Biology* 5:2090-2103.
- Soltis DE Soltis PS & Tate JA 2003 Advances in the study of polyploidy since *Plant Speciation*. *New Phytologist* 161: 173-191.
- Soltis DE Bell CD Kim S & Soltis PS 2008 Origin and early evolution of the angiosperms. *Annals of the New York Academy of Sciences* 1133: 3-25.

Combined Bibliography

- Song SU Kurkulos M Boeke JD & Corces VG 1997 Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of *Drosophila*. *Development* 124: 2789–2798.
- Sorek R Ast G & Graur D 2002 Alu-containing exons are alternatively spliced. *Genome Research* 12: 1060–1067.
- Stadleman B Lin LK Kunz TH & Ruedi M 2007 Molecular phylogeny of New World *Myotis* (*Chiroptera*, *Vespertilionidae*) inferred from mitochondrial and nuclear DNA genes. *Molecular Phylogenetics & Evolution* 43: 32–48.
- Stanley SM 1981 *The New Evolutionary Timetable*. Basic Books Inc New York.
- Stanyon R Yang F Cavagna P O'Brien PCM Bagga M et al 1999 Reciprocal chromosome painting shows that genomic rearrangement between rat and mouse proceeds ten times faster than between humans and cats. *Cytogenetic and Genome Research* 84: 150-155.
- Stavenhagen JB & Robins DM 1988 An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 55:247-254.
- Stebbins GL 1971 *Chromosomal Evolution in Higher Plants*. London: Edward Arnold.
- Steel G & Lutz EM 2007 Characterisation of the mouse vasoactive intestinal peptide receptor type 2 gene, *Vipr2*, and identification of a polymorphic LINE-1-like sequence that confers altered promoter activity. *Journal of Neuroendocrinology* 19:14-25.
- Steele EJ 2009 Lamarck and immunity: somatic and germline evolution of antibody genes. *Journal of the Royal Society of Western Australia* 92: 437-446.
- Steele EJ Lindley RA & Blanden RV 1998 *Lamarck's Signature: How Retrogenes are Changing Darwin's Natural Selection Paradigm*. Allen & Unwin St Leonards.
- Sternberg R von & Shapiro JA 2005 How repeated elements format genome function. *Cytogenetic and Genome Research* 110: 108-116.

Combined Bibliography

- Stokes HW & Gillings MR 2011 Gene flow, mobile genetic elements and the recruitment of antibiotic genes into Gram-negative pathogens. Federation of European Microbiological Societies Microbiological Review 35: 790-819.
- Stoltzfus A & Yampolsky LY 2009 Climbing mount probable: mutation as a cause of nonrandomness in evolution. Journal of Hereditary 100:637-647.
- Sugino H Toyama T Taguchi Y Esumi S Miyazaki M & Yagi T 2004 Negative and positive effects of an IAP-LTR on nearby Pcdalpha gene expression in the central nervous system and neuroblastoma cell lines. Gene 337:91-103.
- Sugiyama A Noguchi K Kitanaka C Katou N Tashiro F et al 1999 Molecular cloning and chromosomal mapping of mouse intronless myc gene acting as a potent apoptosis inducer. Gene 226:273-283.
- Suzuki S Ono R Narita T Pask AJ Shaw G et al 2007 Retro- transposon silencing by DNA methylation can drive mammalian genomic imprinting. PLoS Genetics 3: e55.
- Szabó Z Levi-Minzi SA Christiano AM Struminger C Stoneking M et al 1999 Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution. Journal of Molecular Evolution 49:664-671.
- Tarlinton R E Meers J & Young P R 2006 Retroviral invasion of the koala genome. Nature 442: 79-81.
- Tate JA Soltis DE & Soltis PS 2005 Polyploidy in plants. Chapter 7, *In: The Evolution of the Genome*. Gregory TR editor. Elsevier Academic Press San Diego USA.
- Taylor PD 2004 Extinction in the fossil record. *In: Extinctions in the History of Life*. PD Taylor ed Cambridge University Press New York.
- Teeling EC Springer MS Madsen O Bates P O'Brien SJ & Murphy WJ 2005 A molecular phylogeny for bats illuminates biogeography and the fossil record. Science 307: 580-584.

Combined Bibliography

- Than NG Romero R Goodman M Weckle A Xing J et al 2009 A primate subfamily of galectins expressed at the maternal-fetal interface that promote immune cell death. *Proceedings of the National Academy of Science USA* 106:9731-9736.
- Thomas J Schaack S & Pritham E 2010 Pervasive horizontal transfer of rolling circle transposons among animals. *Genome Biology and Evolution* 2: 656-664.
- Thomas J Sorourian M Ray D Baker RJ & Pritham EJ 2011 The limited distribution of *Helitrons* to vesper bats supports horizontal transfer. *Gene* 474: 52-58.
- Thomson SJ Goh FG Banks H Krausgruber T Kotenko SV et al 2009 The role of transposable elements in the regulation of IFN-lambda1 gene expression. *Proceedings of the National Academy of Science USA* 106:11564-11569.
- Ting CN Rosenberg MP Snow CM Samuelson LC & Meisler MH 1992 Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes and Development* 6:1457-1465.
- Togu H & Sota T 2009 Do arms races punctuate evolutionary stasis? Unified insights from phylogeny, phylogeography and microevolutionary processes. *Molecular Ecology* 18: 3940-3954.
- Toll-Riera M Bosch N Bellora N Castelo R Armengol L et al 2009 Origin of primate orphan genes: a comparative genomics approach. *Molecular Biology and Evolution* 26: 603-612.
- Trujillo MA Sakagashira M & Eberhardt NL 2006 The human growth hormone gene contains a silencer embedded within an Alu repeat in the 3'-flanking region. *Molecular Endocrinology* 20:2559-2575.
- Tsiantis M 2011 A transposon in the *tb1* drove maize domestication. *Nature Genetics* 43: 1048-1050
- Ullu E & Tschudi C 1984 Alu sequences are processed 7SL RNA genes. *Nature* 312:171-172.
- Vaknin K Goren A & Ast G 2009 TEs or not TEs? That is the evolutionary question. *Journal of Biology* 8: 83.

Combined Bibliography

- van de Lagemaat LN Landry JR Mager DL & Medstrand P 2003 Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* 19: 530–536.
- van den Hurk JA Meij IC Seleme MC Kano H Nikopoulos K et al 2007 L1 retrotransposition can occur early in human embryonic development. *Human Molecular Genetics* 16: 1587–1592.
- Van-Regenmortel MHV Fauquet CM Bishop DHL Carsten EB Estes MK et al 2000 *Virus taxonomy : seventh report of the International Committee on Taxonomy of Viruses*. Academic Press San Diego CA.
- Vansant G & Reynolds WF 1995 The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proceedings of the National Academy of Science USA* 92:8229-8233.
- Velasco R Zharkikh A Affourtit J Dhingra A Cestaro A et al 2010 The genome of the domesticated apple (*Malus x domestica* Borkh.) *Nature Genetics* 42:833-9.
- Venditti C Meade A & Pagel M 2010 Phylogenies reveal new interpretation of speciation and the Red Queen. *Nature* 463: 349-352.
- Veyrunes F Dobigny G Yang F O'Brien PCM Catalan J et al 2006 Phylogenomics of the genus *Mus* (Rodentia; Muridae): extensive genome repatterning is not restricted to the house mouse. *Proceedings of the Royal Society B*. 273: 2925-2934. doi:10.1098/rspb.2006.3670
- Vicient CM Kalendar R & Schulman AH 2001 Envelope-Class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Research* 11: 2041-2049.
- Villarreal LP 1997 On viruses, sex, and motherhood. *Journal of Virology* 71: 859-865
- Villarreal LP 2003 Can viruses make us human? *Proceedings of the American Philosophical Society* 148: 296-323.
- Villarreal LP 2005 *Viruses and the Evolution of Life*. ASM Press Washington DC.
- Villarreal LP 2009 *Origin of Group Identity: Viruses, Addiction and Cooperation*. Springer springer.com.

