GENETIC VARIATION AND HEALTH IN A RURAL CARIBBEAN VILLAGE

A dissertation presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

---

by

MONICA H. KEITH

JULY 2020

The undersigned, appointed by the Dean of the Graduate School,

have examined the dissertation entitled:

GENETIC VARIATION AND HEALTH

IN A RURAL CARIBBEAN VILLAGE

Presented by Monica H. Keith, a candidate for the degree of Doctor of Philosophy,

and hereby certify that it is worthy of acceptance.

Gregory E. Blomquist	Mark V. Flinn
Lisa Sattenspiel	Jared E. Decker

To my husband, Shayne

Thank you for your support, patience, and partnership.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the community of Bwa Mawego for facilitating this research and recognize each of the participants for their contributions. I am especially grateful to local research assistants Nethilia Durand and Nakisha Wiltshire for helping with data collection.

This work would not have been possible without the mentorship of my adviser, Greg Blomquist, who has provided statistical insight, programming expertise, and invaluable guidance. Mark Flinn was instrumental to the planning and data collection for this research, and I thank him for connecting me with the community to address health concerns of local importance in Bwa Mawego. I am grateful to all members of my dissertation committee for their support, and I thank Lisa Sattenspiel for sharing her knowledge of fieldwork in Dominica. Jared Decker's guidance and expertise were essential to the genotypic analyses of admixture and population structure.

A network of researchers, mentors, and collaborators facilitated the data collection and analyses for this project. I thank Tom Jaskowiec for his logistic help and field advice in Dominica. Kristen Taylor provided lab space, equipment, and resources for DNA extractions, and I thank Alex Stuckel for his help with lab setup and sample processing advice. I am grateful for Jerry Taylor's feedback and expertise in the research design and genome-wide association analyses. Many thanks to Harly Durbin and Troy Rowan for their help designing and executing analyses of admixture and population structure. Shane Macfarlan provided the 11-generation pedigree for Bwa Mawego. I am grateful for Lori Eggert's mentorship and lab training, especially in DNA extractions. I would also like to thank the Life Sciences Fellowship for funding my first four years at

MU and the Dorothy (Dot) Tompkins Gelvin Memorial Scholarship for funding my travel and field expenses for data collection in 2017.

Finally, I thank my husband, friends, family, and cats for their encouragement, patience, and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABSTRACT

Bwa Mawego is a small-scale horticultural community (~500 people) on the island of Dominica that has been the site of a longitudinal health research project for more than 30 years. Cardiovascular diseases and metabolic health are growing local concerns. Here we analyze longitudinal growth data, cardiometabolic metrics, and genome-wide single nucleotide polymorphism (SNP) data from this population to investigate sources of variation in anthropometric and cardiometabolic outcomes. Mixed-effect heritability models indicate that (1) variation in body mass index (BMI) is significantly shaped by genetic variation, and (2) variation between longitudinal BMI curves has not been consistently impacted by secular environmental trends from 1997-2017. In order to assess genetic variation in more detail, we first characterize the population structure and admixture in this Caribbean community using high-density SNP data and global reference samples in the Human Genome Diversity Panel. We detect four distinct family clusters and admixture from African, European, and Amerindian ancestral populations that occurred 5-6 generations ago (~130-150 years). Amerindian haplotypes represented in Bwa Mawego associate with deeply diverged lineages in Karitiana and Surui peoples, highlighting the regionally variable nature of admixture throughout the Caribbean and unique historical outcomes in Dominica. Genome-wide association tests of cardiometabolic phenotypes identify a genomic region of interest downstream of the *ANK3* gene that associates with BMI in Bwa Mawego, after controlling for confounding variation from ancestral population structure and relatedness. Any functional relationship between *ANK3* and BMI is currently uncharacterized, and there is unique potential to further explore complex gene-environment-phenotype landscapes in Bwa Mawego.

CHAPTER ONE

Introduction

The detail with which genetic data capture patterns of human variation is unparalleled, and our genomes simultaneously reflect demographic histories and impact current biology and health. The physical mapping of the human genome in the early 2000's (Lander et al., 2001) expanded our ability to characterize sequence variation (Sachidanandam et al., 2001), protein-coding functionality (MacArthur et al., 2012), non-coding regulatory systems (Zhang and Lupski, 2015), and a myriad of complex interactions between genetic and environmental sources of variation (Portela and Esteller, 2010; Marsit, 2015). Interpretations of genetic, health, and biocultural variation are limited by the individuals represented in research datasets, and more inclusive work is needed to reach underserved populations and represent global human diversity in health-oriented research (Bustamante et al., 2011).

Human Genomic Diversity

Several large-scale projects have aimed to capture global population genetic diversity, including the International HapMap Project (International HapMap Consortium, 2003), the Human Genome Diversity Panel (HGDP) (Cann et al., 2002), and the 1,000 Genomes Project (Delaneau and Marchini, 2014). Analyses of global datasets detect strong associations between genomic similarity and geographic distance (Relethford, 2004), which reflects what is referenced as genetic "population structure". Globally, population structure is in part shaped by continental geography (Rosenberg et al., 2002), and demographic processes (e.g. fertility, migration, mortality) have shaped

global genomic variation through time and space via generational recombination of inherited genetic sequences (haplotypes) and migrations throughout human evolutionary history.

Global genomic patterns show concentrated haplotype heterozygosity (i.e. high amounts of genomic variation) in sub-Saharan Africa where our species first evolved; genetic variation decreases proportionally with distance from eastern Africa, showing evidence of successive migrations of small groups who dispersed out from this region over time (Li et al., 2008). Different migrations saw different (and reduced) representations of parental genetic variation out of Africa, and factors such as population size, geographic and/or cultural boundaries, and natural selection have subsequently shaped population structures and sub-structures. Structure of this nature varies regionally and is related to patterns of linkage disequilibrium (Campbell and Tishkoff, 2008).

Linkage disequilibrium (LD) refers to nonrandom association between alleles at different loci, largely impacted by the physical distance between genetic markers on a chromosome. Alleles that are located close to one another are less likely to be separated by recombination from one generation to the next as inherited chromosome pieces (haplotypes) are shuffled during meiosis, but linkage erodes over time as haplotypes recombine and break into smaller segments each generation. Assuming negligible inbreeding, increasingly diverse haplotypes emerge in a population with each generational recombination event. Thus, African populations with the most haplotype diversity also have the lowest levels of LD, and both of these attributes result from our species' evolutionary origin in sub-Saharan Africa where human genomes have been recombining for the largest number of generations (Campbell and Tishkoff, 2008).

Genotype-Phenotype Associations

In a biological context, LD is a critical factor that determines our ability to detect associations between genetic variants and observed traits (Zondervan and Cardon, 2004). "Complex" traits and diseases that result from interactions between many genetic variants and environmental factors present structural and statistical challenges. Many genetic variants of small to moderate phenotypic effect contribute, and markers are scanned genome-wide in an untargeted manner to identify genetic regions of interest. Efficient, cost-effective genotype datasets can sequence up to 1-2 million variants per array (panel of markers sequenced for each individual), which may or may not actually include causal loci for observed traits (LaFramboise, 2009). Thus, genome-wide association studies (GWAS) rely on sufficient levels of LD between tagged markers and actual causal variants to detect informative genotype-phenotype associations. Several factors impact our ability to detect associations in this manner, including the allele frequencies of associated variants, phenotypic effect sizes of causal variants, sample size, confounding sources of genetic and/or phenotypic variation, and sufficient array coverage of variation across the genome. GWAS can detect associations between markers in high LD with common alleles of small phenotypic effect (given sufficient sample size), or with rare- but not very rare (freq < 0.01)- alleles of moderate phenotypic effect (Zondervan and Cardon, 2004).

GWAS have had unprecedented successes identifying functionally verified associations for several phenotypes, including type II diabetes (Zeggini et al., 2007), body mass index (BMI) (Frayling et al., 2007), asthma (Moffatt et al., 2007), and some cancers (Gudmundsson et al., 2007; Stacey et al., 2007). With higher mapping resolution

and broader potential than other methods such as linkage mapping or candidate gene studies, GWAS remain the primary tool for identifying associations between genetic variants and phenotypes (Wu et al., 2017). Despite these successes, GWAS have only been able to explain small proportions of estimated heritabilities for complex traits and diseases, and the majority of genetic variance that underlies phenotypic variation remains uncharacterized (Manolio et al., 2009). This is also known as "missing heritability".

