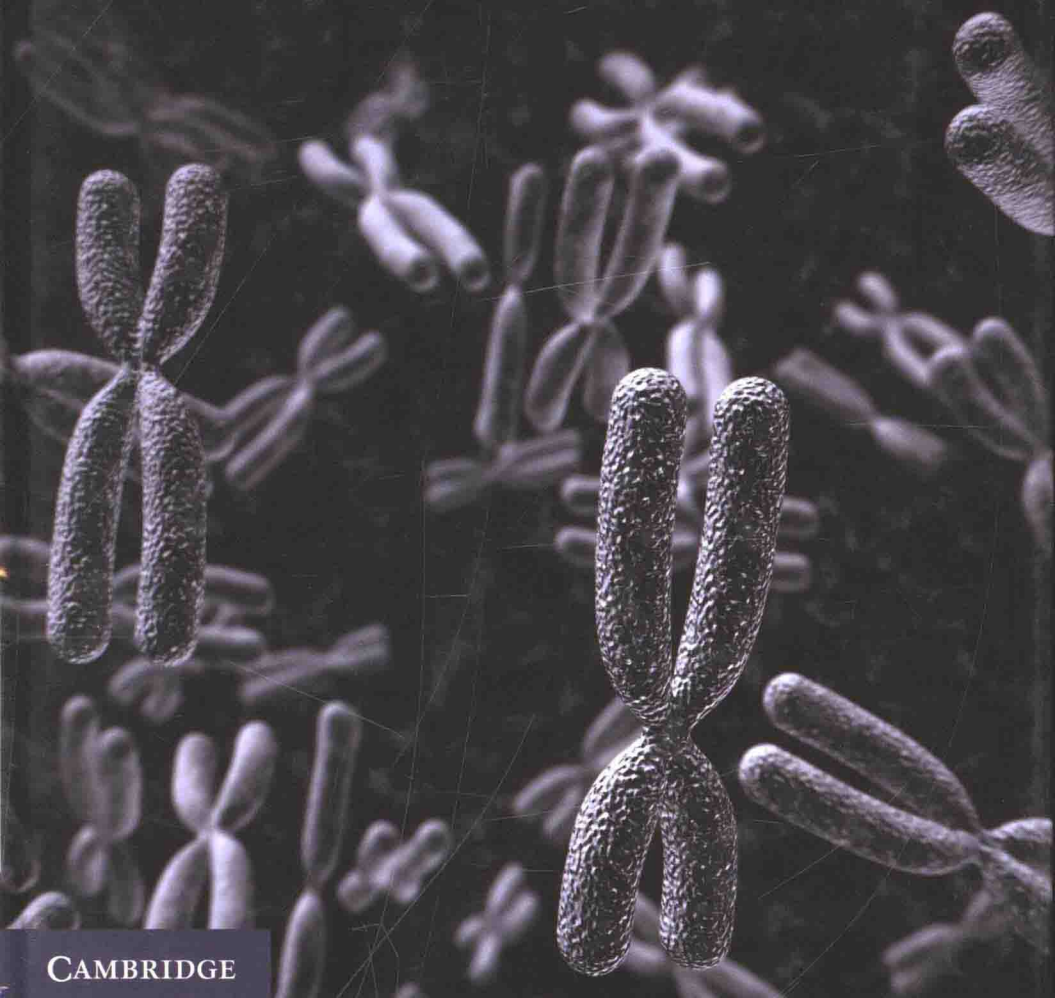
HUMAN EVOLUTION

Genes, Genealogies and Phylogenies

GRAEME FINLAY



CAMBRIDGE

Human Evolution

Genes, Genealogies and Phylogenies

GRAEME FINLAY

Department of Molecular Medicine and Pathology,
Auckland Cancer Society Research Centre,
University of Auckland, New Zealand

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107040120

First published 2013

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Finlay, Graeme, 1953–
Human evolution : genes, genealogies and phylogenies / Graeme Finlay, Department of Molecular Medicine and Pathology, Auckland Cancer Society Research Centre, University of Auckland, New Zealand.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-04012-0 (hardback)
1. Human evolution. 2. Human population genetics. 3. Evolutionary genetics. 4. Genetic genealogy. I. Title.
GN281.F54 2013
599.93'8–dc23 2013015863

ISBN 978-1-107-04012-0 Hardback

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.

