

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings

Dan Dediú

Frontmatter

[More information](#)

An Introduction to Genetics for Language Scientists

During the last few decades we have discovered an enormous amount about our genomes, their evolution and, importantly for linguists and language scientists, the genetic foundations of language and speech.

Accessible and readable, this introduction is designed specifically for students and researchers working in language and linguistics. It carefully focuses on the most relevant concepts, methods and findings in the genetics of language and speech, and covers a wide range of topics such as heritability, the molecular mechanisms through which genes influence our language, and the evolutionary forces affecting them.

Filling a large gap in the literature, this essential guide explores relevant examples including hearing loss, stuttering, dyslexia, brain growth and development, as well as the normal range of variation. It also contains a helpful glossary of terms and a wide range of references so the reader can pursue topics of interest in more depth.

DAN DEDIÚ is Senior Investigator in the Language and Genetics Department at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands.

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts,
Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts,
Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

An Introduction to Genetics for Language Scientists

Current concepts, methods and findings

Dan Dediu



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

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/9781107001299

© Dan Dediu 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Printed in the United Kingdom by TJ International Ltd., Padstow, Cornwall

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

ISBN 978-1-107-00129-9 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts,
Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

Dedicated to my mother, Maria Dediu, who managed to
foster curiosity, interest in science and optimism even in the
darkest of times.

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts,
Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

Contents

<i>List of illustrations</i>	<i>page xi</i>
<i>Acknowledgements</i>	<i>xiii</i>
1 Introduction	1
1.1 Why is genetics relevant for me?	4
2 Nature, nurture, and heritability	7
2.1 Phenotype, genotype and environment	7
2.2 Innateness, a slippery and complex concept	9
2.3 Some basic notions of statistics	12
2.3.1 Mean, variance and standard deviation	14
2.3.2 The normal distribution	15
2.3.3 Statistical populations and samples	19
2.3.4 Covariance and correlation	21
2.3.5 Regression	24
2.4 Basic notions of quantitative genetics	30
2.4.1 Partitioning the phenotype and broad-sense heritability, H^2	31
2.4.2 Partitioning the genotype and narrow-sense heritability, h^2	32
2.4.3 Partitioning the environment	33
2.4.4 Estimating heritability	34
2.4.5 Heritability: what it does and does not mean	37
2.4.6 The heritability of language and speech	39
2.4.7 Relationships between genotype and environment	40
2.4.8 Sharing genes between phenotypes: genetic correlations and “generalist genes”	42
3 The molecular bases of genetics	44
3.1 We are composed of cells	45
3.2 The molecules of life	48
	vii

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

viii	Contents	
3.3	The genetic material	49
3.4	DNA: storing and transmitting information	50
3.5	Chromosomes are DNA molecules	53
3.6	Genetic loci and recombination	58
3.7	What is a gene?	59
3.7.1	Decoding genes: transcription and translation	61
3.7.2	Genes have structure: introns and exons	68
4	Effects of genes on phenotype	72
4.1	Dominance and recessiveness	72
4.2	Autosomal dominance and recessiveness	75
4.2.1	A dominant speech and language disorder	75
4.2.2	Recessive hearing loss	80
4.3	Sex-linked dominance and recessiveness	83
4.3.1	X chromosome inactivation	84
4.3.2	Anomalous colour perception	84
4.3.3	Variation in colour perception and language	88
5	Linkage disequilibrium and its role in finding genes	90
5.1	What is linkage disequilibrium?	90
5.2	Using linkage disequilibrium	93
5.3	Association studies	94
5.3.1	Statistical concepts: power, multiple testing correction and effect size	97
5.3.2	Controlling for other factors using regression	110
5.3.3	Accounting for population stratification	113
5.3.4	Replication as the gold standard	115
5.3.5	Combining multiple markers	116
5.3.6	The Transmission Disequilibrium Test	118
5.3.7	Examples of association studies for language and speech	120
5.4	Linkage studies	124
6	What do genes actually do?	131
6.1	Structural proteins: <i>TECTA</i> in the inner ear	131
6.2	Energy production: mitochondrial gene <i>MTRNR1</i>	133
6.3	The transportation system: <i>MYO15A</i>	134
6.4	Unexpected processes: stuttering and the lysosome	135
6.5	Guiding axons: <i>ROBO1</i> in dyslexia and normal variation	137
6.6	Brain growth and development: <i>ASPM</i> and <i>MCHP1</i>	139
6.7	Regulating the expression of other genes: <i>FOXP2</i>	142
6.8	Neural development and neural impulse conduction: <i>CNTNAP2</i>	149