More than 90% of GWAS have only included individuals of recent European ancestry (Need and Goldstein, 2009). There is vast potential to both reduce missing heritability and improve mapping resolution to reveal causal variants by conducting GWAS with more diverse populations (McCarthy et al., 2008). As robustly demonstrated by variants of the *FTO* gene, meaningful genotype-phenotype associations may extend across populations but only be detectable in specific regions due to population structure, low allele frequencies, and other sample-specific factors. Multiple studies have confirmed the causal effects of *FTO* variants on fat storage and body weight regulation across people of diverse ancestries, but targeted investigation of causal alleles has also shown that their frequencies are much lower in Asian and African populations than among European cohorts, rendering this biologically informative association less likely to be detected in non-European samples (Loos and Yeo, 2014). Nevertheless, the vast majority of GWAS contain *only* European genotypes, thus our current knowledge of genotype-phenotype relationships is biased and largely limited to those variants that meet detectable criteria in European populations (Bustamante et al., 2011).

Most GWAS are limited to include individuals of a single continental ancestry, and many have more specific inclusion criteria than that (Rosenberg et al., 2010). Such

parameters aim to reduce heterogeneity in a study's sample from population structure and variable patterns of LD, which are sources of genetic variation that confound the detection of variant-phenotype associations. Case-control ratios and quantitative phenotype distributions may also vary across race and ethnicity for myriad reasons, creating another potential confounding pattern that limits sample diversity in many studies.

Several factors that have contributed to European sample bias in GWAS can now be easily mediated in order to improve inclusivity across global populations. Higher LD, reduced haplotype diversity, and relatively homogenous genetic population structure across most people of European descent were beneficial when SNP arrays were more limited in size, because fewer markers are needed to capture variation across genomes with less haplotype diversity. Technological advances and commercialization have lowered costs and enabled researchers to genotype study participants at higher densities across their genomes with SNP arrays commonly ranging from 500,000 to more than 2.5 million variants now, improving the detection of associations on shorter haplotypes (Ha et al., 2014). Additionally, dimension-reduction techniques efficiently allow researchers to control for some confounding sources of genetic variation, and principal component analyses enable us to account for heterogeneous population structure with the inclusion of principal component loadings in GWA models as covariates (Price et al., 2006).

Large-scale GWAS typically exclude related individuals because genetic variants that are identical by descent (IBD) between relatives are another confounding influence on patterns of genomic variation. However, in a community-based sample of related individuals, we can include pairwise measures of relatedness (e.g. kinship coefficients) as

a GWA covariate to account for this population attribute (Laird and Lange, 2006). Furthermore, family-based designs have some benefit in the GWA framework as genotype-phenotype associations that are enriched in family lineages will be more detectable at higher frequencies (Li et al., 2006). GWAS among related participants in unique, localized populations have identified genetic variants of physiological importance (e.g. for glucose metabolism), demonstrating the potential of these methods in diverse communities and ecologies (Li et al., 2007). Reduced environmental heterogeneity between individuals from localized study populations may also enhance detection of meaningful genotype-phenotype associations compared with association tests in more clinical research designs that often include participants with broad lifestyle variation among them (Kulminski et al., 2016). GWAS in localized populations also capture associations within specific ecological contexts, adding another potential set of environmental interactions to explore (McCarthy et al., 2008).

Ancestral variation within a GWA sample can be confounding if it is not accounted for (e.g. with principal component analysis loadings), but recently mixed ancestry (admixture) can actually enhance our ability to detect genotype-phenotype associations in a population with relatively long haplotypes produced by recent ancestral recombination and diverse representation of continental ancestries (Medina-Gomez et al., 2015). Admixture mapping takes advantage of this long-range haplotype structure as well as structure in phenotypic outcomes that vary by ancestry to detect genotype-phenotype associations (Winkler et al., 2010), thus demonstrating that admixed populations have unique potential to expand our understandings of human variation. Large proportions of the world's population are admixed (e.g. African Americans, communities throughout

Latin America and the Caribbean), and more inclusive research is needed to mitigate existing European bias in health and genetic research.

Study Population

Bwa Mawego is a small-scale horticultural community in the Commonwealth of Dominica, a small island in the Lesser Antilles with a unique colonial history. This community is one of the most isolated and remote on the island, located on the steep windward coast near the largest indigenous reserve (Kalinago Territory) in the Caribbean (Flinn et al., 1999; Quinlan, 2004). The majority of Bwa Mawego's residents (~500 people) have been engaging in anthropological and psychological research for more than 30 years, producing more than 20 years of longitudinal health data and a population-wide, 11-generation genealogy amidst a multitude of health and behavioral research (Flinn, 1999; Flinn, 2009; Flinn and England, 1997; Flinn et al., 1999; Flinn et al., 2012; Quinlan, 2004; Quinlan and Flinn, 2005; Quinlan and Hagen, 2008; Macfarlan et al., 2012; Ponzi et al., 2015).

The ecology of Bwa Mawego has been characterized in detail, and the community has an environmentally-shaped history beginning in the mid-19th century following the emancipation of enslaved peoples (mostly of African descent) and indentured servants (mostly of European descent) on the island who sought protection and refuge by settling in the island's steepest, heavily forested terrain (Quinlan, 2004). Bwa Mawego's specific ecology is exceptionally steep and difficult to traverse, even by Dominica standards. Flexible matrifocal structure shapes the social environment in Bwa Mawego such that households are relatively fluid in their compositions over time, and resources (including products of horticultural labor) are commonly shared beyond immediate kin (Quinlan,

2006). Longitudinal ethnographic data, health reports, and hormone profiles (cortisol, testosterone, etc.) have characterized specific aspects of family-based social environments in Bwa Mawego and established associations between child growth, stress, and household composition (e.g. living with step-parents) (Flinn et al., 1999; Flinn et al., 2012). Previous research highlights the unique potential to investigate biocultural interactions and trends in Bwa Mawego, especially across generations.

Cardiometabolic traits, including type II diabetes and hypertension, have become top local health priorities in Bwa Mawego over the past several decades. Technological developments throughout the community such as piped water, electricity, and internet access have altered lifestyles to an extent beginning in the late 1990's (Quinlan and Flinn, 2005), yet most residents continue to practice traditional subsistence horticulture full-time and are curious as to why cardiometabolic conditions appear to cluster as they do in the community despite a lack of obvious environmental or behavioral explanation.

This dissertation examines anthropometric, cardiometabolic, and genetic variation in Bwa Mawego, Dominica in an integrated manner. First, a pedigree-based analysis of longitudinal growth data estimates the extent to which variation in height, weight, and BMI is shaped by additive genetic variation and environmental secular trends. This content (Chapter Two) was published in the *American Journal of Physical Anthropology* in 2019 (Keith et al., 2019). Second, population genetic analyses of high-density SNP data provide the first genome-wide characterization of admixture and population structure in this unique Caribbean region. Lastly, GWAS of cardiometabolic and SNP data test for genotype-phenotype associations while controlling for ancestral population structure and family relatedness. Together, these analyses address health traits of local importance in

Bwa Mawego to enhance our understandings of human variation both in the community's unique ecology and in a broader genetic context that increases diverse representation.

References

Bustamante, C. D., De La Vega, F. M., & Burchard, E. G. (2011). Genomics for the world. *Nature*, 475(7355), 163–165. https://doi.org/10.1038/475163a

Campbell, M.C., & Tishkoff, S.A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics,* 9, 403–433. https://doi.org/10.1146/annurev.genom.9.081307.164258

Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., … Cavalli-Sforza, L.L. (2002). A human genome diversity cell line panel. *Science*, 296(5566), 261–262. https://doi.org/10.1126/science.296.5566.261b

Delaneau, O., & Marchini, J. (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nature Communications*, 5(1), 3934. https://doi.org/10.1038/ncomms4934

Flinn, M.V. (2009). Are cortisol profiles a stable trait during child development? *American Journal of Human Biology* 21(6), 769-771.

Flinn, M.V., Duncan, C.M., Ponzi, D., Quinlan, R.J., Decker, S.A., & Leone, D.V. (2012). Hormones in the Wild: Monitoring the Endocrinology of Family Relationships. *Parenting*, 12(2–3), 124–133.

Flinn M.V. & England, B.G. (1997). Social economics of childhood glucocorticoid stress response and health. *American Journal of Physical Anthropology* 102(1): 33-53.

Flinn, M.V. (1999). Family environment, stress, and health during childhood. In: Hormones, health, and behavior, C. Panter-Brick & C. Worthman (Eds.), pp. 105-138. Cambridge: Cambridge University press.

