Human Evolution
Genes, Genealogies and Phylogenies

Controversy over human evolution remains widespread. However, the Human Genome Project and genetic sequencing of many other species have provided myriad precise and unambiguous genetic markers that establish our evolutionary relationships with other mammals. *Human Evolution* identifies and explains these identifiable rare and complex markers, including endogenous retroviruses, genome-modifying transposable elements, gene-disabling mutations, segmental duplications, and gene-enabling mutations. The new genetic tools also provide fascinating insights into when, and how, many features of human biology arose: from aspects of placental structure; vitamin C-dependence and trichromatic vision; to tendencies to gout, cardiovascular disease and cancer.

Bringing together a decade’s worth of research and tying it together to provide an overwhelming argument for the mammalian ancestry of the human species, this book will be of interest to professional scientists and students in both the biological and biomedical sciences.

**GRAEME FINLAY** is Senior Lecturer in Scientific Pathology at the Department of Molecular Medicine and Pathology, and Honorary Senior Research Fellow at the Auckland Cancer Society Research Centre, University of Auckland, New Zealand.
Human Evolution
Genes, Genealogies and Phylogenies

GRAEME FINLAY
Department of Molecular Medicine and Pathology,
Auckland Cancer Society Research Centre,
University of Auckland, New Zealand
# Contents

<table>
<thead>
<tr>
<th>Preface</th>
<th>page ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prologue</td>
<td></td>
</tr>
<tr>
<td>1 Darwin's science</td>
<td>2</td>
</tr>
<tr>
<td>2 Genetics arrives on the scene</td>
<td>4</td>
</tr>
<tr>
<td>3 Theological responses to Darwin</td>
<td>6</td>
</tr>
<tr>
<td>4 Interpretations of evolution today</td>
<td>10</td>
</tr>
<tr>
<td>5 Evolution and the genome revolution</td>
<td>12</td>
</tr>
<tr>
<td>6 The scope of this book</td>
<td>18</td>
</tr>
<tr>
<td>1 Retroviral genealogy</td>
<td>21</td>
</tr>
<tr>
<td>1.1 The retroviral life cycle</td>
<td>22</td>
</tr>
<tr>
<td>1.2 Retroviruses and the monoclonality of tumours</td>
<td>26</td>
</tr>
<tr>
<td>1.3 Endogenous retroviruses and the monophylicity</td>
<td></td>
</tr>
<tr>
<td>of species</td>
<td>32</td>
</tr>
<tr>
<td>1.4 Natural selection at work: genes from junk</td>
<td>47</td>
</tr>
<tr>
<td>1.4.1 ERVs and the placenta</td>
<td>48</td>
</tr>
<tr>
<td>1.4.2 ERVs that contribute to gene content</td>
<td>55</td>
</tr>
<tr>
<td>1.5 Natural selection at work: regulatory networks</td>
<td>56</td>
</tr>
<tr>
<td>1.6 Are there alternative interpretations of the data?</td>
<td>58</td>
</tr>
<tr>
<td>1.7 Conclusion: a definitive retroviral genealogy for simian primates</td>
<td>68</td>
</tr>
<tr>
<td>2 Jumping genealogy</td>
<td>70</td>
</tr>
<tr>
<td>2.1 The activities of retroelements</td>
<td>73</td>
</tr>
<tr>
<td>2.1.1 LINE-1 elements</td>
<td>74</td>
</tr>
<tr>
<td>2.1.2 Alu elements</td>
<td>77</td>
</tr>
<tr>
<td>2.1.3 SVA elements</td>
<td>78</td>
</tr>
</tbody>
</table>
vi CONTENTS

2.2 Retroelements and human disease 78
2.3 Retroelements and primate evolution 84
  2.3.1 LINE-1 elements 84
  2.3.2 Alu elements 88
  2.3.3 Retroelements and phylogeny: validation 97
2.4 More ancient elements and mammalian evolution 101
  2.4.1 Euarchontoglires: the primate–rodent group 103
  2.4.2 Boreoeutheria: incorporating the primate–rodent group and the Laurasian beasts 105
  2.4.3 Eutheria 107
  2.4.4 Mammals 111
  2.4.5 TE stories on other branches of the tree of life 114
2.5 Exaptation of TEs 116
  2.5.1 Raw material for new genes 117
  2.5.2 Raw material for new exons 118
  2.5.3 Raw material for new regulatory modules 120
2.6 The evolutionary significance of TEs 124
  2.6.1 TEs, genomic reorganisation and speciation 124
  2.6.2 TEs and evolvability 128

3 Pseudogenealogy 132
  3.1 Mutations and the monoclonal origins of cancers 135
  3.2 Old scars on DNA 138
    3.2.1 Classical marks of NHEJ 139
    3.2.2 LINEs and Alus 141
    3.2.3 NUMTs 142
    3.2.4 Interstitial telomeric sequences 145
  3.3 Pseudogenes 148
    3.3.1 Human-specific pseudogenes 152
    3.3.2 Ape-specific pseudogenes 157
    3.3.3 Simian-specific pseudogenes 163
    3.3.4 Pseudogenes and sensory perception 172
    3.3.5 Pseudogenes from further afield 180
  3.4 Processed pseudogenes 183
### Contents vii

