Our Genes

Situated at the intersection of natural science and philosophy, *Our Genes* explores historical practices, investigates current trends, and imagines future work in genetic research to answer persistent, political questions about human diversity. Readers are guided through fascinating thought experiments, complex measures and metrics, fundamental evolutionary patterns, and in-depth treatment of exciting case studies. The work culminates in a philosophical rationale, based on scientific evidence, for a moderate position about the explanatory power of genes that is often left unarticulated. Simply put, human evolutionary genomics – our genes – can tell us much about who we are as individuals and as collectives. However, while they convey scientific certainty in the popular imagination, genes cannot answer some of our most important questions. Alternating between an up-close and a zoomed-out focus on genes and genomes, individuals and collectives, species and populations, *Our Genes* argues that the answers we seek point to rich, necessary work ahead.

RASMUS GRØNFELDT WINTHER is a philosopher of science, researcher, writer, educator, diver, and explorer. He is Professor of Humanities at University of California, Santa Cruz and Affiliate Professor of Transformative Science at the GLOBE Institute at University of Copenhagen.

Photo credit: Marie Raffn

Our Genes

A Philosophical Perspective on Human Evolutionary Genomics

RASMUS GRØNFELDT WINTHER

Professor of Humanities at University of California, Santa Cruz

Affiliate Professor of Transformative Science
at the GLOBE Institute at University of Copenhagen

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

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/9781107170407
DOI: 10.1017/9781316756324

First published 2022

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall

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

Library of Congress Cataloging-in-Publication Data
Names: Winther, Rasmus Grønfeldt, author.
Title: Our genes : a philosophical perspective on human evolutionary genomics / Rasmus Grønfeldt Winther.
Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2022. | Includes bibliographical references and index. | Contents: Introduction – Origins and Histories – The Mind, the Lab, and the Field : Three Kinds of Populations – Metrics and Measures – Models and Methodologies – Six Patterns of Human Genomic Variation – Natural Selection – Intelligence, Female Orgasm, and Future Discovery – Is Race Real? – The Conscious Universe : Genes in Complex Systems.
Identifiers: LCCN 2021029747 (print) | LCCN 2021029748 (ebook) | ISBN 9781107170407 (hardback) | ISBN 9781316621509 (paperback) | ISBN 9781316756324 (epub)
Subjects: MESH: Genomics | Philosophy | Genetic Phenomena | Human Genetics | Genome, Human
Classification: LCC QH447 (print) | LCC QH447 (ebook) | NLM QU 460 | DDC 572.8/6–dc23
LC record available at https://lccn.loc.gov/2021029747
LC ebook record available at https://lccn.loc.gov/2021029748

ISBN 978-1-107-17040-7 Hardback
ISBN 978-1-316-62150-9 Paperback

For Us.

And with extreme gratitude to
Aage Bisgaard Winther, Helen Longino, and Amir Najmi.

experienced the most bottlenecks (though Oceanian populations experienced almost as many, some of them nonoverlapping with Indigenous Americans; see Figure 2.3). Over time, random genetic drift and genetic bottlenecks tend to lead to loss of (a) nucleotide diversity, (b) alleles of various kinds, and (c) genetic heterozygosity. It is therefore Africa, the second-largest continent on Earth, that retains the greater, or ancestral genetic variation, making it the putative capital of Planet Unity (see Figures 4.1, 6.1, and 6.3).

4. Most Genetic Variation Is Among Individuals Within Populations, Not Across Populations Within Continental Regions, Nor Across Different Continental Regions (Lewontin's Distribution)

What I have called *Lewontin's distribution* in Chapters 4 and 5 helps answer the question of how much less genetically different two randomly chosen individuals from the same population are, on average, than a randomly chosen individual from that population and a randomly chosen individual from *another* population – whether from the same continental region or from anywhere on the globe.