Flinn, M.V., Leone, D.V., & Quinlan, R.J. (1999). Growth and Fluctuating Asymmetry of Stepchildren. *Evolution and Human Behavior*, 20(6), 465-479.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., … McCarthy, M.I. (2007). A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science*, 316(5826), 889–894. https://doi.org/10.1126/science.1141634

Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J.T., Thorleifsson, G., Manolescu, A., … Stefansson, K. (2007). Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nature Genetics*, 39(8), 977–983. https://doi.org/10.1038/ng2062

Ha, N.T., Freytag, S., & Bickeboeller, H. (2014). Coverage and efficiency in current SNP chips. *European Journal of Human Genetics*, 22(9), 1124–1130. https://doi.org/10.1038/ejhg.2013.304

International HapMap Consortium. (2003). The International HapMap Project. *Nature*, 426(6968), 789–796. https://doi.org/10.1038/nature02168

Keith, M.H., Blomquist, G.E., & Flinn, M.V. (2019). Anthropometric heritability and child growth in a Caribbean village: A quantitative genetic analysis of longitudinal height, weight, and body mass index in Bwa Mawego, Dominica. *American Journal of Physical Anthropology*, 170(3), 393–403. https://doi.org/10.1002/ajpa.23924

Kulminski, A.M., Loika, Y., Culminskaya, I., Arbeev, K.G., Ukraintseva, S.V., Stallard, E., & Yashin, A.I. (2016). Explicating heterogeneity of complex traits has strong potential for improving GWAS efficiency. *Scientific Reports*, 6(1), 35390. https://doi.org/10.1038/srep35390

LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research*, 37(13), 4181–4193. https://doi.org/10.1093/nar/gkp552

Laird, N.M., & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, 7(5), 385–394. https://doi.org/10.1038/nrg1839

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., … International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. https://doi.org/10.1038/35057062

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., … Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319(5866), 1100–1104. https://doi.org/10.1126/science.1153717

Li, M., Boehnke, M., & Abecasis, G.R. (2006). Efficient Study Designs for Test of Genetic Association Using Sibship Data and Unrelated Cases and Controls. The *American Journal of Human Genetics*, 78(5), 778–792. https://doi.org/10.1086/503711

Li, S., Sanna, S., Maschio, A., Busonero, F., Usala, G., Mulas, A., … Nagaraja, R. (2007). The GLUT9 Gene Is Associated with Serum Uric Acid Levels in Sardinia and Chianti Cohorts. *PLOS Genetics*, 3(11), e194. https://doi.org/10.1371/journal.pgen.0030194

Loos, R.J.F., & Yeo, G.S.H. (2014). The bigger picture of FTO – the first GWAS-identified obesity gene. Nature Reviews. *Endocrinology*, 10(1), 51–61. https://doi.org/10.1038/nrendo.2013.227

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., … Tyler-Smith, C. (2012). A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science*, 335(6070), 823–828. https://doi.org/10.1126/science.1215040

Macfarlan, S.J., Remiker, M., & Quinlan, R. (2012). Competitive Altruism Explains Labor Exchange Variation in a Dominican Community. *Current Anthropology*, 53(1), 118–124. https://doi.org/10.1086/663700

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., … Visscher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753. https://doi.org/10.1038/nature08494

Marsit, C.J. (2015). Influence of environmental exposure on human epigenetic regulation. *Journal of Experimental Biology*, 218(1), 71–79. https://doi.org/10.1242/jeb.106971

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. A., & Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, 9(5), 356–369. https://doi.org/10.1038/nrg2344

Medina-Gomez, C., Felix, J.F., Estrada, K., Peters, M.J., Herrera, L., Kruithof, C.J., … Rivadeneira, F. (2015). Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *European Journal of Epidemiology*, 30(4), 317–330. https://doi.org/10.1007/s10654-015-9998-4

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., … Cookson, W.O.C. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, 448(7152), 470–473. https://doi.org/10.1038/nature06014

Need, A.C., & Goldstein, D.B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics*, 25(11), 489–494. https://doi.org/10.1016/j.tig.2009.09.012

Ponzi, D., Muehlenbein, M.P., Geary, D.C., & Flinn, M.V. (2016). Cortisol, salivary alpha-amylase and children's perceptions of their social networks. *Social Neuroscience*, 11(2), 164-174. DOI: 10.1080/17470919.2015.1045988

Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, 28(10), 1057–1068. https://doi.org/10.1038/nbt.1685

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909. https://doi.org/10.1038/ng1847

Quinlan, M.B. (2004). From the Bush: The Front Line of Health Care in a Caribbean Village. Belmont, CA: Wadsworth.

Quinlan, R.J. (2006). Gender and Risk in a Matrifocal Caribbean Community: A View from Behavioral Ecology. *American Anthropologist*, 108(3), 464–479.

Quinlan, R.J., & Hagen, E.H. (2008). New Genealogy: It's Not Just for Kinship Anymore. *Field Methods*.

Quinlan, R.J., & Flinn, M. V. (2003). Intergenerational Transmission of Conjugal Stability in a Caribbean Community. *Journal of Comparative Family Studies*, 34(4), 569–583.

Quinlan, R.J., & Flinn, M.V. (2005). Kinship, sex, and fitness in a Caribbean community. *Human Nature*, 16(1), 32–57. https://doi.org/10.1007/s12110-005-1006-3

Relethford, J.H. (2004). Global Patterns of Isolation by Distance Based on Genetic and Morphological Data. *Human Biology*, 76(4), 499–513.

Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I., & Boehnke, M. (2010). Genome-wide association studies in diverse populations. *Nature Reviews*. Genetics, 11(5), 356–366. https://doi.org/10.1038/nrg2760

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., …Kidd, K.K. (2002). Genetic structure of human populations. *Science.* 298:2381–2385

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., … Altshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822), 928–934.

Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., … Stefansson, K. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor–positive breast cancer. *Nature Genetics*, 39(7), 865–869. https://doi.org/10.1038/ng2064

Winkler, C.A., Nelson, G.W., & Smith, M.W. (2010). Admixture Mapping Comes of Age. *Annual Review of Genomics and Human Genetics*, 11(1), 65–89. https://doi.org/10.1146/annurev-genom-082509-141523

Wu, Y., Zheng, Z., Visscher, P.M., & Yang, J. (2017). Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biology*, 18(1), 86. https://doi.org/10.1186/s13059-017-1216-0

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., … Hattersley, A. T. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316(5829), 1336–1341. https://doi.org/10.1126/science.1142364

Zhang, F., & Lupski, J.R. (2015). Non-coding genetic variants in human disease. *Human Molecular Genetics*, 24(R1), R102–R110. https://doi.org/10.1093/hmg/ddv259

Zondervan, K.T., & Cardon, L.R. (2004). The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*, 5(2), 89–100. https://doi.org/10.1038/nrg1270

CHAPTER TWO

Anthropometric Heritability and Child Growth in Rural Dominica

Abstract

Body size and composition vary widely among individuals and populations, and long-term research in diverse contexts informs our understanding of genetic, cultural, and environmental impacts on this variation. We analyze longitudinal measures of height, weight, and BMI from a Caribbean village, estimating the extent to which variation in these anthropometrics is shaped by genetic variance in a small-scale population of mixed ancestry. An 11-generation pedigree enables us to estimate the proportions of phenotypic variation in height, weight, and BMI attributed to genetic variation. We assess variation within individual growth curves as well as heritabilities of these traits for 260 individuals using Bayesian variance component estimation. Analyses of longitudinal anthropometrics show high repeatabilities (>0.75) within individual growth curves independent of age or sex. Moderate heritabilities (h^2_{height}=0.68, h^2_{weight}=0.64, h^2_{BMI}=0.49) reveal clear genetic signals, accounting for large proportions of the variation in body size observed between families. Secular trends indicate that height has increased by approximately 2 inches and weight by 13 pounds (averaged across age and sex) between 1997 and 2017. Body mass varies widely between individuals in this population without a clear secular trend and is significantly shaped by genetic variation, warranting further exploration with other physiological correlates and associated genetic variants.

Introduction

Variation in body size and child growth is shaped by combinations of biological, cultural, and environmental factors that are often ecologically dependent and population-specific. Body size and growth patterns vary widely among populations in developing countries and small-scale societies (Walker et al., 2006), many of whom are transitioning nutritionally and behaviorally to more Westernized, processed foods and decreased physical activity (Popkin et al., 2012). Cardiometabolic health is a growing concern in the Caribbean, yet we know little of the specific biological, cultural, and environmental influences on anthropometric variation in this region, particularly in rural areas (Boyne, 2009; Rueda-Clausen et al., 2008). Longitudinal measures of height, weight, and body mass index (BMI) reflect secular trends in human growth and development (Cole, 2003) and inform public health concerns of undernutrition and overnutrition (Monteiro et al., 2003). We assess the repeatabilities and heritabilities of longitudinal height, weight, and BMI in a rural Caribbean village during a period of nutritional transition to quantify variation within individual growth trajectories and to estimate the proportion of variation between individuals attributed to genetic versus non-genetic variance.