Contents	ix
6.9 Post-transcriptional gene regulation: microRNAs	151
6.9.1 microRNAs and <i>FOXP2</i>	151
6.9.2 <i>miR-96</i> and hearing loss	152
7 The way forward: exome and genome sequencing	153
7.1 Exome and genome sequencing	153
7.2 The “missing heritability” and the genetic architecture of complex traits	155
8 Population and evolutionary genetics	160
8.1 Foundations of population genetics: loci, alleles, individuals and populations	161
8.2 A useful baseline: the Hardy-Weinberg equilibrium	162
8.3 Genetic drift: the power of chance	164
8.4 Mutation: the creator of diversity	168
8.5 Selection: when differences do matter	170
8.5.1 Positive, stabilizing and disruptive selection	171
8.5.2 Hiking the fitness landscape	173
8.5.3 When selection seemingly fails	175
8.5.4 What is selected: de-focusing the individual	177
8.5.5 The (nearly) neutral theory of evolution	178
8.5.6 Detecting selection from genetic data	179
8.5.7 The evolutionary tale of <i>FOXP2</i>	185
8.6 Population structure: love is not blind	187
8.6.1 Inbreeding: mating among genetic relatives	187
8.6.2 (Dis-)assortative mating: choosing partners (un)like yourself	188
8.6.3 Structured populations: choosing the partners you can actually choose	188
8.6.4 The genetic structure of modern humans	190
8.7 Recent and ongoing evolution in humans	195
8.7.1 Skin colour	195
8.7.2 Hair, sweat and ear wax	196
8.7.3 Living at high altitude	197
8.7.4 Is culture shielding or selecting us?	198
9 Interactions between genetic and cultural evolution	199
9.1 Fighting pathogens	200
9.2 Eating well	202
9.3 Creating new languages	204
9.4 Genetically biased cultural evolution	207
9.4.1 Genetic biases in language	209
10 Conclusions, topics not covered, future directions	212
10.1 Resources for further study	213

Cambridge University Press
978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts,
Methods and Findings
Dan Dediu
Frontmatter
[More information](#)

x	Contents	
	Appendix The computer code	216
	A.1 Simulating contingency tables	216
	A.2 Log odds ratio	217
	A.3 Simulating genetic drift	218
	<i>References</i>	221
	<i>Index</i>	252
	<i>Glossary</i>	257

Illustrations

2.1	The distribution of height in 25,000 children	13
2.2	Probability density function of the normal distribution	17
2.3	The Central Limit Theorem	19
2.4	The distribution of weight in 25,000 children	21
2.5	Relationship between height and weight in 25,000 children	22
2.6	Relationship between height and weight in 1,035 baseball players	25
2.7	Pairwise relationships between weight, height and age in 1,035 baseball players	27
2.8	3D plot of the relationship between weight, height and age in 1,035 baseball players	27
2.9	Regression to the mean	29
2.10	Estimating heritability in a twin design	36
3.1	The structure of a generalized animal cell	47
3.2	The Tree of Life	47
3.3	DNA replication	52
3.4	Point mutation and its transmission	53
3.5	The position of <i>FOXP2</i> on Chromosome 7	54
3.6	Gametogenesis	56
3.7	Fertilization	57
3.8	Chromosomal crossover	58
3.9	The structure of a chromosome	60
3.10	The structure of the simplest imaginable gene	61
3.11	The transcription of DNA into messenger RNA	63
3.12	The translation of mRNA into protein	66
3.13	Reading frames	66
3.14	The structure of a more complex gene: exons and introns	68
3.15	Alternative splicing	68
3.16	Promoters, enhancers, suppressors and insulators	70
4.1	Transmission of flower colour in the sweet pea	73
4.2	Pedigree of the KE family	76
4.3	Pedigree of the Bengkala kindred K1	81