Lewontin deployed data from especially Giblett (1969) and Mourant (1954) in his influential 1972 study, and from Cavalli-Sforza and Bodmer (1971) in his 1974 book, wherein he partly corrected calculation errors from his earlier results.[32] Lewontin (1972) drew on reasonably extensive allele frequency data tables of serum proteins and red blood cell enzymes[33] and blood groups,[34] including two blood lipoproteins, *Duffy*. the Rh factor, and the ABO blood group complex. With varying degrees of breadth (numbers of populations sampled) and depth (numbers of individuals sampled in each population), Lewontin collected – and sometimes also had to finish calculating – global distributions of the allele frequencies of each of 17 genes or loci. He then used his particular form of the technical methodology of variance partitioning, which we reviewed in Chapter 5, on the allele variation for each of these genes or loci, doing so within populations, among populations on the same continent, and among continents.

[32] The earliest published data that explicitly give something close to this statistical result can be found in Table 12 in Cavalli-Sforza (1966, p. 367), which in column 4 indeed "showed that only about 15% of the overall gene-frequency variation occurred between populations compared with 85% within them" (Bodmer, 2018, p. 322).

[33] Giblett (1969). [34] Mourant (1954).

Table 6.1. *Allele frequencies of three distinct genes across continental regions*[35]

Gene	Alleles	Africa	East Asia	Europe
Duffy	*Fy*	0.94	0.10	0.03
	Fy^a	0.06	0.90	0.42
	Fy^b	0.00	0.00	0.55
Auberger	Au^a	0.64		0.62
	Au	0.36		0.38
Xg	Xg^a	0.55	0.54	0.67
	Xg	0.45	0.46	0.33

Frequencies are rounded from four to two significant figures. Empty cells indicate lack of data. Following Lewontin (1974), the "a" and "b" superscripts differentiate alleles.

Table 6.1 focuses on three major continental regions to show three types of allele distributions, represented by three genes, that gave rise to the true 85.7%/7.1%/7.2% distribution of Lewontin's analysis. His results imply that at most variable loci, different human groups tend to have relatively similar allele frequencies. Thus, the *Duffy* gene is an atypical example, as it is more extremely diverged than average – 0.94% (Africa)/ 0.10% (East Asia)/0.03% (Europe) for the *Fy* "null" allele; based on similar Lewontin (1972) allele frequency data, *Duffy* has a true diversity apportionment of 63.6%/10.5%/25.9% (Table 6.2), which is a much higher among continents variance component than almost all other loci Lewontin studied in 1972 (think Galápagos-Writ-Large).[36] In contrast, *Auberger* indicates less variation across populations than the average human locus (think Planet Unity). The *Xg* gene is typical of the human

[35] From Cavalli-Sforza and Bodmer (1971), as presented in Table 33 in Lewontin (1974), a table with only these three regions represented (p. 153). As we saw in Chapter 5, Lewontin (1972) analyzed allele frequencies globally, including, at the highest continental regional level, South Asian Aborigines, Indigenous Americans, Oceanians, and Australian Aborigines, in addition to African, East Asian, and European populations (cf. Winther, 2022a). Depicting three major continental regions in this table is, however, sufficient for illustrating the range of allele frequency differences across continental regions – from extremely different (e.g., *Duffy*) to highly similar (e.g., *Auberger*). Figure 3 in Rosenberg (2011) selects three microsatellite loci with analogous levels of allele frequency geographic differentiation, with *Duffy* corresponding to D12S2070 (bottom row), *Auberger* to D6S474 (top row), and *Xg* to D10S1425 (middle row) (p. 664).

[36] See Table 1.6 in Winther (2022a, p. 36). The allele distribution for *Duffy*, and concomitant diversity apportionment, is unsurprising, as homozygous *Fy* individuals are resistant to the malarial parasite *Plasmodium vivax*, historically common in Africa and, to a lesser extent, in Southeast Asia (see, e.g., Szpak et al., 2019, pp. 1432–1435). This allele has thus been under selection, a topic to which we turn in Chapter 7.