Body size and composition vary among geographic regions (de Onis et al., 2004), in part reflecting climate adaptations such that temperature and BMI generally show inverse relationships in indigenous populations (Wells, 2012). Small-scale societies that share similar tropical climates show considerable variation among their growth patterns, shaped partially by life-history trade-offs that balance growth rates with mortality risk and fertility to produce taller individuals through faster growth (Walker et al., 2006). The extent to which height, weight, and BMI are phenotypically plastic depends upon genetic

variants, environmental inputs, and developmental/epigenetic backgrounds (Godfrey et al., 2010). The relative impacts of these factors vary within and between populations as well as across age over the course of the human lifespan (Visscher et al., 2008).

Heritabilities quantify the proportions of variation in observed phenotypes that are explained by genetic variation in a population (Falconer, 1960, Lynch & Walsh, 1998). Heritability estimates do not reflect the extent to which a trait's phenotypic outcome is determined by an individual's genes, but instead partition the variance in an observed trait into genetic and non-genetic components (Lewontin, 1974; Vitzthum, 2003). Methods for estimating heritabilities use known genetic relationships to assess the extent to which the proportion of alleles shared among individuals associates with phenotypic variation (Vandemark, Kelly, & Eckhardt, 1985), and larger pedigrees with many generations provide varied kinship coefficients that produce more robust estimates of genetic variance than those based on only ancestor-descendant pairs, sib-ships, etc. (Kruuk, 2004; Wilson et al., 2010).

Quantities of phenotypic and genotypic variation are population-specific, and heritability estimates range from 0.26-0.90 for height, 0.22-0.85 for weight, and 0.17-0.90 for BMI (Dubois et al., 2012; Elks et al., 2012; Nan et al., 2012; Starkweather and Keith, 2018; Yang et al., 2015). Few heritabilities are published from Caribbean populations, but an analysis from Jamaica reports estimates of 0.74 for height, 0.63 for weight, and 0.53 for BMI (Luke et al., 2001). Larger heritabilities can result from larger quantities of genetic variation or from relatively low amounts of environmental variance. The proportional impact of genetic variance generally increases under stable and more favorable environmental conditions (Hoffmann et al., 1999), however heritabilities may

also decrease when environmental conditions change drastically between generations. Lower anthropometric heritabilities in some immigrant populations reflect the context-dependent nature of genetic variance components when environmental conditions differ dramatically between ancestors and their descendants (Gravlee et al., 2003). Across the lifespan, Elks et al. (2012) found higher heritabilities for BMI in twin children than adult twins but found no detectable relationships between BMI heritability and age among other types of family studies. We estimate heritabilities for height, weight, and BMI in a Caribbean village population to capture the proportional influence of genetic variance on body size and child growth during a population-wide period of nutritional transition using longitudinal anthropometric data.

Longitudinal data require within-individual analyses to account for variation in repeated measures over time in addition to between-individual analyses of phenotypic variation. Repeatabilities reflect how consistent traits remain for an individual as they age by regressing an individual's measurements against themselves over time, and any aspects of an individual's identity (including genetic and non-genetic attributes) that impact the observed phenotypes are captured in repeatability ratios (Wilson et al., 2010). Repeatability estimates generally indicate the upper limits for heritability estimates in a population given that individuals share 100% of their genetic variation with themselves (Dohm, 2002; Falconer and Mackay, 1996). Our repeatability estimates of height, weight, and BMI measure phenotypic variation within-individual growth curves during a period of nutritional and behavioral transition.

We expect repeatabilities and heritabilities to be higher for height than for weight or BMI given that height is less variable as an additive, long-term measure that remains

constant once adult height is reached until bone loss occurs at elderly ages (Dey et al., 1999). Weight and BMI are constrained by height but can fluctuate in response to short-term conditions across the lifespan. BMI is a composite measure of height and weight used to define clinical underweight, overweight, and obese categories, despite its variable relationship with adiposity and metabolic health across ethnicities and ancestries (Carroll et al., 2012; Hall & Cole, 2006; Prista et al., 2003). BMI remains a readily available metric of body size that may be more useful in diverse populations when tracked longitudinally over time to assess population-specific trends in changing body mass rather than relying on standardized cut-offs to categorize metabolic status (Hall & Cole, 2006). We capture secular trends in height, weight, and BMI from 1997-2017 in rural Dominica to assess how the global nutrition transition (Popkin et al., 2012) has influenced growth and body size in a small-scale horticultural population.

Body size, composition, and metabolic health in the Caribbean are uniquely impacted by sugar cultivation, other aspects of historic colonialism, alcohol production, tourism, and recent economic transition (Cherry et al., 2014; Mintz, 1985). Cardiometabolic health is an increasing concern in this region as cardiovascular disease and type II diabetes climb in prevalence (Rueda-Clausen et al., 2008). Sugar was widely cultivated throughout the Caribbean from the 18-20th centuries, mostly for export to European countries who transported enslaved laborers to the islands from west Africa in the 17-18th centuries (Mintz, 1985). Sugarcane is still grown throughout much of the Caribbean, and gene flow from Europe and Africa into indigenous Caribbean communities has shaped genetic variation throughout the region. Middle-income/wealthier Caribbean nations such as St. Lucia report negative metabolic

outcomes characteristic of the global nutrition transition (Popkin et al. 2012) that result from decreasing physical activity levels, increased alcohol consumption, urbanization, and changing diets due to imports and increased tourism (Cherry et al., 2014). Poorer, less developed Caribbean nations appear to be suffering similar health outcomes, but their data are sparse and secular trends poorly understood, particularly in rural areas (Boyne, 2009).

The Commonwealth of Dominica is one of the least developed Caribbean islands, and the village of Bwa Mawego is one of the most remote communities on the island. There are approximately 500-600 residents in Bwa Mawego, most of whom continue to practice traditional horticulture in tandem with increased access to cash goods and modern technology (e.g. cell phones, high-speed internet) (Decker and Flinn, 2011). Several varieties of taro, yams, and other root vegetables are the primary components of the traditional diet, which are boiled and eaten with plantains, other crops, and sometimes fish (Quinlan, 2004). Observational data from several decades of research at this field site indicate that processed foods, sweets, sugary beverages, and meat are increasingly available in local rum shops, transforming diets population-wide to include a combination of horticultural products and foods with more caloric sweeteners, oils, and animal products since the 2000's. Coinciding with dietary shifts characteristic of the global nutrition transition (Popkin et al., 2012), the transport of piped water and electricity to most homes in the village as of the early-mid 2000's has decreased physical activity demands (Decker & Flinn, 2011). We analyze longitudinal anthropometric data spanning 20 years (1997-2017) by capturing secular trends in growth curves during this transitional

period and estimating the relative influence of genetic variation on observed variation in height, weight, and BMI.

Materials and Methods

Longitudinal anthropometric data were collected in Bwa Mawego, Dominica at varying timepoints between 1997-2017 in accordance with procedures approved by the University of Missouri's Institutional Review Board. All participants provided informed consent, and parental consent was also obtained for all individuals under the age of 18 at the time of data collection. This study includes data for 260 individuals (126 males and 134 females) for whom there were repeated measures of height and weight over the study period that met our quality-control criteria, and ages of this sample range from birth to 27 years (Table 2.1). The number of repeated measures per individual ranges from 2-16 with a mean of 7.56, and the average time between a person's data points is 0.90 years, ranging from 4 months to 10 years (Table 2.1). The height of individuals old enough to stand upright was measured with a stadiometer on a flat surface; those too young to stand were laid on a flat surface and measured to the nearest millimeter by stretching a tape measure from heel to crown. Weight was measured using an electronic scale, and children too young to be weighed independently were weighed with a parent, whose weight was then subtracted from the combined total.

Table 2.1. Sample characteristics of longitudinal growth data.

N	260
Males	126 (48.5%)
Females	134 (51.5%)
Age range (years)	0-27
Data collection period	1997-2017 (mean=2002, SD=3.5 years)
Number of repeated measures	2-16 (mean=7.56, SD=4.85)
Time between data points (years)	0.36-10.01 (mean=0.90, SD=1.01)

Growth data were visually inspected and cleaned using the `Sitar` package in `R v.3.4.3` to account for errors in data collection over the 20-year study period and to remove outliers exceeding three standard deviations in an individual's growth curve (Cole, 2015; R Core Team, 2017). BMI was calculated using the standard equation (weight (kg)/height(m)^2). A pedigree that includes 11 generations was compiled for this village in the 2000's from interview data and historical records (Quinlan & Hagen, 2008), providing kinship coefficients needed for estimating trait heritabilities (Table S2.1; Figure S2.1).