Combined Bibliography

- Villarreal LP & Witzany G 2010 viruses are essential agents within the roots and the stem of the tree of life. *Journal of Theoretical Biology* 262: 698-710.
- Vinckenbosch N Dupanloup I & Kaessmann H 2006 Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Science USA* 103:3220-3225.
- Vinogradov AE 2003 Selfish DNA is maladaptive: evidence from the red plant list. *Trends in Genetics* 19: 609-614.
- Vitte C & Bennetzen J L 2006 Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proceedings of the National Academy of Science USA* 103: 17638-17543.
- Volff JN 2006 Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* 28: 913–922.
- Von Sternberg R & Shapiro JA 2005 How repeated retroelements format genome function. *Cytogenetics and Genome Research* 110: 108–116.
- Voytas DF Cummings MP Konieczny A Ausubel FM & Rodermel SR 1992 *cop**ia*-like retrotransposons are ubiquitous among plants. *Proceedings of the National Academy of Science USA* 89: 7124-7128.
- Wagner GP & Lynch VJ 2005 Molecular evolution of evolutionary novelties: the vagina and uterus of therian mammals. *Journal of Experimental Zoology B Molecular and Developmental Evolution* 304B: 580-592.
- Wagner GP & Lynch VJ 2010 Evolutionary novelties. *Current Biology* 20: R48-R52.
- Wagner GP Amemiya C & Ruddle F 2003 Hox cluster duplications and the opportunity for evolutionary novelties. *Proceedings of the National Academy of Science USA* 100: 14603–14606.
- Walbot V & Evans MMS 2003 Unique features of the plant life cycle and their consequences. *Nature Reviews Genetics* 4: 369-379.
- Wallace MR Andersen LB Saulino AM Gregory PE Glover TW & Collins FS 1991 A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 353: 864–866.

Combined Bibliography

- Walters RD Kugel JF Goodrich JA 2009 Invaluable junk: the cellular impact and function of Alu and B2 RNAs. *IUBMB Life* 61:831-837.
- Wang H Xing J Grover D Hedges DJ Han K et al M A 2005 SVA elements: a hominid-specific retroposon family. *Journal of Molecular Biology* 354: 994-1007.
- Wang R-L Stec A Hey J Lukens L & Doebley J 1999 The limits of selection during maize domestication. *Nature* 398: 236-239
- Wang T Zeng J Lowe CB Sellers RG Salama SR et al 2007 Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Science USA* 104:18613-18618.
- Wang Y Liška F Gosele C Šedová L Křen V et al 2010 A novel active endogenous retrovirus family contributes to genome variability in rat inbred strains. *Genome Research* 20: 19-27.
- Wang Z Miyake T Edwards SV & Amemiya CT 2006 *Tuatara* (Sphenodon) genomics: BAC library construction, sequence survey, and application to the DMRT gene family. *Journal of Heredity* 97: 541–548.
- Warren WC Hillier LDW Marshal-Graves JA Birney E Ponting CP et al 2008 Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 2008 453:175-183.
- Warren WC Clayton DF Ellegren H Arnold AP Hillier LW et al 2010 The genome of a songbird. *Nature* 464: 757-762.
- Waterston RH Lindblad-Toh K Birney E Rogers J Abril JF et al 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.
- Watson JB & Sutcliffe JG 1987 Primate brain-specific cytoplasmic transcript of the Alu repeat family. *Molecular and Cell Biology* 7:3324-3327.
- Wendel JF 2000 Genome evolution in polyploids. *Plant Molecular Biology* 42: 225-249.
- Wendel JF & Doyle JJ 2004 Polyploidy and evolution in plants. *Diversity and Evolution in Plants* (ed Henry R) CAB Publishing Oxon UK.
- Wessler SR 1988 Phenotypic diversity mediated by the maize transposable elements *Ac* and *Spm*. *Science* 242: 399-405.

Combined Bibliography

- Wessler SR 2001 Plant transposable elements. A hard act to follow. *Plant Physiology* 125: 149-151.
- Wessler SR 2006 Eukaryotic transposable elements: teaching old genomes new tricks. In: *The Implicit Genome*. Caporale L editor. Oxford University Press New York.
- Wessler SR Bureau TE & White SE 1995 LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetics & Development* 5: 813-821.
- Wheelan SJ Aizawa Y Han JS Boeke JD 2005 Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Research* 15:1073-1078.
- Werner A & Swan D 2010 What are natural antisense transcripts good for? *Biochemical Society Transactions* 38: 1144-1149.
- White SE Habera LF & Wessler SR 1994 Retrotransposons in the flanking regions of normal plant genes- a role for copia-like elements in the evolution of gene structure and expression. *Proceedings of the National Academy of Science USA* 91:11702-11796.
- Whitham TG & Slobodchikoff CN 1981 Evolution by individuals, plant-herbivore interactions, and mosaics of genetic variability; the adaptive significance of somatic mutations in plants. *Oecologia* 49: 287-292.
- Wicker T Sabot F Hua-Van A Bennetzen JL Capy P et al 2007 A field classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973–982.
- Wiens D & Slaton MR 2012 The mechanism of background extinction. *Biological Journal of the Linnean Society* 105: 255-268.
- Wilson DE & Reeder DM 2005 *Mammal species of the world: a taxonomic and geographic reference* (3rd ed). Baltimore Johns Hopkins University Press.
- Wilson TG 1993 Transposable elements as initiators of insecticide resistance. *Journal of Economic Entomology* 86: 645-651.
- Wing SL 2004 Mass extinctions in plant evolution. *In: Extinctions in the History of Life*. PD Taylor ed Cambridge University Press New York.

Combined Bibliography

- Wooding S & Jorde LB 2006 Duplication and divergence in humans and chimpanzees. *BioEssays* 28: 335–338.
- Woods RJ Barrick JE Cooper TF Shrestha U Kauth MR & Lenski RE 2011 Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 331: 1433-1436.
- Wright DA & Voytas DF 2002 *Athila4* of *Arabidopsis* and *Calypso* of Soybean define a lineage of endogenous plant retroviruses. *Genome Research* 12: 122-131
- Wright S 1931 Evolution in Mendelian populations. *Genetics* 16: 97-159.
- Wu J Grindlay GJ Bushel P Mendelsohn L & Allan M 1990 Negative regulation of the human epsilon-globin gene by transcriptional interference: role of an Alu repetitive element. *Molecular and Cell Biology* 10:1209-1216.
- Wu M Li L & Sun Z 2007 Transposable element fragments in protein-coding regions and their contributions to human functional proteins. *Gene* 401:165-171.
- Xie D Chen CC Ptaszek LM Xiao S Cao X et al 2010 Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Research* 20:804-815.
- Xing J Wang H Belancio VP Cordaux R Deininger PL & Batzer MA 2006 Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proceedings of the National Academy of Science USA* 103:17608-17613.
- Yamada T Ohtani S Sakurai T Tsuji T Kunieda T & Yanagisawa M 2006 Reduced expression of the endothelin receptor type B gene in piebald mice caused by insertion of a retroposon-like element in intron 1. *Journal of Biological Chemistry* 281:10799-10807.
- Yang N & Kazazian HH Jr 2006 L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Structural and Molecular Biology* 13: 763–771.
- Yates PA Burman RW Mummaneni P Krussel S & Turker MS 1999 Tandem B1 elements located in a mouse methylation center provide a target

Combined Bibliography

- for de novo DNA methylation. *Journal of Biological Chemistry* 274: 36357–36361.
- Yi P Zhang W Zhai Z Miao L Wang Y & Wu M 2003 Bcl-rambo beta, a special splicing variant with an insertion of an Alu-like cassette, promotes etoposide- and Taxol-induced cell death. *FEBS Letters* 534:61-68.
- Yoder JA Walsh CP & Bestor TH 1997 Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* 13: 335– 340.
- Young ND Debellé F Oldroyd GE Geurts R Cannon SB 2011 The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480:520-4.
- Yousoufian H & Lodish HF 1993 Transcriptional inhibition of the murine erythropoietin receptor gene by an upstream repetitive element. *Molecular and Cellular Biology* 13:98-104.
- Yu S Li J & Luo L 2010 Complexity and specificity of precursor microRNAs driven by transposable elements in rice. *Plant Molecular Biology Reporter* 28: 502-511.
- Zeh DW Zeh JA & Ishida Y 2009 Transposable elements and an epigenetic basis for punctuated equilibria. *BioEssays* 31: 715–726.
- Zaiss DM & Kloetzel PM 1999 A second gene encoding the mouse proteasome activator PA28beta subunit is part of a LINE1 element and is driven by a LINE1 promoter. *Journal of Molecular Biology* 287: 829-835.
- Zhang R Wang YQ & Su B 2008 Molecular evolution of a primate-specific microRNA family. *Molecular Biology and Evolution* 25:1493-1502.
- Zheng JH Natsuume-Sakai S Takahashi M & Nonaka M 1992 Insertion of the

Combined Bibliography

B2 sequence into intron 13 is the only defect of the H-2k C4 gene which causes low C4 production. *Nucleic Acids Research* 20:4975-4979.