3.5 Rare mutations that conserve protein-coding function  187
3.6 Conclusions  189

4 The origins of new genes  194
4.1 New genes in cancer  195
4.2 Copy number variants  198
4.3 Segmental duplications  201
4.3.1 Some early pointers  201
4.3.2 Systematic studies of SDs  203
4.4 New genes  206
4.4.1 Reproduction  207
4.4.2 Hydrolytic enzymes  218
4.4.3 Neural systems  220
4.4.4 Blood  224
4.4.5 Immunity  228
4.4.6 Master regulators of the genome  236
4.5 Retrogenealogy  238
4.5.1 Reverse-transcribed genes in primates  239
4.5.2 Reverse-transcribed genes in mammals  246
4.6 DNA transposons  249
4.7 De novo origins of genes  254
4.8 Generating genes and genealogies  261

Epilogue: what really makes us human  265
1 Immune systems  267
2 Nervous systems  270
2.1 Critical periods  272
2.2 Learning from neglect  273
3 Features of personhood  277
4 Stories and narrative identity  279

References  284
Index  351
Histories are subject to different interpretations. We would expect biological history to conform to this variety of understandings. But the strange thing is that the very existence of biological history is denied in some quarters. This field of science has acquired a ‘more than scientific’ aura to it. People argue about it as if it were an ideology. Vast resources, including a lot of goodwill, have been expended in the debate. To have achieved this notoriety, we must conclude that biological history (or evolutionary biology) is widely misunderstood. But the evidence for it is there; and a vast volume of fresh genetic data has been added recently. Such data are compelling.

This is a history book, and for two reasons. It attempts to describe, in a very limited and situated sense, a spectacular period in the history of science. Its timeframe covers, with somewhat fuzzy edges, the first decade of the twenty-first century. This is the period during which the human genome sequencing project has been elaborated to ever increasing degrees of detail, and during which myriad fascinating insights into the biological basis of our humanness have been revealed.

Secondly, it describes the evolutionary history of our species, as inscribed in great detail in our genomes. The DNA that we carry around as part of our bodies is an extraordinary library of genetic information. But it is more than simply a blueprint for the human body plan; it also carries, inscribed in its base sequence, a record of its own formative history. Multiple other mammal and vertebrate genomes have also been sequenced over the last decade or so, and this means that we have access to their histories too. When our genomic history is laid out, side by side with those of other species, particular discrete changes in the historical records can be identified.
Preface

in our genome and in the genomes of cohorts of other species. We can thus infer, unambiguously and with a great deal of confidence, that most of our genetic history has been shared with the genetic histories of other primates and, more inclusively, other mammals. Our evolutionary history is well documented.

Molecular evolution is at least as old as the work of Alan Wilson, who used molecular data to infer evolutionary relationships between organisms as long ago as the 1960s. Phylogenetic analyses of DNA and protein sequences have also been used to generate evolutionary trees. Such approaches require expertise in statistics and computation, and require specialist treatments. However, the novel and intuitively appealing approaches surveyed in this book are based, in general, on the identification of particular complex mutations. These arise in unique events. When any such mutation is found in multiple species, it is only because it has been inherited from the one ancestor in which the mutation arose. These are thus very powerful signatures of phylogenetic relatedness.

Along the way, we find out many fascinating things about our biology. We discover that our genome is an entire ecosystem in which semi-autonomous units of genetic material play out their own life cycles. We discover why some people have violent allergic reactions to eating certain animal products. We find out why we must have vitamin C in our diets, whereas other organisms lack this requirement. We learn of the basis of our tendency to suffer from gout. We find clues as to why humans may be particularly cancer-prone. We discover how three-colour vision arose. Indeed many processes through which new genetic functionality has been generated have been laid bare.

Everything that is presented herein is in the public domain. Anything that I have not reported accurately, or that calls for further elaboration, can be fully checked against the source literature. To me, as a cell biologist, the wonder of our DNA-inscribed history is that it requires no logic other than that which is fundamental to all genetics. (Perhaps if I were a palaeontologist, the study of fossils...
would be just as intuitively compelling! But I am not a palaeontologist and I suspect that far fewer people are knowledgeable about fossils than are knowledgeable about the basic mechanisms of heredity.) I believe that the logic of this book will be widely available, although it will require a modicum of biological literacy.

I am very grateful to my superiors in the University of Auckland and the Auckland Cancer Society Research Laboratory, Professors Peter Browett and Bruce Baguley, for allowing me the space and time to work on this book. I thank many senior colleagues who have provided kind and helpful advice: Professor Bill Wilson and Associate Professor Philip Pattemore, Associate Professor Andrew Shelling, Professors Wilf Malcolm, Richard Faull, Malcolm Jeeves and John McClure. Theological input has come from the late Dr Harold Turner, as well as Dr Bruce Nicholls and Dr Nicola Hoggard-Creegan. I am hugely indebted to personnel at the Faraday Institute for Science and Religion, St Edmunds College, University of Cambridge, including Dr Denis Alexander, for sharing their erudition and for their encouragement.

I am deeply grateful to the editorial staff at Cambridge University Press and Out of House Publishing for their unvarying courtesy, patience and helpfulness. It has been a pleasure to work with and learn from them.

I am also grateful to those who have given me scope to work out ideas and evolve ways of expressing them. In particular, I thank the editors of the Paternoster Press periodical Science and Christian Belief, and the multi-author book Debating Darwin: Is Darwinism True & Does it Matter? [2009]. They have allowed me to explore, and reflect upon, earlier phases of an explosively expanding scientific field.