We used Bayesian linear mixed models (LMMs) to analyze repeatabilities, heritabilities, and secular trends in longitudinal height, weight, and BMI among the sample of Bwa Mawego residents described above. Height, weight, and BMI were log-transformed to account for heteroscedasticity as variation increases in these variables with age. All LMMs included three fixed effects as control variables: age modeled as a cubic spline with knot points at 7 and 12 years, sex, and year of data collection (z-scored). Modeling age as a spline provided flexibility in accommodating these longitudinal data by allowing growth trends to vary between early childhood (birth-7 years), middle childhood (7-12 years), and adolescence (12+ years). This cubic polynomial spline allowed us to control for age across different stages of growth such that the relationships between anthropometrics and age later in adolescence were not impacted by trends very early in childhood, balancing complexity in different stages of growth with flexibility in a smooth curve that eases linear model fit (Buja et al., 1989; Harrell et al., 1988). The inclusion of data collection year in these LMMs controlled for secular trends across the 1997-2017 timespan during which these longitudinal data were

collected. This captured period effects independently from the effects of individual aging and indicates how height, weight, and BMI have changed among younger generations across this population during a period of nutritional and behavioral transition.

Repeatability LMMs included individual ID as a random effect, producing variance component estimates that measure the amount of variation in growth curves for height, weight, and BMI explained by an individual's ID, thereby capturing within-individual variation over time. Heritability models included two random effects: an ID variable to control for repeated measures and a second "identity" variable to connect each individual to the population's pedigree. This second random effect produced estimates of additive genetic variance by using Mendelian rules of allele sharing between individuals to explain the observed variation between their heights, weights, and BMIs independent of any variation within individual growth curves.

This method of estimating heritabilities is referenced as the "animal model" and uses complex multigenerational relationships (parents, siblings, half-siblings, grandparents, cousins, etc.) to capture the extent to which proportions of shared alleles influence the variance in observed traits (Kruuk, 2004; Teplitsky et al., 2008). This provides more robust heritability estimates than other common methods that rely on only two generations of parent-offspring relationships or single-generation twin studies. The pedigree for this village population in Dominica goes back 11 generations, ensuring that we capture as many kinship coefficients as possible, including relationships between more extended kin that are less likely to share common household environments (Table S2.1; Figure S2.1).

Although animal models allow for the inclusion of random household effects (Thomson et al., 2018), we do not include common household environment as a variance component in our models because of the flexibility and fluctuation in household composition in this matrifocal community. Households in Bwa Mawego fluctuate in composition as both children and adults change residence frequently related to short-term economic opportunities, temporary migrations, and changing family dynamics (Quinlan, 2004). Many people in this sample resided in more than one household over the timespan of their data points. In many cases, "households" are not discrete units, as many dwellings are organized to varying degrees into larger compounds with extended kin (Quinlan & Flinn, 2003). Additionally, the extensive depth of this population's pedigree reduces confounding effects of common environments since many smaller kinship coefficients contribute to estimates of genetic variance between relatives who do not share household environments.

Bayesian LMMs produce posterior probability distributions of fixed effect beta coefficients and random effect variance components by updating prior probability distributions with data (McElreath, 2015). This method of linear modeling is robust to sample size and accommodates complexity in regression-based analyses by controlling for repeated measures and accounting for other sources of heterogeneity that vary among individuals (Zhao et al., 2006). Fixed effects such as age and sex are simultaneously incorporated in a multivariable fashion such that the coefficients of each variable are measured independently from the rest. We captured Bayesian posterior distributions of repeatability and heritability variance components using Markov chain Monte Carlo sampling with the `MCMCglmm` package in `R` (Hadfield, 2010). This method samples

posteriors in a step-wise fashion rather than directly computing distributions in their entirety, which becomes less feasible as models increase in their complexity (McElreath, 2015).

All LMMs ran for 5,200,000 iterations with a burn-in of 200,000 and thin of 5,000 to produce 1,000 estimates of within-individual variance and additive genetic variance of height, weight, and BMI from the posteriors. We used parameter expanded priors for all models to facilitate chain mixing and obtain sufficient effective sample sizes (>900) by setting prior means to 0 `[alpha.mu=0]` and prior covariance matrices to 1,000 `[alpha.V=1000]` (Hadfield, 2010). Repeatabilities and heritabilities were calculated as variance component ratios, also representing 1,000 retained samples from the posteriors. Repeatability estimates reflect the proportion of total variation in each outcome due to the random effects of individual IDs (Equation 1), controlling for the fixed effects of sex, age, and data collection year. Heritability estimates reflect the proportion of phenotypic variation in each outcome captured by the additive genetic variance components derived from each individual's relatedness to everyone in the pedigree, also controlling for repeated measures, sex, age, and data collection year (Equation 2).

$$r = \frac{V_I}{V_I + V_e} \quad (1)$$

$$h^2 = \frac{V_A}{V_A + V_I + V_e} \quad (2)$$

V_I is the vector of 1,000 retained ID variance components that capture within-individual variation, V_A is the vector of additive genetic variance components, and V_e is the vector of residuals in each model. Ratios of these vectors produced 1,000 estimates of either

repeatability or heritability for each outcome from which we report posterior modes and credible intervals.

Results

The pedigree from Bwa Mawego, Dominica includes 1,455 individuals, spans 11 generations, and dates back to 1899 (Figure S2.1; Table S2.1). We have longitudinal growth data for 260 of those individuals, and the 662 people marked by dots in Figure S2.1 show the members of the pedigree who are related to them such that they contribute to estimates of genetic variance. Pedigree statistics calculated with the `Pedantics` package in `R` (Morrissey, 2018) show that inbreeding is negligible in this population despite the small founding structure of the community (Table S2.1).

Individual growth curves from this population are plotted with WHO percentiles overlaid for comparison (Figures 2.1-2.3). WHO growth percentiles range from birth-19 years for height and BMI, and from birth-10 years for weight (de Onis et al., 2007). Therefore, height and BMI plots include only individuals with 2 or more measurements recorded by age 19, and the weight plot includes only those with 2 or more measurements by age 10. We plot WHO curves for the 5^{th}, 50^{th}, 85^{th}, and 95^{th} percentiles due to their significance in reference to BMI (Barlow, 2007).

Growth curves show that males are slightly taller than females and females slightly heavier than males across all ages (Figures 2.1 and 2.2), and female BMI appears to increase substantially around age 12 (Figure 2.3). BMI is the most variable of the three traits, as we expect for such a composite trait with a less direct relationship to age than either height or weight. This variation is characterized by both fluctuations in individual

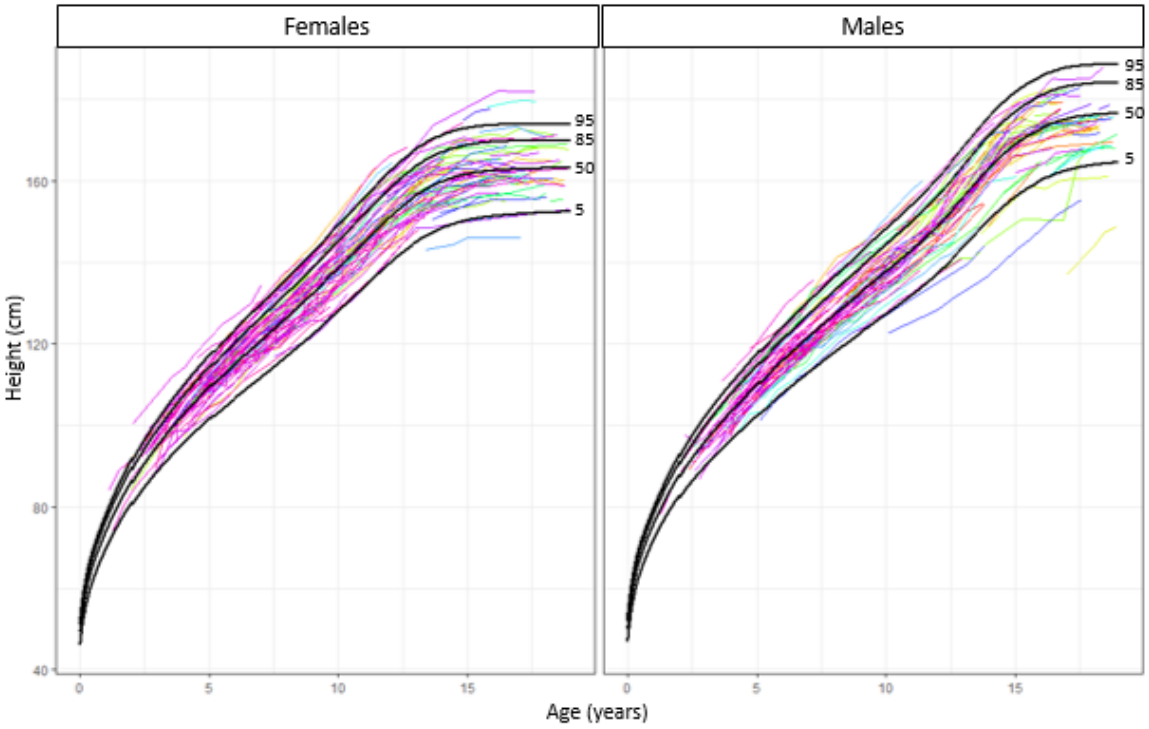

Figure 2.1. Height (cm) curves for 251 children (ages 0-19 years) with WHO percentiles for comparison.