xii	Illustrations	
5.1	The indirect relationship between marker and phenotype	94
5.2	Statistical power of the χ^2 test	106
5.3	Regression approach to genetic association	112
5.4	Family trios illustrating the transmission disequilibrium test	119
5.5	KE family pedigree with genotype information	125
5.6	Example of linkage analysis	127
6.1	The human ear	132
6.2	Transporting enzymes to the lysosome	136
6.3	Neurogenesis: symmetric and asymmetric divisions	140
6.4	The structure of <i>FOXP2</i> and some mutations	146
6.5	Abstract gene regulatory network (downstream)	149
6.6	Gene regulation (up-regulation and down-regulation)	150
8.1	Population, individuals, loci and alleles	161
8.2	Genetic drift function of population size	165
8.3	Population bottlenecks and founder effect	167
8.4	Mutation when only two alleles can exist	169
8.5	Positive selection	171
8.6	Stabilizing selection	172
8.7	Heterozygote advantage	174
8.8	Fitness landscape	175
8.9	Selective sweep	183
8.10	The serial founder effect	193
9.1	Genetic biases in language	210

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

Acknowledgements

This book is the result of a long process that began more than five years ago, with Helen Barton visiting the University of Edinburgh and convincing the young and naïve freshly minted doctor that I was to transform his PhD thesis (a monster of some 460 pages) into a readable book. The idea was that there were no introductions to genetics written for scientists interested in language and speech and the available general introductions were either too long and dense or did not really appeal to the interests of these readers.

Soon after this meeting I moved to Nijmegen, the Netherlands, joining the world-renowned Max Planck Institute for Psycholinguistics with the specific task to begin working on the genetics of language and speech, trying to build upon the enormous amount of expertise scattered across the various departments of the institute, paving the way as it were for the upcoming new department of Language and Genetics. This task was extremely interesting as it effectively meant bridging the gaps between specialties, and I worked with neuroscientists, “classical” psycholinguists, quantitative historical linguists and typologists (for lack of a better name), specialists in child language acquisition and even the odd field linguist. This taught me invaluable lessons about what is interesting, relevant and accessible to all these scientists, and made me discover topics that I did not even know existed and are potentially so important for the genetics of language and speech (such as the emerging village sign languages and the genetics of hearing deficits; thank you Connie). Likewise, teaching several introductory courses in various summer schools and giving invited lectures at conferences and workshops put me in direct contact with the maddeningly complicated and devilishly smart audience this book is designed for, and forced me to sharpen my explanations, pick the right examples and hedge my theoretical positions.

But it was no easy job. Trying to convey the importance of *differences* (as opposed to the always assumed ideal participant that one approximates by having dozens of imperfect replicas in an experimental condition) was a main stumbling block, as was dispelling the fear that what I am “really” interested in was to classify people based on their genes (no, geneticists are some of the

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings

Dan Dediu

Frontmatter

[More information](#)

xiv Acknowledgements

most vehement opponents of classifications as they experience first-hand, day-by-day the gradualness and complexity of human variation), or, equally worse, to find genetic differences that are – therefore one would say – unchangeable, reinforcing the differences between those who have and those who don't (of course, if modern genetics and evolutionary theory teaches us anything it is that the environment is *essential* to everything genetic to the point of blurring this apparently clear-cut distinction).

Of course, as any cursory look at my CV shows, I am no geneticist by training, but I have always been interested in biology (literally, I remember reading about evolution in my 4th or 5th grade thanks to my mother, a biology teacher) and I have been working on questions to do with evolution, genetics and language for the past 10 years or so. The good news is that you don't need to have had biology classes to be able to understand it, but the downside is that, as a self-taught geneticist, you will always have gaps in your knowledge. Thus, I am particularly indebted to my new colleagues – the “real geneticists”, as I sometimes call them – in the Language and Genetics department here, at the Max Planck Institute for Psycholinguistics (whose work oft-times fills me with awe), especially to Prof. Simon Fisher, Dr. Sonja Vernes, Dr. Clyde Francks, Dr. Sarah Graham, Dr. Tulio Guadalupe, Dr. Nicolas Brucato, Paolo Devanna, Alessandro Gialluisi, and Amaia Carrión Castillo.