Preface

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

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.

Contents

Prologue

Charles Darwin did not discover biological evolution. The concept had been brewing in people's minds for decades and Darwin grew up in an ambience of evolutionary speculation. His own grandfather, Erasmus, who died seven years before Charles was born, had ventured the possibility that all warm-blooded animals had evolved from a single ancestor. Erasmus undoubtedly had a great influence on his grandson through family links and his book *Zoonomia*.

In the first half of the nineteenth century, many biologists propounded the idea that humans had evolved from single-celled microbes. The physician-turned-biologist Robert Grant embraced evolutionary ideas from both Erasmus Darwin and the French evolutionary theorist Lamarck (who had proposed that organisms generated adaptive responses when presented with environmental challenges, and that these were heritable). Grant, in turn, passed these ideas on to the young Charles Darwin when he was studying medicine at Edinburgh. Grant then moved to University College London where he continued to popularise evolutionary thinking.

A book promoting the idea that humans evolved from simple ancestors (*Vestiges of the Natural History of Creation*) was published in 1844. It was published anonymously, but was later revealed as the work of a journalist, Robert Chambers. It was derided by its reviewers, but remained hugely popular during the rest of the nineteenth century. The philosopher Herbert Spencer (who coined the term 'survival of the fittest') also wrote on themes of human and social evolution. Spencer contributed to the wider intellectual environment of receptivity to evolutionary ideas. These works prepared popular thinking for Darwin's *Origins* when it was finally published in 1859 [1].

1 DARWIN'S SCIENCE

Darwin was the first to offer a plausible *mechanism* for evolutionary development [2]. In this he was closely followed by Alfred Russel Wallace, who had spent time exploring the Amazonian and South East Asian rainforests. The outline of this scheme, known as *natural selection*, is elegantly simple.

- Resource limitations will always prevent a population from increasing at the rate that it is potentially capable of. In every generation, the individuals that become parents are a subset of the individuals that were born into that generation.
- The individuals of a species vary in many features. When a population is presented with environmental challenges or opportunities, the individuals endowed with variations that enable them to best tolerate or exploit those conditions will have a better chance of producing offspring. Parents are a *selected* group.
- Offspring tend to inherit their parents' characteristics. Features conferring reproductive success will become progressively more widely represented or more strongly developed in the population. Continuously changing conditions will drive the continuous modification of the biological features possessed by populations.

Darwin drew parallels between natural selection and the *artificial selection* performed by breeders of domesticated plants and animals. The characteristics of cereals and fruits, and of dogs and horses, are progressively altered as breeding is limited to those individuals that display the characters people desire. A spectacular example (not known to Darwin) is the way in which humans transformed the grass teosinte into maize in a few thousand years. The kernels of teosinte are few (no more than a dozen per ear), attached to long stalks and protected by a hard case. The kernels of maize are many, attached to a cob (peculiar to maize) and unprotected. A large number of genes underwent selection during the transformation from teosinte to maize [3]. Dramatic as these effects are, the particular features established by selective breeding are retained only as long as the appropriate selective pressures are applied.

Darwin identified another source of selection known as *sexual selection*. Male and female individuals of a species are often highly distinctive. The sexual dimorphism of the Indian peafowl is a classical example. In such cases, the factor driving evolutionary change is a behavioural one: choice by potential mates. The genes favoured in the case of the peacock are genes for glamour, not for usefulness.