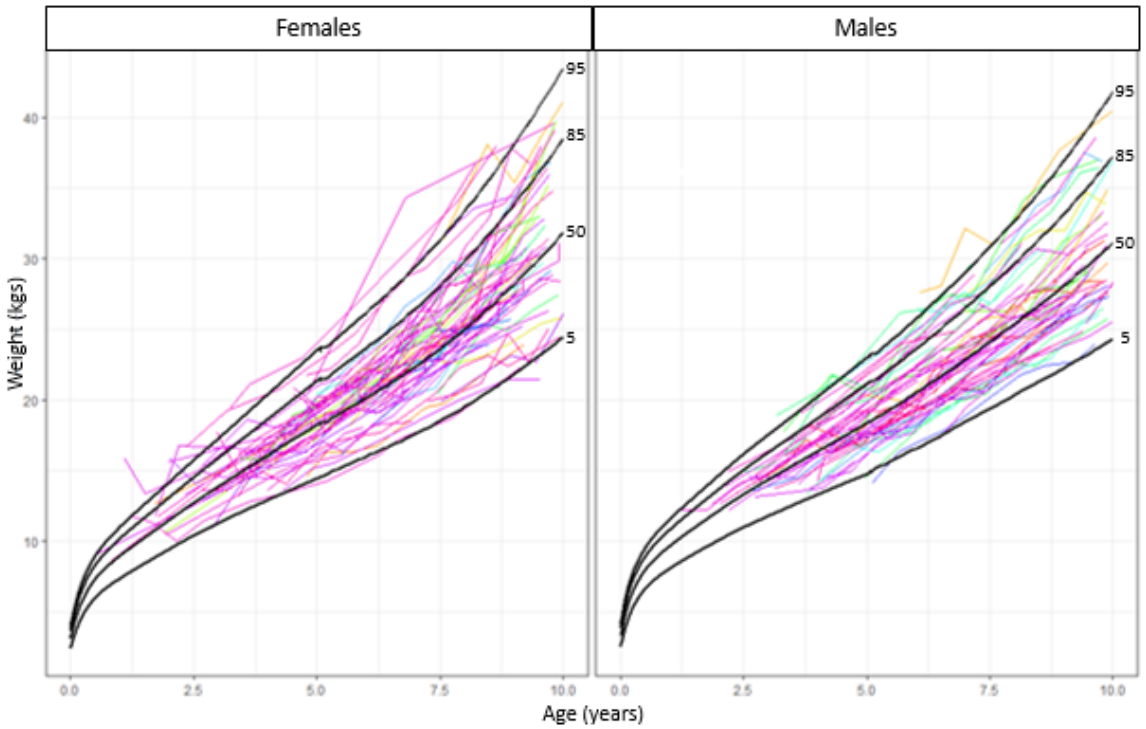

Figure 2.2. Weight (kg) curves for 158 children (ages 0-10) years with WHO percentiles for comparison.

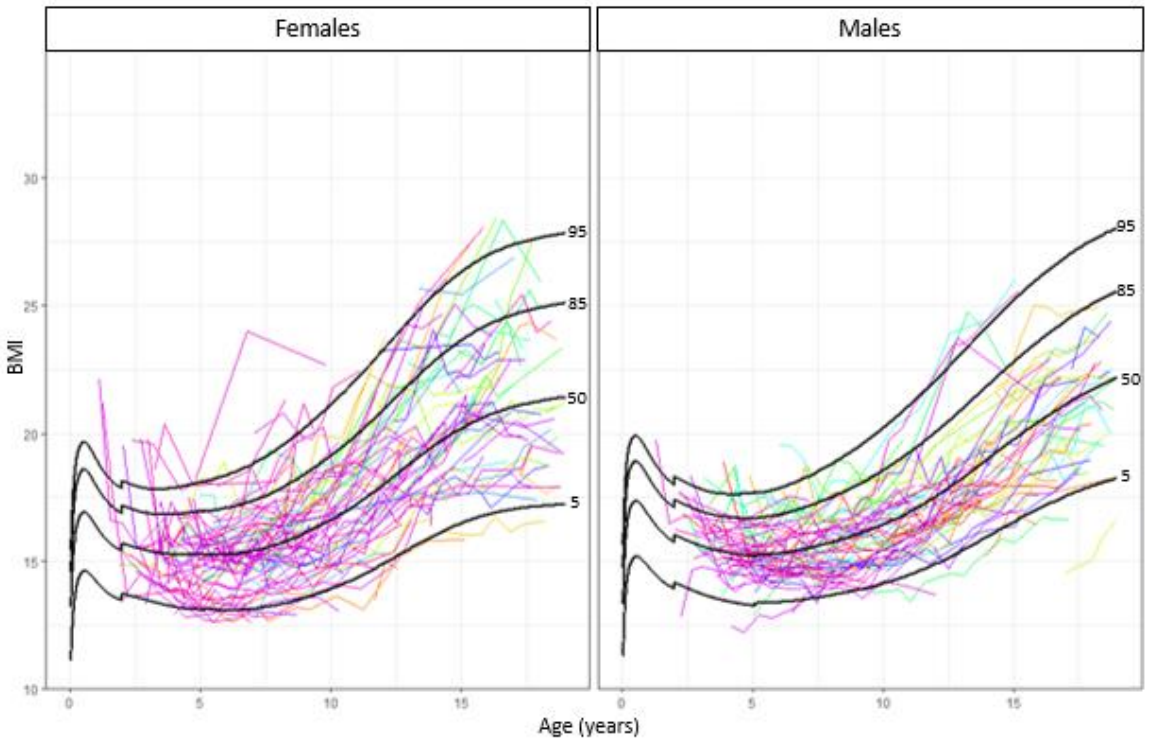

Figure 2.3. BMI curves for 251 children (ages 0–19) years with WHO percentiles for comparison.

BMI curves and the large population-wide spread of BMI measurements that increases substantially throughout adolescence (Figure 3). Children and teens who fall below the 5th BMI percentile are considered underweight by one broadly accepted clinical standard, those between 5-85% are healthy, those between 85-95% are considered overweight, and those above 95% are considered obese (Barlow, 2007). Table 2.2 displays descriptive statistics for child height, weight, and BMI in Bwa Mawego alongside height-for-age, weight-for-age, and BMI-for-age Z-score statistics based on the 2007 WHO reference tables (de Onis et al., 2007). Stunting (height-for-age < -2SD), underweight (weight-for-age < -2SD), and wasting (BMI-for-age <-2SD) appear to be uncommon in this population given that no more than 2 children fall into any one of these categories in any age class (Table 2.2). Higher proportions of children fall above 1 standard deviation in

Table 2.2. Descriptive anthropometrics with WHO-reference z-score statistics.