Zylka MJ Dong X Southwell AL & Anderson DJ 2003 Atypical expansion in mice of the sensory neurone specific Mrg G protein-coupled receptor family. *PNAS* 100: 10043-10048.

Appendices (Table of Contents)

Additional Relevant Published Works Forming a Part of this Thesis:

Oliver KR and Greene (2009b) The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations. *Journal of the Royal Society of Western Australia* 92: 447-451. (Appendix One).

Oliver K.R. and Greene W. K. (2009c) Jumping Genes Drive Evolution. *Australasian Science* 30(8): 29-31. (Appendix Two).

Oliver K. R. and Greene W. K. (2012a) Jumping Genes: How They Drove Primate Evolution. *Australasian Science* 33(1): 18-21. (Appendix Three).

An Additional Published Work Relevant to this Thesis

Proceedings of an Evolutionary Biology Symposium, held in October 2009, convened by V. Semeniuk and K. Oliver. *Journal of the Royal Society of Western Australia*. 92(4): 1-487.

Appendix One

The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

Keith R. Oliver & Wayne K. Greene

Abstract

Orthodox evolutionary theory does not accord with what palaeontologists usually find in the fossil record, which mainly indicates long periods of stasis, interspersed with relatively short periods of rapid change, that is, macro or micro punctuational evolutionary taxonations. This is usually known as punctuated equilibrium. A novel hypothesis we have called Genomic Drive, points towards Transposable Elements (TEs) as powerful facilitators of evolution and as essential for induction of periodic changes in the rate of evolution. The Genomic Drive hypothesis, which is supported by current data, if confirmed, will open the way for the reconciliation of evolutionary theory with the findings of most palaeontologists. It may also help to explain the extraordinary fecundity of some orders, and the paucity of species in others, and why there are “fossil species”.

Keywords: Evolution, the Genomic Drive hypothesis, transposable elements, taxonations, punctuated equilibrium, gradualism, stasis, extinction, fossil record

Box 1:

A Brief Summary of two Major Types of Transposable Elements

Type I Transposable Elements are retrotransposons or retroposons (retro-TEs) which transpose via an RNA intermediate, in a “copy and paste manner” by means of an encoded reverse transcriptase protein (RT), and other protein(s) encoded in the autonomous LINEs (Long Interspersed Nuclear Element), LTRs (Long Terminal Repeats), and ERVs (Endogenous Retroviruses). The non-autonomous SINEs (Short Interspersed Nuclear Element) which have internal promoters use the reverse transcriptase of the LINEs to transpose and multiply. Retrocopies (Processed pseudogenes or PPs) can also be transposed by reverse transcriptase, helping to facilitate evolution, but cannot multiply further, as they lack promoters.

Type II DNA Transposons (DNA-TEs) of most superfamilies transpose by a “cut and paste” mechanism. They encode various transposases, according to their family, which recognises their terminal inverted repeats (TIRs) and transposes them from one location in the genome to another, sometimes with an increase in copy number.

1.0 Introduction

Discovered by Barbara McClintock in the 1950s (McClintock 1950; 1956; 1984), TEs were authoritatively written off thirty years later as parasitic, junk, or selfish DNA, which we would be better off without (Orgel & Crick 1980; Doolittle & Sapienza 1980). However, during the last decade a large number of researchers

Appendix One: The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

have noted that the evolutionary potential of taxa can benefit from the presence of TEs (Kidwell & Lisch 2001; Bowen and Jordon 2002; Kazazian 2004; Feschotte & Pritham 2007; Goodier & Kazazian Jr. 2008) and many others. Building on this foundation, we developed the Genomic Drive hypothesis, the major elements of which were recently published as an unnamed synthesis (Oliver & Greene 2009).

Genomic Drive, according to our hypothesis, is a powerful facilitator of evolution in sexually reproducing eukaryotes. It is the process by which germ line or early embryo genomes engineer coding, regulatory, karyotypic, or other changes to their own genome. Transposable elements (TEs) (see Box 1) are the major facilitators of evolution by Genomic Drive (Oliver & Greene 2009). Other genomic content, such as simple sequence repeats (SSRs) also make some contribution (Kashi & King 2006; King et al. 2006). Of course we do not deny that many other factors also facilitate evolution and may possibly result in punctuation events on some occasions. Some examples are: whole genome duplications, endosymbiosis, horizontal gene transfer (especially in bacteria) point mutations, insertions and deletions, and other such well known phenomena. We also acknowledge that there may be some as yet unknown phenomena that help to facilitate evolution and that may also give rise to punctuation events. We present Genomic Drive as a hypothesised major facilitator of evolution, but certainly do not claim that there are no other significant facilitators of evolution. Indeed, some phenomena such as point mutation highly complement Genomic Drive by allowing newly engineered DNA sequences to diversify. A notable

example of this is the ape-specific *GLUD2* gene, which encodes a glutamate dehydrogenase enzyme involved in neurotransmitter recycling. Derived from a retrotransposed copy of *GLUD1* that has undergone critical nucleotide substitution events, *GLUD2* appears to have significantly increased the cognitive powers of the apes (Burki & Kaessmann, 2004).

2.0 Major Principles of the Genomic Drive Hypothesis

TEs (Transposable Elements) are ubiquitous, comprising 20% to 80% of most genomes, and are extremely ancient; they are powerful facilitators of evolution. We have proposed this powerful facilitation of evolution by TEs, as the Genomic Drive hypothesis. Successful taxa do not destroy TEs, but strongly control transposition of TEs in the soma, where they are often damaging and cannot be inherited. However, they allow some TE activity in the germ line and the early embryo, where they can generate potentially useful variation in progeny, for natural selection to work on. Thus Genomic Drive can cause genetic changes of great magnitude and variety within germ line genomes, making such genomes flexible and dynamic, so that they drive their own evolution and the evolution of their resultant phenotype. Genomic Drive can cause many genomic alterations that cannot be caused by any other mutagens. The de novo synthesis of new, or orphan, genes from TEs, and perhaps other non-coding DNA, has recently been confirmed in fruit fly and humans. Genomic Drive can build, sculpt, and reformat genomes by both active and passive means. Active Genomic Drive is due to the active transposition of TEs, from either a heterogeneous or homogeneous population of TEs. Passive Genomic Drive is due to ectopic

Appendix One: The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

recombinations between homologous TE insertions. Such ectopic recombinations are common only when there are large homogeneous populations of TEs. TEs can infiltrate germ lines by endogenous de novo synthesis, e.g. SINEs, SVAs, by exogenous invasions by retroviruses e.g. ERVs and LTRs, and by horizontal transfer between taxa, mostly by DNA-TEs. New infiltrations of germ line genomes by TEs, or modifications to existing TEs, or various stresses experienced by the phenotype, can result in intermittent bursts of TE activity. We propose that this can result in punctuational evolutionary taxonations or radiations, usually known as punctuated equilibrium.