Table 6.2. *Two genetic diversity apportionments from Lewontin (1972): Lewontin's and the true one*[37]

	Within populations		Among populations		Among races	
	Lewontin	True	Lewontin	True	Lewontin	True
Hp	0.893	0.893	0.051	0.050	0.056	0.057
Ag	0.834	0.835	–	0.003	–	0.162
Lp	0.939	0.942	–	0.025	–	0.033
Xm	0.997	0.997	–	0	–	0.003
APh	0.927	0.919	0.062	0.059	0.011	0.023
6PGD	0.875	0.877	0.058	0.055	0.067	0.068
PGM	0.942	0.942	0.033	0.032	0.025	0.026
Ak	0.848	0.740	0.021	0.154	0.131	0.105
Kidd	0.741	0.763	0.211	0.218	0.048	0.020
Duffy	0.636	0.636	0.105	0.105	0.259	0.259
Lewis	0.966	0.965	0.032	0.033	0.002	0.001
Kell	0.901	0.903	0.073	0.072	0.026	0.025
Lutheran	0.694	0.696	0.214	0.215	0.092	0.089
P	0.949	0.949	0.029	0.029	0.022	0.022
MNS	0.911	0.906	0.041	0.042	0.048	0.052
Rh	0.674	0.682	0.073	0.068	0.253	0.250
ABO	0.907	0.923	0.063	0.047	0.030	0.030
True mean	**0.857**		**0.071**		**0.072**	
Lewontin (1972) written mean	0.854		0.083		0.063	
Lewontin (1972) realculated mean (Table 4)	0.861		0.076		0.076	

[37] Lewontin's reported genetic diversity apportionment values, for each locus, from Table 4 in Lewontin (1972), as well as the true, correct apportionment of genetic diversity recalculated from Lewontin's Table 3, together with the true mean (for recalculations, see Winther 2022a). It remains unclear how Lewontin calculated his own written mean values, as they do not match a simple recalculation of his own Table 4. His overstatement of the among populations but within races diversity component at the expense of the among races diversity component is also somewhat curious. A note on nomenclature: while gene names and alleles today are typically italicized, as I have done in this book, also in order to distinguish them from their protein products (whose names are typically *not* italicized), Lewontin was not consistent – Lewontin (1974) italicizes genes and alleles (e.g. Table 33, p. 153) but Lewontin (1972) does not (e.g., Table 1, p. 384). Again, I italicize genes and alleles, except for Table 6.2 and Figures 6.4 and 6.5, which directly engage Lewontin (1972).

genome, showing some variation across continental regions. A key consequence of the hierarchical structure in genetic variation is that we can still use small allele differences globally to identify the population or cluster to which an individual belongs (cf. empirical pattern #5).

As we saw in pattern #3, we now know – thanks to the work of Noah Rosenberg, John Novembre, their respective colleagues, and many others – that most of the common alleles present in our species are globally distributed and that there are relatively few common, private alleles. Such a pattern was already presaged in Lewontin's meta-analysis, in which almost all the alleles of the 17 genes under study were found to be present in almost all populations.[38]

Interestingly, Lewontin's classic article is rather telegraphic in neither listing data sources nor explaining foundational evolutionary genetic theory, both of which I have now done in Winther (2022a), with data analysis available in Winther (2021b). I also found that Lewontin (1972) was replete with calculation errors for all of the 17 genes it covered, except one (P).[39] Importantly, Lewontin's calculation errors are not systematic (Figures 6.4 and 6.5; Table 6.2), although learning this required redoing all of his calculations and finding appropriate visualizations. However, the overstatement (by 0.7%) of the among populations diversity component and the understatement (by 1.3%) of the among races component, relative even just to Table 4 in Lewontin (1972), merits further discussion. Unfortunately, the erroneous distribution (even by his own data tables) of 85.4%/8.3%/6.3% is widely and influentially cited.[40] Now, Lewontin (1974) did correct some of these small calculation errors in its Table 34 (e.g., for the *Lewis* gene), for the same 17 genes as Lewontin (1972), and revised the overarching diversity apportionment to 84.9%/7.5%/7.5%, which is closer to his own data, even if still not quite consistent with it, and ends up rounding in the wrong direction. Even so, work focusing on the apportionment of human genetic variation that cites Lewontin (1972) rarely *also* cites Lewontin (1974).