Darwin developed many other insights that have been validated subsequently. He promoted the idea of common descent, ultimately represented by the image of a single tree of life. He perceived that an authentic taxonomic system simply reflects the branching patterns of this tree, and that extant species are a mere sample of all those that have existed, because of the wholesale extinction of linking intermediate species. He accounted for the geographical distributions of species in terms of patterns of adaptive radiation, according to which organisms evolve to take advantage of all available habitats.

He developed the concept of the vastness of time required for evolution. He accepted that the concept of gradual evolutionary change encompasses stepwise innovations, anticipating the discovery of punctuated equilibrium in the late twentieth century. Other areas of Darwin's prescience included the concerted evolution of mutually interacting species (*co-evolution*). He recognised that complex interactions occur between species (the economy of nature), and so anticipated ideas that would find their place in the science of ecology.

Darwin compiled a huge volume of evidence supporting his evolutionary paradigm. Such evidence featured comparative anatomy, physiology and behaviour, the illuminating – but necessarily incomplete – fossil record, the geographical distributions of plants and animals, and analogies with artificial breeding. These approaches have been the staple of evidential discussion (almost) to the present day [4]. The cumulative evidence for evolution was impressive, but inherently circumstantial. No-one had seen a wing evolve.

But the idea of natural selection faced one huge hurdle. Darwin knew no genetics. He did not know how heredity worked. He and most of his contemporaries considered that hereditary information was somehow distilled from throughout the parents' bodies and imprinted on to the appropriate sites of the developing embryo. This system of inheritance entailed that distinctive parental characteristics would be blended in their offspring. Such blending of inherited features engendered an unfortunate consequence. Useful adaptations would be diluted out with each succeeding generation, and ultimately lost. This was argued cogently on mathematical grounds by Fleeming Jenkin in the late 1860s.

Blending inheritance presented what appeared to be an intractable problem to Darwin's theory. As he wrestled with it, he reverted increasingly to the idea that environmental challenges could induce adaptive features in organisms, and that these were transmissible to the next generation. To get around the problem of blended inheritance, he suggested that environmental conditions might affect all the individuals in a population in a concerted manner. For much of his life, Darwin was more a Lamarckian than a Darwinian [5].

2 GENETICS ARRIVES ON THE SCENE

In the early 1900s, Gregor Mendel's work was rediscovered. It provided a first hint of the existence of units of inheritance that would later be known as genes. The answer to the problem of blending inheritance is that inheritance is quantised. Darwinian evolution only became established in the 1920s with the synthesis of natural selection and genetics. But the biochemical substance that acted as the repository of genetic information remained unknown until 1944. In that year, the material of inheritance was shown to be a constituent of cells, called DNA. People had not thought DNA particularly interesting up until that time.

In 1953, James Watson and Francis Crick proposed a model of the chemical structure of DNA, and revealed how it could embody genetic information. A DNA molecule contains myriad

chemical units called *bases*, arranged in linear sequence, which are information-bearing. Watson and Crick showed how DNA could be faithfully copied and transmitted from generation to generation. And their model revealed – at last! – how DNA could undergo structural changes that would account for heritable (and non-blending) variation. Changes in the chemical units (and information content) of DNA would be transmitted from parents to their children, and thence to succeeding generations.

An important corollary of the heritability of DNA variants is that particular novelties in genetic information identify organisms connected by descent. DNA constitutes a record of family relationships. Indeed, the genetic information inscribed in DNA is an archive of long-term (evolutionary) histories. But a digression is first necessary. This book is written for biologists, and for people in medical and allied sciences who are familiar with biological concepts. But, hopefully, it will be read by all sorts of interested people – teachers, students, pastors and theologians – and so the conventions used to depict the nature of genetic information should first be reviewed.