Although sometimes harmful to some individuals, TEs can be very beneficial to lineages. The result of this is lineage selection for lineages endowed with a suitable repertoire of TEs; this endows such lineages with enhanced evolutionary potential. Taxa or lineages deficient in active TEs, and with heterogeneous populations of inactive TEs, tend to be non-fecund, tend to prolonged stasis, and eventually may become extinct. Conversely, taxa or lineages well endowed with such TEs tend to be fecund, or species rich, as they taxonate readily. This could be called the evolution of evolvability. Cellular defences against excessive TE activity have resulted in the capacity of genomes to evolve epigenetic controls of TEs, which may further facilitate evolution or adaptation by epigenetic means.

In short, TEs, which we propose constitute the main engine of Genomic Drive, can result in the generation of widely divergent new taxa, fecund lineages, lineage selection, and punctuated

equilibrium (Oliver & Greene 2009). Others, to give but one example, Goodier & Kazazian Jr. (2008), have clearly recognised the likelihood that bursts of TE transposition could accelerate taxonations, and that evolution has been adept at changing “junk” into treasure. However, they have stopped short of developing any hypotheses to this effect, and exploring the implications of such hypotheses.

3. 0 Gradualism and Punctuated Equilibrium

Gradualism and Punctuated Equilibrium are two possible modes of evolution. Current orthodox evolutionary thought is dominated by an assumption that biological lineages evolve by the slow and gradual accumulation of adaptive mutations, that is, by gradualism, and that macroevolution (the origin of higher taxa) can be explained by an extrapolation of microevolution (the origin of races, varieties and species) into the distant past (Kutschera and Niklas 2004; and many others). This line of thought has been mostly dominant since Charles Darwin who, influenced by Lyell’s concept of very slow changes in geology, regarded gradualism as fundamental to his theory. Darwin unreservedly said *Natura non facit saltum* (nature does not make a leap). Despite a number of early dissenters who strongly advocated evolution by discontinuous variation or sudden leaps, gradualism was eventually incorporated into neoDarwinism and the Modern Synthesis (Bowler 2003). However, many palaeontologists have found that gradualism does not concur with the majority of the fossil record. Instead, new species are found to arise abruptly and periodically and there are intermittent and often long periods of stasis, punctuated by periods of rapid change and branching

Appendix One: The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

speciation. These punctuations often occur during different periods in diverse lineages, so are apparently not always related to environmental changes. The observed persistence of ancestors in stasis, following the abrupt appearance of a descendant, is an indicator of punctuated equilibrium (Eldridge and Gould 1972; Stanley 1981; Eldridge 1986, 1995; Gould 2002). This is not to be confused with the hypothesis that fairly rapid, but gradualistic, evolution can occur in peripheral isolates, followed by the movement of the resulting new taxon back into the main population, giving the erroneous appearance of a gap in the fossil record.

Punctuated equilibrium, as detailed by the palaeontologists cited above, has been observed in certain very fine grained strata, and entails intermittent periods of rapid evolutionary change, over an estimated 15,000 to 40,000 years (Gould 2002), which gives birth to a new taxon that remains little changed (i.e. in a period of stasis) until it becomes extinct, usually four to ten million years later. This taxon is often the progenitor of other taxa in the same lineage, while it is still extant. Contemporaneous, or successor taxa, in the same lineage eventually suffer the same fate. Of course, mass extinction events can interrupt this pattern, but they only account for less than 5% of all extinct species and recovery from them tends to be slow, about 5 million years in the Early Triassic, after the end of Permian great mass extinction (Erwin 2001). This seems to make the “Cambrian explosion” seem all the more remarkable. That gradualism occurs sometimes is not denied, and Fortey (1985), from a study of Ordovician trilobites, estimated that the ratio of punctuated equilibrium type speciation

Appendix One: The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

to gradualist speciation is 10:1, while Ridley (2004) posits that although both occur, and punctuated equilibrium appears to be the more common, they may be extremes of a continuum. It seems, therefore, that the ratio of these types of speciation events, one to the other, is somewhat uncertain. According to Gould (2002), punctuated equilibrium should not be confused with the hypothesised evolution of “hopeful monsters” by saltations (Goldschmidt 1940). Whereas many palaeontologists have observed punctuated equilibrium, they could not explain it satisfactorily in terms of the Modern Synthesis. Now, however, intermittent waves of transposable element activity have very recently been hypothesised to be a major causal factor of punctuated equilibrium (Oliver and Greene 2009; Zeh et al. 2009; Parris 2009), seemingly finally reconciling evolutionary theory to punctuated equilibrium, and the fossil record. However, whereas Zeh et al. (2009) place heavy emphasis on environmental stress as a trigger for TE activity, we additionally consider *de novo* emergence (e.g. SINEs) activating modifications, especially to promoter regions (e.g. LINEs and SINEs), horizontal transfer of TEs, and germ line invasions by retroviruses, as intermittent events that can trigger new waves of TE activity (Oliver and Greene, 2009). Parris (2009) suggests intermittent germ line invasions by retroviruses, possibly in concert with environmental change, as an example of a trigger for intermittent rapid taxonations. There are many examples of such intermittent waves of TE activity temporarily accelerating evolution, such as the amplification peaks of the now extinct L2 LINEs and MIR SINEs roughly coinciding with the marsupial-eutherian split ~120-150 Mya, and the peak activity of the L1 LINEs corresponding to the

early eutherian radiation ~100 Mya (Kim et al. 2004). *Myotis* (a genus of microbats) is one of the most species rich mammalian genera and Stadlemann et al. (2007) found that a burst of *Myotis* diversification occurred ~12-13 Mya corresponding well to the estimated time during which the most active DNA-TE families were expanding in the *Myotis* genome (Ray et al. 2008).

4.0 Transposable Elements have Periodic Waves of Activity

Although retro-TEs can horizontally transfer between taxa only very rarely, new infiltrations of germ lines by retro-TEs can occur by *de novo* synthesis of SINEs, the fusion of SINEs into dimers, or the fusion of SINEs with other complex elements. Activating modifications to SINEs, and to the untranslated (promoter) regions of LINEs can also occur, and so on. Invasions of the germ line by exogenous retroviruses are also common, as are modifications to existing endogenous retroviruses or retroviral remnants. All of these intermittent events can result in transient waves of retrotransposition. Such intermittent waves of transposition often result in contemporaneous waves of retrocopies or retrogenes (these are sometimes called processed pseudogenes) which can sometimes be converted into useful new genes by other mutations. SINEs (especially Alu SINEs) are also thought to be typically activated in response to stresses on the host organism (Oliver & Greene 2009; Zeh et al. 2009). These waves of retrotransposition can therefore activate periods of rapid evolution punctuating the more normal near stasis of a taxon, giving punctuated equilibrium type evolution. In contrast to retro-TEs, many DNA-TEs can readily transfer horizontally from one taxon to another, sometimes between widely divergent lineages. There is

often a dramatic wave of TE activity following on from a horizontal transfer event (Pace et al. 2008). However, all transposable elements, both Type I and Type II, can eventually succumb to the increased effectiveness of cellular controls such as methylation and interfering RNAs, and to crippling mutations which can eventually result in much reduced activity, and a possible return to near stasis in the affected taxon. But there are nearly always periodic new infiltrations of TEs in genomes by one means or another, to keep Genomic Drive going and to intermittently result in rapid evolution. If this does not occur then, according to the Genomic Drive hypothesis, extinction of the affected taxon is likely, if all other things are equal, or alternatively the taxon could become a fossil species. This is in agreement with the fossil record where extinction, aside from the mass extinctions, is the normal eventual fate of a taxon, but usually the lineage to which it belongs, survives. An example of this is the hominid lineage, where a number of earlier successful hominids which were in stasis, showing little variation over time, such as *Homo erectus*, succumbed to extinction but where *Homo sapiens* both survives and thrives (Eldridge 1986).