There is reason for concern about Lewontin's own inappropriate reporting of his (true) results, and of the significantly greater influence of Lewontin (1972) compared to Lewontin (1974): When apportionments are commensurable among populations but within races, and

[38] See Table 3 in Lewontin (1972, pp. 390–394). [39] Winther (2022a).

[40] See, e.g., Latter (1980, p. 220), Barbujani et al. (1997, p. 4517), Brown and Armelagos (2001, p. 38), Long and Kittles (2003, p. 450), and Long et al. (2009, p. 23). None of these papers cites Lewontin (1974).

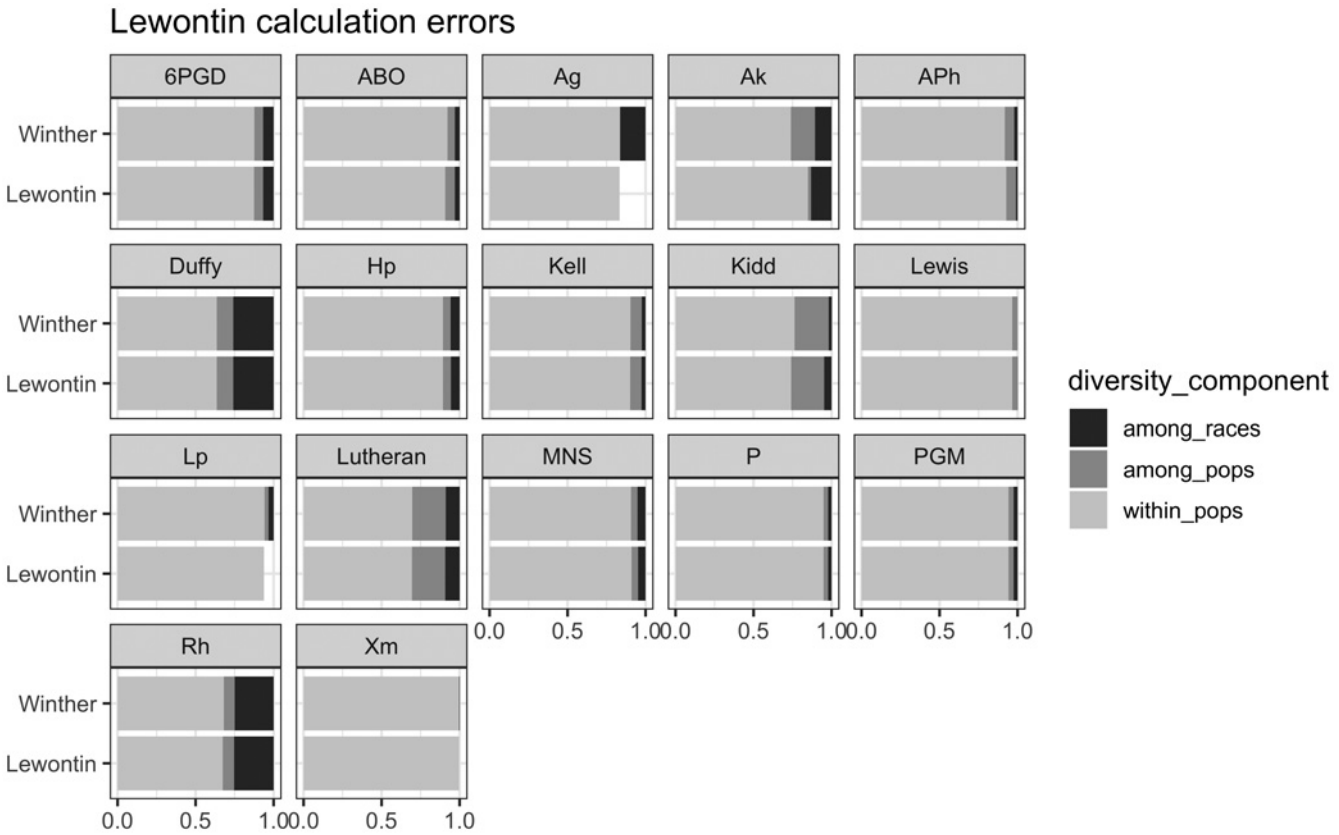

Figure 6.4. Lewontin–Winther bar plots
Bar plots of the three diversity components for each gene (on the same scale), as presented in Table 4 in Lewontin (1972), compared to the true recalculated values from Table 1.6 in Winther (2022a), juxtaposed here in Table 6.2. Lewontin did not calculate "among_pops" and "among_races" values for Ag, Lp, and Xm genes, hence there is some white empty space for those genes. Illustrated by Amir Najmi using ggplot2 in R. © 2021 Rasmus Grønfeldt Winther.