HEIGHT										
Age (years)	**N**	**Mean (cm)**	**SD (cm)**	**Range (cm)**	**Mean HAZ**	**HAZ SD**	**%<-2SD**	**%<-1SD**	**%>+1SD**	**%>+2SD**
Females										
0-3	21	90.13	6.35	(74.0, 100.3)	0.70	1.70	0 (0%)	3 (14.3%)	2 (9.5%)	4 (19.0%)
3-5	26	103.34	5.73	(95.4, 116.6)	0.00	0.97	1 (3.8%)	2 (7.7%)	3 (11.5%)	1 (3.8%)
5-10	32	119.95	9.49	(107.7, 142.2)	0.07	0.78	0 (0%)	1 (3.1%)	4 (12.5%)	0 (0%)
10-15	30	154.98	8.45	(142.5, 175.0)	0.03	1.03	1 (3.3%)	3 (10.0%)	6 (20.0%)	1 (3.3%)
15-19	18	162.38	5.94	(155.0, 178.1)	-0.06	0.91	0 (0%)	2 (11.1%)	1 (5.6%)	1 (5.6%)
Males										
0-3	13	90.62	5.27	(78.10, 98.55)	0.08	1.15	1 (7.7%)	1 (7.7%)	1 (7.7%)	1 (7.7%)
3-5	33	103.56	6.78	(90.6, 122.8)	0.00	1.19	1 (3.0%)	4 (12.1%)	3 (9.1%)	3 (9.1%)
5-10	29	124.30	10.30	(101.1, 138.5)	0.00	0.90	1 (3.4%)	2 (6.9%)	3 (10.3%)	1 (3.4%)
10-15	22	151.10	11.79	(122.7, 167.5)	-0.34	1.01	1 (4.5%)	4 (18.2%)	2 (9.1%)	0 (0%)
15-19	22	169.34	9.52	(137.3, 183.7)	-0.63	1.23	1 (4.5%)	4 (18.2%)	1 (4.5%)	0 (0%)
WEIGHT										
Age (years)	**N**	**Mean (kgs)**	**SD (kgs)**	**Range (kgs)**	**Mean WAZ**	**WAZ SD**	**%<-2SD**	**%<-1SD**	**%>+1SD**	**%>+2SD**
Females										
0-3	20	13.15	2.57	(8.62, 17.69)	0.99	1.37	0 (0%)	1 (5.0%)	6 (30.0%)	4 (20.0%)
3-5	27	16.70	2.55	(13.15, 21.77)	0.13	1.00	1 (3.7%)	3 (11.1%)	4 (14.8%)	1 (3.7%)
5-10	32	22.45	4.66	(15.06, 33.57)	-0.04	0.80	0 (0%)	3 (9.4%)	3 (9.4%)	0 (0%)
Males										
0-3	14	13.66	0.96	(12.25, 14.97)	0.53	0.59	0 (0%)	0 (0%)	2 (14.3%)	0 (0%)
3-5	33	16.74	2.56	(12.70, 24.04)	0.07	1.00	0 (0%)	7 (21.2%)	4 (12.1%)	1 (3.0%)
5-10	29	24.74	4.93	(14.06, 34.02)	0.03	1.04	1 (3.4%)	3 (10.3%)	5 (17.2%)	1 (3.4%)
BMI										
Age (years)	**N**	**Mean**	**SD**	**Range**	**Mean BAZ**	**BAZ SD**	**%<-2SD**	**%<-1SD**	**%>+1SD**	**%>+2SD**
Females										
0-3	16	17.04	2.83	(13.29, 22.13)	0.88	1.73	0 (0%)	2 (12.5%)	1 (6.3%)	6 (37.5%)
3-5	26	15.72	1.54	(12.95, 18.59)	0.24	1.06	0 (0%)	4 (15.4%)	6 (23.1%)	1 (3.8%)
5-10	32	15.48	1.69	(12.71, 20.05)	-0.16	1.00	1 (3.1%)	5 (15.6%)	2 (6.3%)	1 (3.1%)
10-15	30	19.10	2.65	(13.47, 25.64)	-0.01	1.00	1 (3.3%)	1 (3.3%)	6 (20.0%)	0 (0%)
15-19	18	21.90	4.27	(17.73, 33.85)	0.12	1.09	0 (0%)	2 (11.1%)	4 (22.2%)	1 (5.6%)
Males										
0-3	13	16.82	1.70	(12.83, 19.72)	0.65	1.34	1 (7.7%)	0 (0%)	5 (38.5%)	1 (7.7%)
3-5	33	15.57	1.46	(12.44, 18.89)	0.11	1.10	2 (6.1%)	1 (3.0%)	3 (9.1%)	3 (9.1%)
5-10	29	15.87	1.57	(12.97, 19.55)	0.01	1.09	1 (3.4%)	3 (10.3%)	2 (6.9%)	1 (3.4%)
10-15	22	18.01	2.66	(14.61, 25.42)	-0.18	0.99	1 (4.5%)	4 (18.2%)	0 (0%)	1 (4.5%)
15-19	22	20.59	2.08	(14.56, 23.93)	-0.23	0.90	1 (4.5%)	1 (4.5%)	0 (0%)	0 (0%)

BMI-for-age z-scores, but less than 10% are considered overweight (> +2SD) after age 3 by WHO standards (WHO, 2010).

Bayesian LMMs characterize the variation observed in height, weight, and BMI. Although sex and age were modeled as fixed effects in all LMMs, we do not report their coefficients since these control variables are better visualized (Figures 2.1-2.3), and age was modeled as a spline which complicates interpretations of its model coefficients. We report beta coefficient estimates for the period effect control (secular trends) in Table 2.3. The collection year variable used to measure this effect was z-scored (mean=2002, SD=3.5) and height, weight, and BMI were log-transformed as outcome variables.

Table 2.3. Growth LMM posterior modes and 90% credible intervals for within-individual variance (V_I), additive genetic variance (V_A), repeatability and heritability ratios, and secular trends.

	Height	**Weight**	**BMI**
V_I	0.019 (0.016, 0.021)	0.199 (0.170, 0.233)	0.107 (0.097, 0.131)
Repeatability	0.817 (0.790, 0.838)	0.813 (0.787, 0.841)	0.772 (0.743, 0.802)
V_A	0.014 (0.009, 0.022)	0.162 (0.105, 0.246)	0.071 (0.028, 0.112)
Heritability	0.683 (0.450, 0.836)	0.640 (0.453, 0.837)	0.487 (0.213, 0.704)
Beta (period)	0.011 (0.008, 0.014)	0.026 (0.016, 0.035)	0.005 (-0.003, 0.012)
R^2_m	0.924 (0.914, 0.934)	0.891 (0.876, 0.905)	0.580 (0.534, 0.614)
R^2_c	0.986 (0.986, 0.987)	0.981 (0.979, 0.982)	0.905 (0.893, 0.913)

R^2_m measures the proportion of observed variation explained by only the fixed effects of each heritability model (sex, age, and period effect), and R^2_c is that explained by all of the fixed and random effects (sex, age, period effect, within-individual variance, and additive genetic variance).

Exponentiating the beta coefficients indicates that for every 3.5-year increase over the 1997-2017 data collection period, height increased by approximately 1.01 cm, weight increased by approximately 1.03 kgs, and these secular trends are independent of age or sex (Table 2.3). Credible intervals indicate that height and weight show clearly increasing secular trends, whereas the 90% interval for BMI spans zero and does not show a significantly positive trend. Height, weight, and BMI in 5-year-olds plotted from 1997-

2007 show secular trends in the raw anthropometric data for a single age of significance and also support these model coefficients (Figures 2.4-2.6). We show secular trends at this age because body fat is typically at its lowest percentage between 5 and 6 years, and children who are overweight at this age show increased risks of metabolic disorders later in life (Moore et al., 2003; Nader et al., 2006). Height and weight show increasing secular trends in 5-year-olds during the 1997-2007 decade that are slightly higher in males than females (Figures 2.4-2.5). BMI trends are less clear, and there is a larger spread of variation in this metric than for either height or weight at this specific age (Figure 2.6). BMI appears to increase slightly in 5-year-old females between 1997-1999 before plateauing, and males show a modest increase between 2005-2007.

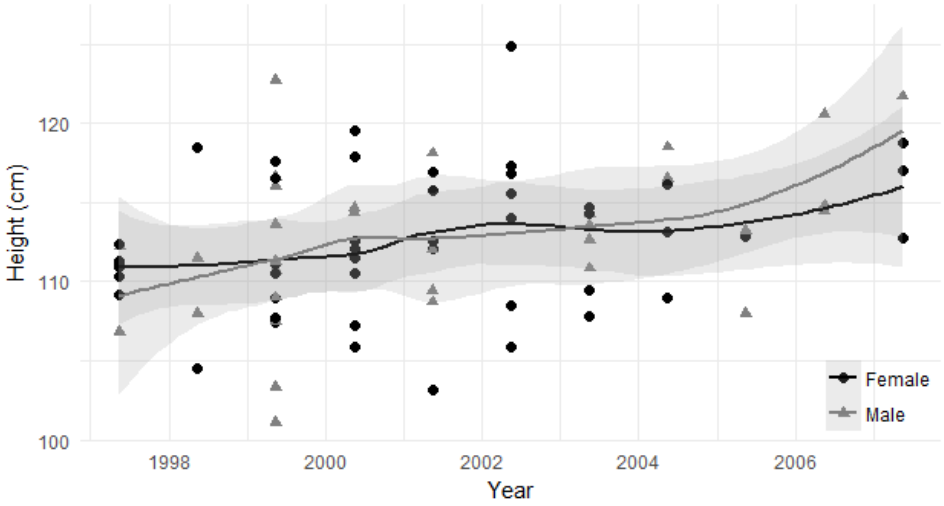

Figure 2.4. Height of children between ages 5 and 6 years from 1997-2007. 128 data points for 80 children (33 males and 47 females) are plotted with loess curves showing moving averages and 95% confidence intervals.