5.0 Genomic Drive in Mammals

The Human Genome is composed of about 45% TEs: ~42% are retro-TEs, made up of ~21% LINEs, ~13% SINEs, ~8% LTRs and ERVs. Most, but by no means all, of these are inactive, so some active Genomic Drive is continuing in humans. However, as nearly all of the LINEs are L1s and nearly all of the SINEs are the primate-specific Alu SINEs, there is a good potential for passive Genomic Drive, by means of such repetitious DNA promoting

Appendix One: The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

ectopic recombinations. A wide variety of DNA-TEs make up the other ~3% of TEs in the human genome, all of which are inactivated molecular fossils, but some have been exapted for cellular functions and are under positive selection, as in the *SETMAR* chimeric primate (anthropoids only – not prosimians) gene (Cordaux et al. 2006). TEs do cause disease in individuals, but in humans only slightly more than 0.5% of known genetic diseases are attributable to TEs. With ~356 extant species the primates are moderately fecund. Bursts of considerably increased TE activity have been associated with the major separations and divergences in the primate lineage, such as those of prosimians and Old World monkeys, and of Old World monkeys and apes (Oshima et al. 2003; Kim et al. 2004; Khan et al. 2006).

Quite atypically for mammals, the bats have many recently active DNA-TEs, some of which may be still active, as well as retro-TEs (Ray et al. 2007; Ray et al. 2008; Pritham and Feschotte 2007). Bats are correspondingly fecund, and with approximately 1000 extant species they comprise over 20% of all mammalian species. Bats evolved in an early Eocene “big bang” ~52 million years ago (Simmons 2005) and appear to have taxonated rapidly. Rodents exhibit even greater fecundity, comprising close to 40% of all extant mammals, and these are well endowed with retro-TEs, but in at least some rodents (mice) individuals apparently pay a high price for the success of their lineage, as they have very many more TE-caused genetic diseases in individuals than do humans (Maksakova 2006). The bat and rodent orders contrast with the colugos, or “flying lemurs” (order Dermoptera) with only 2 to 4 species, but little is known about their TEs at present, except that

they lack the 7SL derived SINEs that may have been a major factor in the successful radiations of the rodents and primates. More data on the genomes of colugos would be very valuable in assessing the consistency of their paucity of species with the Genomic Drive hypothesis.

With punctuated equilibrium evolution, as evidenced in the fossil record, stasis is the normal condition, and rapid change occurs rarely. Stasis is data, which must be accounted for in a satisfactory theory of evolution. “Living fossils” such as the Tuatara and the Coelacanth have been more or less in stasis for hundreds of millions of years. Information about the TEs in their genomes is scarce, but what there is suggests that these species have a paucity of TE activity, which is entirely consistent with the Genomic Drive hypothesis (Oliver & Greene 2009). The robustness of the Genomic Drive hypothesis as applied to the evolution of other phyla has scarcely even been contemplated, but there is certainly evidence of long periods of stasis in some insects, and in crocodilians.

6.0 The Genomic Drive Hypothesis and Plants

TEs are also very active in plants and some angiosperm genomes are comprised of up to ~80% TEs. We have not yet extensively investigated the validity of the Genomic Drive hypothesis for plant evolution, but preliminary investigations indicate that plants also show punctuated equilibrium type taxonations. Plants have their fecund lineages, such as orchids in the monocots and daisies in the dicots, among the very numerous angiosperms, which evolved ~130 million years ago. Plants also have “fossil species” like the

gymnosperm *Ginkgo biloba*, which has leaves similar in form and venation to those found in rocks deposited in the Mesozoic era (248-65 Mya) when ginkgo-like plants apparently had a worldwide distribution (Foster & Gifford 1974). Gymnosperms, which evolved ~350 million years ago, have relatively few species and seem to be in stasis, but are still very successful in terms of biomass, forming extensive forests in both the northern and southern hemispheres. Hybridisation giving sterile plants, followed by polyploidy to give fertile plants, is an observable and well documented example of punctuated equilibrium. The endosperm has also played a significant role in the evolution of many angiosperms by means of possible reproductive isolation caused by mismatches of the endosperm balance number (Johnson et al. 1980). The Genomic Drive hypothesis, when its probable application to plants is thoroughly investigated, may also help to explain the major transition of the angiosperms from their possible seed fern progenitors. However, another factor to consider is that the large advances in plant evolutionary forms are all associated with high global CO₂ levels at different periods of the earth's history (Calver et al. 2009).

7.0 Conclusions

Genomic Drive is a new hypothesis which still needs much development and testing, but it powerfully portrays the profound effects that waves of transposable element activity produce in intermittently driving evolution, and also the possible passive effects of homogenous repertoires of inactive TEs. Much evidence suggests that it offers an explanation for stasis, and for rapid punctuational evolutionary taxonations and/or radiations, or

Appendix One: The Genomic Drive Hypothesis and Punctuated Evolutionary Taxonations, or Radiations

punctuated equilibrium as it is usually called, as has been found in the fossil record. However, like all hypotheses it needs to be subjected to testing. If it is confirmed, it will offer a new conceptual foundation for much of evolutionary theory and will probably enable a reconciliation of the findings of palaeontologists with biological evolutionary theory which has not been possible previously. It also seems likely that it, when fully developed, will be able to help to explain why some lineages are very fecund, while other lineages are quite non-fecund and why some lineages evolve rapidly while others linger in stasis and terminate in “living fossils,” and similar puzzles. Working out the relationship between TEs and evolution, in terms of cause and effect, seems likely to be a fruitful area of research well into the foreseeable future.

Jumping Genes Drive Evolution

Orthodox evolutionary theory does not tally with the fossil record, but a new school of thought points towards so-called jumping genes as essential agents of periodic changes in the rate of evolution.

Current evolutionary thought is dominated by an assumption that biological lineages evolve by the slow and gradual accumulation of adaptive mutations. However, this does not match up with most of the fossil record. Instead, new species are found to arise abruptly and periodically, and there are intermittent and often long periods when very little happens, a situation called evolutionary stasis.

Our evolutionary hypothesis, which we call “Transposon Thrust”, states that significant evolution cannot take place without the activity of jumping genes, properly known as transposons or transposable elements. Discovered by Barbara McClintock in the 1950s, they are so-named because of their capacity to jump (or copy themselves) from one position to another in the DNA of an organism. In the 1980s, jumping genes, which are almost universally abundant in genomes, were written off as parasitic, junk, or selfish DNA that we would be better off without.

However, ever-increasing evidence over the past decade has begun to turn this idea on its head, with many studies revealing that jumping genes can generate genetic changes of great variety and magnitude. As with other types of mutations, a proportion of the DNA changes caused by jumping genes will, by chance, be beneficial and be positively selected in evolution. Of course they

can also cause harm, but jumping genes are only a minor source of known genetic disease, causing, for example, just over 0.5% of the total in humans. We argue that this short-term cost to a very small number of individuals is massively outweighed by the longer-term benefits to the evolution of the lineage.

Because they promote adaptability, we consider jumping genes to be extremely useful, if not essential, genomic parasites. This is not to say that jumping genes are the only cause of evolution, but that they are hugely important and powerfully complement other processes such as point mutations, where the wrong DNA bases are inserted at particular locations; horizontal transfer, where one organism transfers genes to another organism that is not its progeny; and polyploidy, where an organism ends up with more than the usual two copies of the genome.

Jumping genes can create useful genetic change, the raw material upon which natural selection acts, in two basic ways. Firstly, they can operate in an active fashion, either by inserting into new locations of the genome to seed new genes or parts of genes, or by inadvertently copying and pasting existing genes or parts of genes from one location to another. Such activity tends to be transient since over time jumping genes become inactive as they succumb to the effects of random mutation. Nevertheless, the mere presence of large numbers of inactive, but similar, jumping gene relics can secondarily cause genetic changes in a passive fashion. This is because they create a “hall of mirrors”; a plethora of virtually identical sites within the genome, which promotes major reorganisations of DNA by confusing the cellular

machinery involved in its propagation. This can result in genes or parts of genes being either duplicated or lost altogether. The loss of genes is not always disadvantageous, but if it is then there will be selection against the affected individuals.