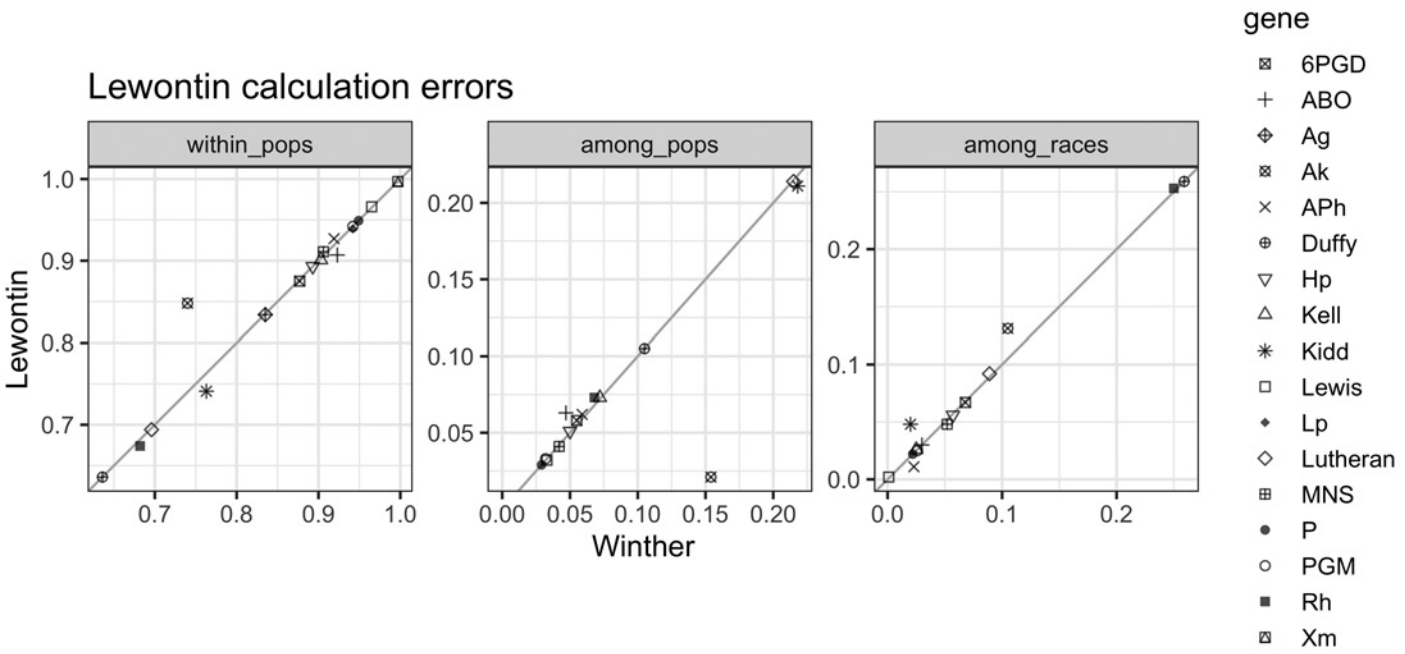

Figure 6.5. Lewontin–Winther scatter plots
Scatter plots of the differences between Lewontin's Table 4 values and the correct values from Table 1.6 of Winther (2022a). The diagonal line maps the identity of value between the two. Scales differ among the three plots because of the different magnitude ranges of the three diversity components (although the among populations and among races components are commensurable in magnitude). The "Winther" (*x*-axis) and "Lewontin" (*y*-axis) component values for each gene can be checked against Table 6.2. Note that while the "within_pops" scatter plot shows 17 genes, the "among_pops" and "among_races" scatter plots show only 14 genes because Lewontin had not calculated those respective values for Ag, Lp, and Xm genes. Illustrated by Amir Najmi using ggplot2 in R. © 2021 Rasmus Grønfeldt Winther
(A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

among races, Lewontin's harsh and critical stance toward the genetic reality of races becomes incongruous with his permissive and even laudatory stance toward the genetic reality of populations, at least in this context.[41]

How has Lewontin's distribution fared since Lewontin's original study? From a bird's-eye view, it has withstood the test of time in a very general way: All subsequent studies show that most genetic variation is within populations. But zooming in on the details of these studies complicates the picture. We saw in Chapter 5 that F_{ST} depends on a variety of assumptions, and even on gene type studied (e.g., microsatellites versus SNPs). Recall also Long's statement about the oddity of Lewontin – and many others – feeling that F_{ST} on the order of 15% was small.[42] What counts as large or small in this context would seem to be a matter of judgment, as I argue through the conventionalist philosophical position defended in Chapters 4 and 9. And three points bear mentioning regarding Lewontin's actual percentages:[43]