The DNA double helix is an icon of biology. DNA consists of two helical strands, each of which consists of a backbone from which projects a succession of bases. There are four different bases, designated A (adenine), T (thymine), G (guanine) and C (cytosine). Each base hanging off one backbone interfaces with a base hanging off the opposite backbone. But size and shape considerations mean that A must pair with T, and G must pair with C. In a moment of exhilarating intuition, Watson perceived how this arrangement underlies the mechanism of heredity. Genetic information is inscribed in the order (or *sequence*) in which the bases occur. If the two strands of a DNA molecule (each backbone with its bases) are separated, the base pairing rules ensure that each is able to direct the synthesis of a new strand with its ordered complement of bases. One double helix generates two identical double helices. When cells divide, the DNA of the parent cell is duplicated and an identical copy bequeathed to each daughter cell.

Conceptually, we can unwind the double helix to produce a ladder in which the rungs are the base pairs. By convention, we read the base sequence of the top strand, as set out for the hypothetical sequence below, from left (designated 5′) to right (designated 3′). The bottom strand is read in the opposite direction. If we are thinking about gene sequences, the top strand is called the *coding* or *sense* strand (again, conventionally), because this is the sequence that specifies the order in which amino acids are added to make proteins.

Coding strand:	5′-CATATTACATAGGA-3′
Non-coding strand:	3′-GTATAATGTATCCT-5′

The most economical way of depicting genetic sequence is to present the coding strand, CATATTACATAGGA. We do not need the 5′ or 3′ signs, because we know it reads from left to right; nor do we need to write out the complementary base sequence, because we know that A, T, G and C must specify T, A, C and G as their respective complements. It is in this minimalist form that genetic sequences may be portrayed.

3 THEOLOGICAL RESPONSES TO DARWIN

Humanity had formulated no plausible scientific theory to account for the development of new species (including humans) and the diversity of life forms until Darwin. In the absence of scientific knowledge, the default position had been to account for *physical* realities (the adaptations and diversity of organisms) by using *metaphysical* concepts. It was sufficient to say that living species possess their particular constellations of characteristics because God made them that way. But such reasoning transgresses category boundaries.

The Darwinian revolution exploded this long-held conflation of concepts. The spectacular diversity of life was for the first time explained in physical cause-and-effect terms. The development of evolutionary theorising simply illustrated the dictum that scientific questions require scientific answers. Theologians had to rethink

the relationship between the God whom they perceived as being at work in human history, and physical or biological mechanisms. The question of whether the cosmos was *creation* had to be accepted (or rejected) on the basis of considerations other than scientific ones.

Theologians had to recognise that the biblical concept of 'creation' referred to *ontological* origin (God creates all things at all times), not *temporal* origin (God creates particular things at particular times) [6]. A biblical creator had to be understood as the cause of everything but scientifically the explanation of nothing [7]. Such a creator could not be conceived as a component of, or an alternative to, any scientific formulation. No process – and certainly no aspect of cosmic or biological history – could be out of bounds to empirical investigation. The created order had an authentic evolving history [8], and such histories were open to empirical investigation, and on their own terms.

Many Christians accommodated their thinking to Darwin's new scientific paradigm. Darwin agreed with the Reverend William Whewell, Master of Trinity College, Cambridge (and inventor of the word *scientist*), that in the material world, 'events are brought about not by insulated interpositions of divine power, exerted in each particular case, but by the establishment of general laws' (1859). The Reverend Charles Kingsley (later Professor of History at Cambridge) articulated similar sentiments: it is 'just as noble a conception of Deity, to believe that he created primal forms capable of self-development' as to believe that God had to make a fresh act of intervention to fill every taxonomic gap (1859).

Darwin was religiously agnostic but advocated strategies of reconciliation. He did not see how evolution should shock the religious feelings of anyone. His chief supporter in America was the Christian, Asa Gray (Professor of Natural History at Harvard). They shared the conviction that evolution was 'not at all necessarily atheistical' (1860). Towards the end of his life, Darwin rejected (in private correspondence) any reason why the disciples of religion and of science 'should attack each other with bitterness' (1878). He stated that