In their active mode, even small numbers of jumping genes will have a great impact on their host genome, and high activity is likely to reoccur with every new invasion of jumping genes into a lineage. New invasions can occur either by horizontal transfer, such as through viruses or bacteria, or by the natural origination of jumping gene activity from within a genome. By contrast, to have significant passive effects on a genome, near-identical copies of jumping gene relics must be present in great numbers. This is the situation in humans and other primates, for example, whose genomes are roughly half comprised of jumping gene relics of two major varieties. These are the so-called LINE-1 and Alu elements, which in the human genome are present in a whopping 0.5 and 1.1 million copies, respectively.

A central tenet of our Transposon Thrust hypothesis is that lineages which have active jumping genes, or alternatively large populations of the same type of jumping gene relic (that can act passively), are adaptable and spawn new species readily. Conversely, species whose genomes are deficient in jumping genes, or which possess a great mixture of different types of jumping gene relics, tend to undergo evolutionary stasis (become frozen) and may risk extinction by lacking the capacity to adapt and change, or diversify. Transposon Thrust can provide answers to six key mysteries in evolutionary biology, namely:

1) Why do species appear suddenly in the fossil record?

New species appear suddenly because jumping genes can cause major genetic changes in a lineage rather rapidly, rather than gradually. They do this by creating new genes, altering the control switches of existing genes or rearranging chromosomes. These large changes are thought to be the major means by which new species-specific traits evolve and a significant number of them cannot be caused in any other way.

2) What is the cause of punctuated equilibrium?

Punctuated equilibrium is rapid evolution followed by slow evolution, or a stoppage in evolution, as is observed in the fossil record. This can be explained by the fact that jumping gene activity does not occur at a low and uniform rate over time. Instead, it sporadically occurs in sudden bursts resulting in rapid evolution, followed by decreasing activity and slowing evolution. These rapid bursts of evolution can happen when a new type of jumping gene is suddenly transferred into a lineage from some other lineage or when a new type of jumping gene naturally emerges from within a genome. Jumping gene activity can also increase as a response to stress, temporarily increasing the rate of evolution. Successive waves of jumping gene activity thus account for alternating periods of rapid evolution and stasis, and can thereby reconcile evolutionary theory with palaeontology and the fossil record.

3) Why are some lineages of organisms species-rich and others species-poor?

Species-rich lineages, which among the mammals include rodents, bats and primates, have had successive bursts of jumping gene activity over evolutionary time, extending into recent times or to the present. Species-poor lineages such as the primate cousins known as flying lemurs, have not had recent bursts of activity, but probably had them in the very distant past. Such waves of activity may also help to explain why certain other groups of animals are particularly diverse, such as the songbirds, which account for over half of all bird species and the perciform (perch-like) fish, which account for 40% of all fish species, although there is insufficient data to verify this at present.

4) Why do living fossil species change little over millions of years while other lineages evolve rapidly?

Living fossils such as the lobe-finned coelacanth fish and the reptilian tuatara of New Zealand, have remained virtually unchanged for 410 and 220 million years, respectively. As examples of evolutionary stasis, these fossil species appear to have had no new infiltrations of jumping genes, except in the very distant past. What little jumping gene relics they do possess are in low numbers and/or very diverse leaving little scope for passive effects either. As a result, they are effectively frozen in time. In contrast, most lineages of mammals have evolved rapidly following the extinction of the dinosaurs 65 million years ago.

5) Why do species have differing controls on jumping genes in reproductive cell DNA compared to ordinary body cell DNA?

Jumping gene activity in normal body cells is heavily restricted by multiple mechanisms including a chemical modification to jumping gene DNA called methylation. In reproductive cell (sperm, egg and early embryo) DNA, these controls are temporarily relaxed, which creates a window of opportunity to allow some jumping gene activity. This difference between these two cell types can be explained by the fact that genetic changes caused by jumping gene activity in ordinary body cell DNA cannot benefit the lineage because they can't be passed on to the next generation. Rather, they can be damaging to individuals, for example by occasionally causing mutations that lead to cancer. By contrast, jumping gene activity in reproductive cell DNA can create valuable genetic variation that can be inherited and which natural selection can work on. Thus, successful lineages from single-celled protozoa right through to mammals specifically permit jumping gene activity in reproductive cells for the potential benefit of future generations and strictly minimize it in body cells where it is potentially harmful to the individual.

6) Why do almost all species only suppress jumping genes rather than eliminate them?

Although the types and total amount of jumping genes present vary greatly between different groups of organisms, they often comprise a large, if not massive fraction of the genome. Known mammalian genomes are at least one-third jumping gene DNA in origin, while plant genomes often have an even higher jumping gene DNA content of over two-thirds. It has long been a puzzle as to why many species tolerate having so much of this so-called junk, parasitic or selfish DNA within their genomes. Our answer is

that any species that eliminates its jumping genes cripples its evolutionary potential and greatly increases its chances of extinction, so it is not beneficial for it to do this. It is far better for a species to suppress jumping genes in body cells, while allowing them some activity in reproductive cells in order to promote evolvability, at a cost to a small number of individuals in terms of inherited genetic disorders.

In conclusion, we have no doubt that compelling evidence now indicates that jumping genes have had a major role in evolution as irreplaceable sources of novel genetic changes. Far from being parasites or junk, jumping genes have made their host genomes flexible and dynamic, so that the genomes themselves can promote their own evolution. Their legacy is astounding, ranging from the creation (and sometimes destruction) of genes to the genome-wide seeding of gene control switches and wholesale rearrangement of chromosomes. Periodic bursts of jumping gene activity not only predict punctuated equilibrium as a general characteristic of evolution, but provide an explanation as to how some lineages are able to spectacularly diversify while others are liable to evolutionary stasis. As more data becomes available in the future on jumping genes and their contribution to the genomes of a wide range of species, awareness of their pivotal role in evolution should also grow.

Keith Oliver and Wayne Greene

Appendix Two: Jumping Genes Drive Evolution

Keith Oliver is a biologist and philosopher in the School of Biological Sciences and Biotechnology and Dr Wayne Greene is an Associate Professor in the School of Veterinary and Biomedical Sciences at Murdoch University.

Jumping Genes: How They Drove Primate Evolution

BY KEITH OLIVER AND WAYNE GREENE

Jumping genes have been important in the evolution of higher primates, leading to faster brain function, improved foetal nourishment, useful red-green colour discrimination and greater resistance to disease-causing microbes – and even the loss of fat storage genes in gibbons.

Most DNA is inert, but some DNA sequences are mobile in that they can move, or jump, from one location in a genome to another by copy and paste processes. These so-called transposable elements or “jumping genes” are important because their activity within genomes, past and present, gives them the ability to cause a great variety of genetic changes. While this can be harmful to the occasional individual, for example by causing a genetic disorder, overall it is a boon for the evolution of living things because it increases the amount of potentially beneficial genetic variation upon which natural selection can act. Jumping genes are thus not unlike Rumpelstiltskin, the fairy tale rascal who was somewhat troublesome, yet had the wondrous ability to spin straw into gold.

Jumping genes are ancient and ubiquitous, being found throughout the animal and plant kingdoms. Long regarded as “junk DNA” by some, they can act over and above other known ways by which DNA mutations occur to make genomes more changeable, thereby boosting evolutionary potential. We have recently developed a hypothesis that explains how jumping genes

provide an extra “evolutionary boost.” According to the “Transposon Thrust” hypothesis, jumping genes powerfully promote evolution in one of two major ways.

1) In what we call *active* Transposon Thrust, jumping genes make changes to genomes through insertion into new locations or by inadvertently copying and pasting normal cellular genes from one place to another.