- *Within populations variance component estimates for autosomal genes range from approximately 80% to 95%*: Six key studies indicate the following estimates of the within populations component, showing the maximal ranges, with different values resulting from differences in autosomal gene type, measure, and variance partitioning model used: 80.2–90.8%; 82.7–90.3%; 84.4%; 80.9–87.9%; 89.8–94.6%; and 88.9–94.0%.[44] This is a rather significant range of estimates, even if the average is only slightly higher than Lewontin's 86% within populations component estimate.

[41] Gannett (2022) and Winther (2022a) trace Dobzhansky's influence on his doctoral student, Lewontin, in emphasizing the ontological importance and uniqueness of populations.

[42] Long (2009, p. 801).

[43] For the purposes of simplicity, I here abstract away from indicating – for every case – the precise gene types, measures, and variance partitioning models used. The indicated ranges, or single values, are for the point estimates presented; I have not accounted for the standard errors that some publications give for the point estimates. Table 1 in Brown and Armelagos (2001, p. 38) provides a summary of nine different studies, including Lewontin (1972). Table 5 in Jorde et al. (2000, p. 983) summarizes the variance partitioning results for their article, which was "the first published comparison of within- and between-population genetic diversity in autosomal, mtDNA, and Y-chromosome loci in the same set of individuals" (p. 979). Table 4 in Rosenberg (2011, p. 670) presents variance components (and 95% confidence intervals) for each continental region singly, and for the entire globe, itself classified in different ways (i.e., 1, 5, or 7 regions). The partial lists could be much longer.