The Max Planck Institute for Psycholinguistics in Nijmegen is an amazing place for inter-disciplinary work where we somehow manage to talk across discipline boundaries and effectively collaborate with each other, a feat that is much more difficult to accomplish than one would naïvely think. It has also financially supported me through most of the time spent writing this book and its library allowed me to access all the journals, papers and books I needed to; many thanks to the librarians and especially to Karin Kastens. My current project here, the so-called “G[3]bils” (Genetic biases in language and speech, funded by the Netherlands Organisation for Scientific Research [NWO] through the VIDI project 276-70-022, together with Dr. Scott Moisiik and Rick Janssen) is a living example of how taking into account genetics in an inter-disciplinary framework can shed light on patterns of diversity in phonetics and phonology.

I have learned a lot from discussions with many colleagues across the years, but I want to mention in particular Prof. Stephen C. Levinson, Prof. Simon Fisher and Prof. D. Robert Ladd (University of Edinburgh, UK), who very nicely illustrate how one can successfully bridge multiple disciplines at once and yet still manage to be at the forefront of one's own scientific fields.

Several colleagues took their time to check parts of the manuscript, make many valuable suggestions and point me in the right direction, and I would like to thank Prof. Simon Fisher, Dr. Sonja Vernes, Dr. Clyde Francks, Dr. Sarah Graham, Dr. Tulio Guadalupe, Dr. Nicolas Brucato, Alessandro Gialluisi,

Cambridge University Press

978-1-107-00129-9 - An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings

Dan Dediú

Frontmatter

[More information](#)

Acknowledgements

xv

Amaia Carrión Castillo, Dr. Scott Moisik and Dr. Connie de Vos. I also want to thank one anonymous reviewer who had very important suggestions for improving the book, the most notable being a better coverage of quantitative genetics. Of course, any remaining errors and infelicities are mine!

This would have not been possible without the support of my wife, Dr. Alexandra Dima, who in fact got me to learn statistics in the first place and with whom I would often discuss not only data analysis strategies, but the deeper meaning and assumptions behind various statistical techniques and what *not* to read in a reported result (despite what is claimed in the abstract and conclusions), a very important skill to have. At times she encouraged (and even forced) me to continue working on this project and her suggestions and feedback made this a much better book than it would have otherwise been. Thank you so much!

A special “Dank jullie wel!” to Bernadette and Herman Arts without whom our adaptation to the Netherlands would have been so much more difficult (and we would have indubitably discovered the local cherries from Bommel much later).

Finally, this book would have been much more difficult to write without all the *Open-Source* software I used throughout the years and I wish to thank all those who have contributed, one way or the other, to specific software packages and to the philosophy that makes this possible. My operating system has always been a flavour of GNU/Linux, including Debian (<http://www.debian.org>), Ubuntu (<http://www.ubuntu.com>), ArchLinux (<http://www.ubuntu.com>), OpenSuse (<http://www.opensuse.org>) and CentOS (<http://www.centos.org>). The book was written in L^AT_EX using initially Kile (<http://kile.sourceforge.net>) but mostly TexMaker (<http://www.xmlmath.net/texmaker>) and TexStudio (<http://texstudio.sourceforge.net>) as editors/environments. All data analyses for this book used R (R Development Core Team, 2010), an open-source programming environment focused on statistics available free of charge at <http://www.r-project.org>. The graphs were generated using a wide variety of tools, including R, LaTeXDraw (<http://latexdraw.sourceforge.net>), Inkscape (<http://inkscape.org>), the L^AT_EX packages pstricks, pst-pdgr (pedigrees), mhchem (chemical formulas), xy (blocks and arrows diagrams), and amssymb, MnSymbol, wasysym and tipa (for various symbols including IPA), and even with LibreOffice’s (<http://www.libreoffice.org>) own Draw component.