2) In *passive* Transposon Thrust, following insertion in multiple locations, jumping genes create a profusion of identical DNA sequences - a virtual “hall of mirrors” - that confuse the cellular machinery involved in DNA propagation, leading to an increased rate of duplications, deletions and reorganisations of chromosomal regions.

Through both of these ways, jumping genes can cause very substantial and elaborate changes to genomes by creating new genes or altering, or changing the control of, existing ones. This results in biological lineages that can adapt well to environmental changes or challenges and/or take advantage of new ecological opportunities. It can also pave the way for spectacular radiations of species and the generation of wholly new lineages. By this same reasoning, lineages lacking jumping genes are liable to become “frozen” in evolution, possibly becoming “living fossils” or even extinct.

The activity and types of jumping genes present within genomes varies from lineage to lineage and also over time within any particular lineage. Their activity is usually intermittent, with

periodic bursts of copy-and-paste activity due to either a relaxation of cellular controls (such as after stress), the emergence of new or modified jumping genes within a genome, or their transfer across species. Given enough time, most jumping genes suffer random mutations and eventually become incapable of activity.

Episodic jumping gene activity, and inactivity, helps to explain variations in the rate of evolution over time, why evolution appears to have stalled in some organisms, and why some lineages are highly successful and/or rich in species.

Transposon Thrust is expected to be most effective in lineages in which jumping genes are highly active (for active Transposon Thrust) and/or possess large numbers of the same kind of jumping gene (for passive Transposon Thrust). We have hypothesized four main modes of Transposon Thrust, which help to explain differing modes of evolution that are apparent from the fossil record:

Mode 1: Active Thrust Only.

Periodically active but highly mixed populations of jumping genes would likely result in alternating periods of relatively fast evolution followed by little or no change. Active Transposon Thrust would come into effect during periods of jumping gene activity while there would be little or no passive Transposon Thrust due to the mixed bag of elements present.

Mode 2: Active and Passive Thrust.

Periodically active and highly uniform large populations of jumping genes would likely result in alternating periods of relatively fast evolution followed by more gradual change. Active Transposon Thrust would come into effect during periods of jumping gene activity while in between there would still occur gradual change facilitated by passive Transposon Thrust due to the plethora of identical elements present.

Mode 3: Neither Active Nor Passive Thrust

Inactive and highly mixed populations of jumping genes would likely result in prolonged periods of little or no change, which may lead to eventual extinction and/or the occurrence of living fossils - notable examples being the tuatara and coelacanth. In this situation there is a lack of both active and passive Transposon Thrust.

Mode 4: Passive Thrust Only.

Inactive and highly uniform large populations of jumping genes would likely result in long periods of gradual change. In this situation there is a lack of active Transposon Thrust but there would still be ongoing passive Transposon Thrust.

A key element of our Transposon Thrust hypothesis is that jumping genes can promote the origin of new lineages and subsequently exert a large influence on the course and extent of evolution within such lineages. The evolutionary history of the relatively well- studied primate lineage is a case in point. This was characterised by periodic bursts of jumping gene activity, which have been found to correlate with major divergence points in

primate evolution, including splits between the higher primates and prosimians, the Old and New World monkeys and the apes and Old World monkeys. Over millions of years, the activity of jumping genes was such that, incredibly, they now make up nearly half (45%) of our entire genome! Jumping gene activity is presently much reduced in primates, although higher primate genomes remain well-suited for passive Transposon Thrust, with just two types of jumping gene, the so-called Alu and L1 repeats predominating. These two elements have been amazingly prolific and, within the human genome, now number a whopping 1.1 million and 516,000 copies, respectively. The Alu jumping gene is particularly interesting. Not only is it extremely abundant, but it is only found in primates and it cannot “jump” of its own accord; instead it depends on the copy-and-paste machinery of L1.

Among other things, the higher primates (monkeys, apes and humans) have undergone significant advancements in brain function, reproduction and defence against infectious diseases. By examining the evolution of the primate lineage, one can find some of the strongest specific evidence for the existence of Transposon Thrust. Most evidently, jumping genes have helped drive the separation of the higher primates away from the prosimians, or lesser primates, by engineering changes to DNA sequences that underpin many features that are characteristic of monkeys, apes and/or humans. The advancement, and radiation, of higher primates seems to be, at least in part, due to exceptionally powerful Transposon Thrust, owing especially to the Alu element along with its L1 partner. This evolutionary boost has operated in a variety of ways, most notably by:

- Actively changing the control of pre-existing genes;
- Actively changing the structure of pre-existing genes or creating entirely new genes;
- Actively changing the control of pre-existing genes;
- Passively acting as scattered near identical sequences (a “hall of mirrors”) to cause duplications, deletions or reorganisations of chromosomal regions.

Actively changing the control of pre-existing genes.

After inserting near a pre-existing gene, jumping genes can be very good at acting as control switches to turn genes on or off. Indeed, when Barbara McClintock first discovered jumping genes in the late 1940's, she called them “controlling elements”. Not surprisingly, this is a very major way by which jumping genes have influenced primate evolution. For example, the enzyme amylase, which digests starch, is produced in saliva (in addition to the pancreas) in Old World primates (including humans) because long ago a jumping gene added a switch near the amylase gene that specifically works in the salivary gland. Similarly, in a primate ancestor an *Alu* element pasted a switch near a gene called *FUT1* that allowed it to be turned on in red blood cells. The result: the well-known ABO blood group system found only in apes and humans. Such a mechanism has also helped in our immune defence against microbial invaders. For example, insertion of an *Alu* next to an anti-microbial gene called *CAMP* enabled this gene to be switched on by Vitamin D. Thus, in response to sunlight, the immune response of higher primates has been given a boost in responding to infection.

Actively changing the structure of pre-existing genes or creating entirely new genes.

Jumping genes can contribute to the DNA sequences of genes themselves to create new functions. This appears to have happened many times in primate evolution, although the reason for the changes has, in most cases, not yet been determined.

Less commonly, but more spectacularly, jumping genes can provide the raw material to create entirely new genes from scratch. Two primate genes that are entirely derived from jumping genes are *Syncytin 1* and *Syncytin 2*. These play a crucial role in the formation of the higher primate placenta to help ensure a good connection between the mother and foetus.

Actively using the copy-and-paste mechanism to copy or destroy pre-existing genes.

Certain jumping genes, such as *L1*, can actively create genetic novelties by using their copy-and-paste mechanism to partially or fully copy a pre-existing gene. The duplication of genes is a very important aspect of evolution, as it creates spare copies of genes that can be tinkered with through further mutations. The result can be a new gene with a related, but distinct function, which may be beneficial to the survival and/or reproduction of its host, and thus be retained in evolution.

A good example of this in primates was the creation of the *GLUD2* gene from a copy of *GLUD1*, by jumping gene activity. Only found in the most intelligent of primates (the apes and humans), *GLUD2* is specifically switched on in the brain where it appears to speed up the recycling of the signalling chemical glutamate and hence

improve learning and memory. Of course, jumping genes can also be destructive when they insert into new locations in the genome. This is not always a bad thing though, an example being the destruction of the *CMAH* gene by a jumping *Alu* sequence in a human ancestor about 2 million years ago. It is for this reason that humans lack a particular sialic acid molecule on the surface of their cells. The loss of *CMAH* probably conferred a survival advantage on the human lineage by decreasing the infectious risk from disease-causing microbes known to use this molecule to attack cells.

Passively acting as scattered near identical sequences (a “hall of mirrors”) to cause duplications, deletions or reorganisations of chromosomal regions.

When a single kind of jumping gene is present in very high numbers within a genome it can increase the chance of gain, loss or gross rearrangement of DNA by confusing the cell division machinery. In primates, the highly abundant *Alu* jumping gene, and to a lesser extent *L1*, have been particularly important factors in this process. For example, they have caused much genomic duplication, that is, generated “carbon copies” of existing genes that have subsequently evolved distinct functions through point changes to their DNA sequences.