[44] Respectively, Table 4 in Latter (1980, p. 228), Table 1 in Ryman et al. (1983, p. 97), Tables 1 and 2 in Barbujani et al. (1997, pp. 4517, 4518), Table 5 in Jorde et al. (2000, p. 983), Table 1 in Rosenberg et al. (2002, p. 2382), and Li et al. (2008, p. 1102).

- *Among continents variance component estimates for autosomal genes are sometimes significant, easily exceeding 10% in a number of studies*. Six key studies represent the following among continents component estimates (and, for comparison in parentheses, the among populations but within continents component estimates): 2.8–14.0%; 7.2–15.4% (2.0–5.3%); 10.0–11.7% (3.9–5.5%); 10.4–17.4% (1.3–1.8%); 3.6–5.2% (2.4–5.0%); 3.7–9.0% (2.1–2.3%).[45] Again, these are maximal ranges for each study, with distinct estimates stemming from employing different (autosomal) gene types, measures, and models.

  A number of these studies present a high among continents component: up to 17.4%. Moreover, the sum of variance components above the population level is itself high, up to 19.8% (calculate this as 1 minus the within populations component). Averaging the range value extremes across all the studies, which itself is a strategy requiring discussion and qualification, produces a grand average of 9.2% for the among continents variance component, and of 3.2% for the among populations but within continents variance component. This is a somewhat higher among continents variance component than Lewontin's distribution, even if the sum of the two components is close to Lewontin's distribution 14% sum. Thus, averaging the studies seems to ground the relative stability of Lewontin's distribution, at least for autosomal genes. But the ranges merit ~~further~~ discussion and the consistently higher among continents relative to among populations but within continents estimates belies the typical, easy, interpretation of Lewontin's distribution, calling into question its absolute nature, perhaps even its reproducibility. We might be better off learning the full range of estimates for each of the three variance components, and discussing each of their potential meanings (see Chapter 9).[46]
- *Mitochondrial DNA and Y chromosome among continents variance component estimates are as high as 25%.* The among continents variance component for both mitochondrial DNA (mtDNA) and Y chromosomes can be

[45] Identical sources to the previous note. Because Latter (1980) articulates a four-level hierarchy dividing continents into "regional subgroups" (e.g., Europe into "Northern, Central, Southern" and Oceania into "Australia, Melanesia, and Micronesia," p. 224), I here indicate only within populations and among continents estimates for his study. With the globe as one region, Rosenberg et al. (2002) estimated 94.6% within populations and 5.4% among populations.

[46] Furthermore, F_{ST} varies dramatically for different population pairs. Figure 5C in Bergström et al. (2020, p. 6) gives an immediate visual sense of the enormous variability in F_{ST} for all pairs possible of 54 HGDP-CEPH populations. Values can be as low as effectively 0 for some non-African population pairs and as high as roughly 0.28 for some African–non-African population pairs.

remarkably high (see Chapter 2). For mtDNA, key studies found maximal ranges of among continents component estimates of 15.73–21.99%; 12.5%; 22.0–24.9%; and 14.3%.[47] For the among races component of Y chromosomes, researchers estimated 52.7%, 7.8%, and 21.3%.[48] Wilder et al. argue that Seielstad et al.'s (1998) data sets "varied considerably with regard to sample sizes, populations represented and method used to assay genetic variation"; they found instead that "genetic differentiation between populations was similar for the Y chromosome and mtDNA at all geographic scales that we tested."[49] In a nutshell, the among continents variance component estimates for mtDNA and Y chromosomes are significantly higher than for autosomal genes. Furthermore, some studies find higher among continents, and among populations but within continents, genetic variance components for Y chromosomes than for mtDNA, while others find them to be somewhat comparable.[50]

Two patterns require some explication here: first, the significantly larger among continents component of mtDNA and Y chromosome genetic variance as opposed to nuclear, autosomal genes; and second, the differences (such as they are) between mtDNA and Y chromosome among continents, and among populations but within continents, variance components. Regarding the first, due to their haploidy and their uniparental inheritance, the customary, effective population size for mtDNA and Y chromosome loci is one-quarter that for autosomal loci, making genetic drift and consequential population differentiation much more significant, all other factors being equal.[51] Regarding the second, roughly 70% of human cultures and populations are