A very good example of this was the evolution of red-green colour vision in the Old World primate lineage, which includes apes and humans. Most mammals, including the prosimian primates, have colour-limited vision because they possess just two retina cone photoreceptor genes, one maximally sensitive to blue light and the

other to green. The red-green perception trait apparently had its origin about 40 million years ago from a gene duplication event caused by *Alu* jumping gene sequences. This resulted in three retina cone photoreceptors, with the extra one becoming most sensitive to red light. Among other things, this beneficial change would have immensely improved the ability of the Old World primate lineage to find fruits and other foods.

Periodic bursts of jumping gene activity correlate with major divergence points in primate evolution...

A major focus during primate evolution were changes to reproductive physiology, with the higher primate placenta having developed a number of refinements to ensure efficient nourishment of the growing foetus. Here again, working in a passive manner, jumping genes appear to have played a key role. For example, the growth hormone gene underwent a burst of duplications due to *Alu* sequences, with higher primates now possessing between five and eight copies of the gene. Many of these copies are switched on specifically in the placenta where they help the foetus to acquire resources from the mother by influencing her metabolism. In similar fashion, one of the genes coding for haemoglobin, *HBG*, was duplicated in higher primates by the *L1* jumping gene to generate *HBG1* and *HBG2*. *HBG2* subsequently became switched on specifically in the developing foetus, where it ensures the high oxygen affinity of foetal blood for more efficient oxygen transfer across the placenta. Thus, the important process of gas exchange in the womb has been significantly improved by jumping genes in higher primates, in

contrast to many other mammals, including prosimians, where foetal and adult haemoglobins are the same.

An interesting example of passive gene loss caused by jumping genes was the deletion of 100,000 base pairs of DNA specifically in the gibbon lineage of primates. The culprit behind this genetic mix up was, yet again, the *Alu* sequence and among the genes lost was *ASIP*, which is known to promote the storage of body fat. This may help to explain the wiry build of gibbons, which is so beneficial to their highly active life in the treetops.

Conclusion

A role for jumping genes in evolution has now been recognised by many, yet their importance has often been underestimated. Using primates as an example lineage, the available evidence suggests that jumping genes, via Transposon Thrust, have played an instrumental role in engineering characteristic primate traits and thus have strongly contributed to the divergence of the higher primate lineage away from other types of mammal, including prosimians.

The beneficial features provided by jumping genes in the higher primates include faster brain function, improved foetal nourishment, useful red-green colour discrimination and greater resistance to disease-causing microbes. Such large evolutionary benefits powerfully demonstrate that if jumping genes are “junk DNA” then there is indeed much treasure to be found in the junkyard.

Appendix Three: Jumping Genes: How They Drove Primate Evolution

Keith Oliver is a Biologist and Philosopher in the School of Biological Sciences and Biotechnology and Dr Wayne Greene is Associate Professor of Molecular Genetics in the School of Veterinary and Biomedical Sciences at Murdoch University.

Life's Splendours

Infinite, lovely, and untold,
Life's sacred wonders I declare
More precious far than jewelled gold
Or joyous maidens young and fair;

Whales huge, splash ocean blue,
Harlequin birds carol and sing
Their mating bonds once more renew,
While armoured beetles have a fling:

But death is never ever gone
And agony will ever spill,
As beast slays prey to feed upon,
And dread diseases strike and kill:

Yet life on earth is beautiful,
Pure treasure to enjoy
So quite diverse and wonderful!
A marvellous fount of joy.

© Keith Oliver 2011

Fossils

'The Devil placed fossils in rocks to tempt us' (From a creationist website)

Can't you see him, bright Mephisimo
scratching the curve of his horn into the rock,
the sly segmented bend of his forked tail?
What fun he must have had with
all those awkward animals
the stretched unlikely lizards, the swollen guinea pigs
until the Lord saw what he did, and in horror sent the Flood.

But still he kept on, hopeful Lucifer,
pressing fat stars into stones
as he pressed smaller ones to Galileo's telescope
- eppur si muove -
until God struck the old man blind.

He was with them on the *Beagle*, too,
messing with the finches
tempting them all with barnacles
til the poor lad grew so sick he saw
land rise and fall as queasily as sea.

He's still with us in the labs, young Satan,
whispering in the piled glassware
- Name it after me! -
while the Lord grumbles in the clouds above.

But Life has got away from both of them.
It has tunnelled off in five dimensions,
foxing all their books, dreaming in the ice cores,
and the pulsing membranes of the sun.
We'll find it on Europa next!.

Cecily Scutt 2009

***‘Selection must act on the mechanisms
that generate variation,
much as it does on beaks and bones’***

(Lynn Helena Caporale 2009)

***‘Hypotheses are not statements of
truth,***

***but instruments to be used
in the ascertainment of truth.***

***Their value does not depend
upon ultimate verification,***

***but is to be measured by their effects
upon scientific research.’***

*C. Stuart Gager, University of Missouri 1909
(Translators notes on Intracellular Pangenesis, 1910)*

***‘Doubt is not a pleasant condition,
but certainty is absurd.’***

Voltaire (1694-1778)

'I am convinced that natural selection has been the main but not the exclusive means of modification.'

(Charles Darwin)

'Natural selection is not the all-powerful, all sufficient and only cause of the development of organic forms'

(Alfred Russell Wallace 1901)

'If facts of the old kind will not help, let us seek facts of a new kind'

William Bateson (1861-1926)

'There are still some uncertainties...like the explosive speciation of cichlid fishes...and the stasis of the phenotype in living fossils'

(Ernst Mayr 2004)

'Genomes are not merely passive vehicles of genetic information, but are interactive storage systems'

(James Shapiro 2002)

'Evolvability is a selectable trait'

(David Earl and Michael Deem 2004)

'The union of a vertebrate genome with a viral genome...was potentially far more creative...than the sum of the two components'

(Frank Ryan 2009)

'To conclude that a proposition is true, it is not enough to know that many people find it credible; the proposition itself must be worthy of credence'

(Anonymous)

'The major output of metazoan genomes is non-coding RNA'

(John Mattick 2007)

'Retroposition may represent a dynamic route towards evolutionary progress'

(Jürgen Brosius 1991)

This Thesis represents the culmination of over forty years of mostly intense and passionate interest in evolutionary theory. In my Multidisciplinary Science Degree, I did as many units as I could in this area, and in my Philosophy Degree I wrote extensively on this subject. My Honours Thesis was also about evolutionary theory. In addition to these formal studies I have read widely on the subject, and also had many discussions, sometimes heated, with many people. I am very

grateful to all of the people who have had such discussions or exchanges with me. All of these have helped to shape my present understanding of this very complex area of study.

Curiously, I became interested in evolutionary theory and biological science by a circuitous route. In 1963 I wished to hybridise some different species of an indigenous genus (*Anigozanthos*) of flowering plants commonly called kangaroo paws, so I began reading up on genetics, and polyploidy etc. to help me in this endeavour. This then gave me an interest in biology in general, and especially a passionate interest in evolutionary theory; I expect that this will be my dominant interest for the rest of my life.

The past few decades have been an extraordinarily exciting time in biology, especially with our rapidly expanding flood of data about genomes. Genomes, we are finding, are much more complex than we could ever have imagined, and interact with cells and whole organisms in very complex ways, which may take biologists a very long time to comprehend fully. Greater understanding of genomes could result in major medical, conservation, and social benefits, as well as greatly assist in the further development of evolutionary theory.

In this Thesis, after an introductory chapter on some of the very fascinating history of the realisation of the significance of mobile DNA, I have concentrated on the interactions of Transposable Elements and Endogenised Retroviruses (TEs

and ERVs) with the genotype, and hence with the phenotype. I have proposed, and sought to test, 'The TE-Trust Hypothesis', as a powerful facilitator of evolution, among the many other facilitators for change, that apparently are effective in the process of evolution. It is my hope that the TE-Thrust Hypothesis will help to engender new ways of thinking, and be a positive stimulus to broader future research, in this rapidly developing, and very important study, which constitutes the theoretical basis for understanding the evolution of life on earth.

Keith Oliver 2012