[47] Respectively, Table 3 in Excoffier et al. (1992, p. 486), Table 1 in Seielstad et al. (1998, p. 278), Table 5 in Jorde et al. (2000, p. 983), and Figure 2 in Lippold et al. (2014, p. 7) (this last estimate was not explicitly given in the article and had to be measured from the figure). Again, I abstract away from gene type, measure, and exact methodology.

[48] Respectively, Table 1 in Seielstad et al. (1998, p. 278), Table 5 in Jorde et al. (2000, p. 983), and Figure 2 in Lippold et al. (2014, p. 7) (this last estimate was not explicitly given in the article and had to be measured from the figure). Again, I abstract away from gene type, measure, and methodology.

[49] Wilder et al. (2004, p. 1122).

[50] See, e.g., Gunnarsdóttir et al. (2011). Wilkins and Marlowe (2006) discuss how any such data should be interpreted, why there might be conflicting results for the relative size of the among continents variance components for the Y chromosome and mtDNA, and "how the demographic shift associated with agriculture might affect genetic diversity over different spatial scales" (p. 290; see Chapter 2).

[51] See Hartl and Clark (1989, pp. 425–426), Jorde et al. (2000), and Storz et al. (2001). On effective population size, see Chapter 3.

patrilocal, whereby men tend to stay in their birthplace and women migrate to marry and form new families, at least at a fairly localized scale.[52] A standard anthropological genetics explanation is that "matrilocal groups have high within-group diversity for the Y chromosome and large between-group distances for mtDNA [because diverse men migrate from afar to relatively female-homogeneous villages], whereas patrilocal groups have high within-group diversity for mtDNA and large between-group distances for the Y chromosome [because diverse women migrate from afar to relatively male-homogeneous villages]."[53] Since more human cultures are patrilocal, estimates of the among continents and among populations variance components of Y chromosomes should be larger than the estimates of the same two variance components for mtDNA.[54] However, this might be especially true at more local scales, say, within subregions of continents, whereas at global scales there may be male-biased dispersal and migration patterns, leading to less difference in the Y chromosome and mtDNA among continents variance components.[55]

Lewontin (1972, 1974) did not include mtDNA or Y chromosome data. It is furthermore unclear how we should approach and weight these data in our overall understanding of human population structure. Data on haploid units uniparentally inherited edge us closer toward Galápagos-Writ-Large (see Figure 4.1).

5. Despite Lewontin's Distribution, Clustering Populations And Classifying Individuals Is Possible

Even if most variation is within populations, and Lewontin's distribution roughly and broadly holds, if we accumulate information across loci rather than averaging across loci, we can make somewhat reliable inferences about contemporary populations that exist (clustering analysis) and about the population membership(s) of any particular individual (classification).[56] *Structure*, the clustering analysis computer program

[52] Arias et al. (2018, p. 2719). [53] Oota et al. (2001, p. 21). [54] Seielstad et al. (1998).

[55] See Hammer et al. (2001), Wilkins and Marlowe (2006), and Marks et al. (2012). Indeed, other explanatory factors are necessary for a fuller story. I have here side-stepped selection (but see Sayres, 2018, and Chapter 7). Moreover, Heyer et al. (2012) show how differences in effective population size between females and males is also important here.

[56] See Rosenberg et al. (2002), Edwards (2003), as reprinted in Winther (2018a), Witherspoon et al. (2007), Kaplan (2011), Tal (2012), Edge and Rosenberg (2015), Rosenberg (2018), and Winther (2018b). This is what, somewhat unfortunately, has been termed "Lewontin's fallacy" or, less frequently, "Lewontin's paradox." Even if most genetic variation is within populations, clustering