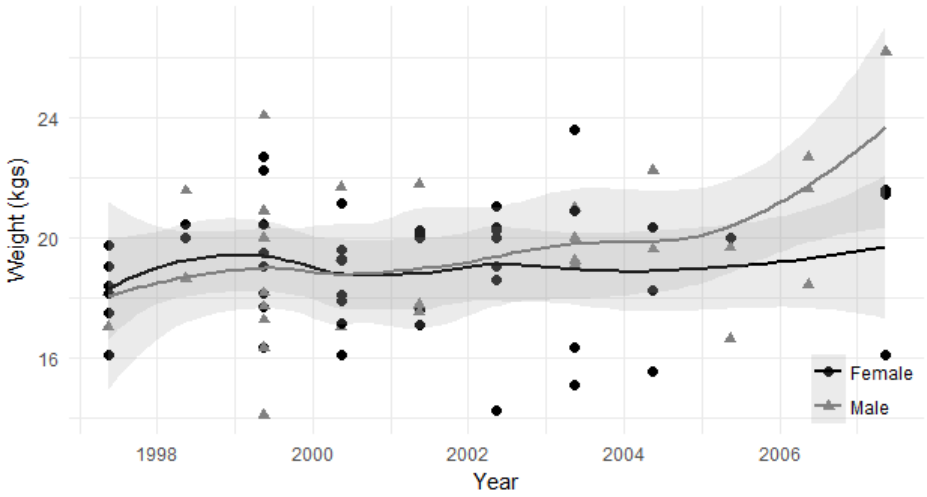

Figure 2.5. Weight of children between ages 5 and 6 years from 1997-2007. 128 data points for 80 children (33 males and 47 females) are plotted with loess curves showing moving averages and 95% confidence intervals.

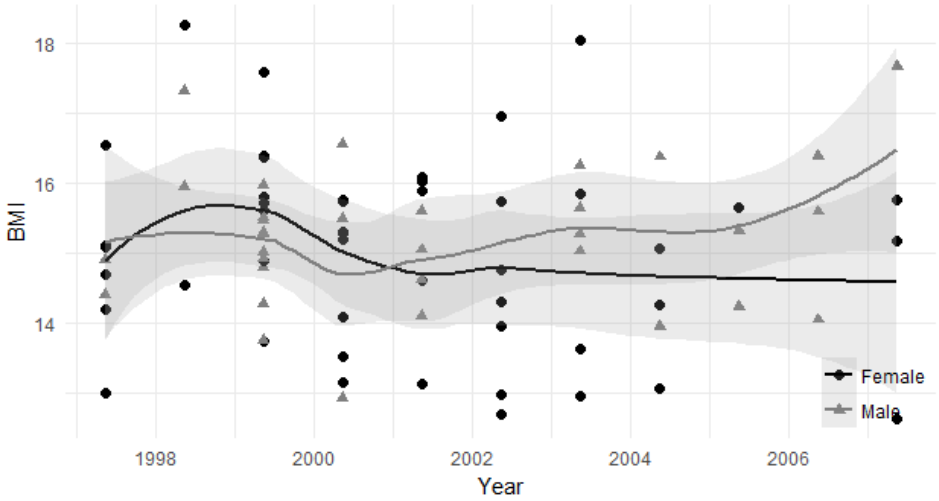

Figure 2.6. BMI of children between ages 5 and 6 years from 1997-2007. 128 data points for 80 children (33 males and 47 females) are plotted with loess curves showing moving averages and 95% confidence intervals.

LMM results show that height, weight, and BMI are all highly repeatable for the 260 individuals in this analysis (Table 2.3; Figure 2.7). We report modes and credible intervals to characterize posterior probability distributions. Unlike confidence intervals that reflect accuracy in reference to theoretical probability distributions, credible intervals

describe ranges of variation in parameter estimates to describe the shape of posterior distributions that have been produced by updating prior probability distributions with observed data (McElreath, 2015). Modes reflect the most probable beta coefficient and variance component values, and 90% credible intervals encompass 90% of the values sampled from posterior distributions. High posterior modes with small credible intervals indicate that approximately 82% of the variation in height and 81% of the variation in weight are explained by variance within individual growth curves in this population when also controlling for sex, age, and secular trends in these longitudinal data. BMI is less predictable than height or weight as individuals age, but variance within individuals still explains 77% of the population-wide variation not captured by age, sex, or secular trends (Table 2.3).

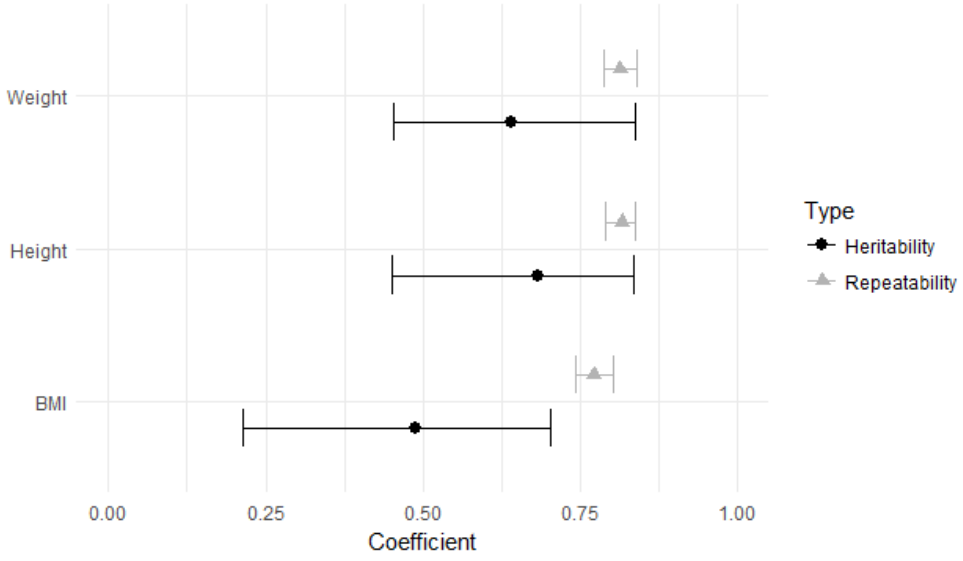

Figure 2.7. Repeatability and heritability estimates from Bayesian growth LMMs. Plotted modes and 90% credible intervals summarize 1,000 samples from posterior distributions of within-individual variance components and additive genetic variance components.

Heritability estimates reveal that, after accounting for the impacts of sex, age, and secular trends, genetic variation explains substantial proportions of variation observed

between individuals' height, weight, and BMI, and also explains large proportions of the repeatabilities modeled in the set of LMMs with only one random effect (Table 2.3). Heritability estimates reflect only the proportions of variation between individuals that are explained by shared genes because these LMMs also included individual IDs as a random effect to account for variation within individual growth curves. Thus, heritability estimates are independent of repeated measures within individuals, but repeatability estimates do encompass what is measured in heritabilities because all aspects of an individual's identity (including genetic and non-genetic attributes) that impact the observed phenotypes are captured in repeatability ratios. Additive genetic variance accounts for approximately 68% of the observed variation between individuals in height, 64% for weight, and 49% for BMI when also controlling for repeated measures within individuals. Although 90% credible intervals are much wider for heritabilities than repeatabilities, the lower limits for all heritability intervals are greater than 0.20, indicating that genetic variation significantly impacts phenotypic variation for all three traits (Figure 2.7).

We report two R^2 statistics defined specifically for LMMs by Nakagawa and Scheilzeth (2013) (Table 2.3). R^2_m values measure the proportion of variation in growth phenotypes explained by only the fixed effects of each model. Sex, age, and secular trends explain approximately 92% of the observed variation in height, 89% of the variation in weight, and 58% of the variation in BMI in this population. R^2_c values measure the proportion of variation in phenotypes explained by both fixed and random effects of each model (Nakagawa & Scheilzeth, 2013). We report conditional R^2 estimates from the set of heritability models, and the combination of sex, age, secular

trends, within-individual variance, and additive genetic variance explains approximately 99% of the observed variation in height, 98% of the variation in weight, and 91% of the variation in BMI (Table 2.3). These statistics indicate that age, sex, and secular trends explain the majority of anthropometric variation, leaving relatively small amounts of variation to be explained by within-individual and additive genetic variances. However, repeatability and heritability variance components account for much more of the variation observed in BMI than for variation observed in height or weight (Table 2.3).

Discussion

We analyzed longitudinal measures of body size in a small-scale Caribbean population that has recently transitioned nutritionally and behaviorally to include more Westernized dietary products and technologies alongside traditional subsistence horticultural practices. Height, weight, and BMI measurements track growth for 260 individuals in Bwa Mawego, Dominica. Individual BMI growth curves show large increases for many females in adolescence and into adulthood while more males appear to be overweight earlier in childhood (Figure 2.3).

Sex-specific differences in growth and variation are population-specific and age-dependent, related to environmental stressors, morbidity, gender-biased resource distributions, and life-history trade-offs (Stinson, 1985). Growth phenotypes from Bwa Mawego follow general patterns observed in other small-scale tropical societies in which males exhibit less variation than females (Walker et al., 2006), a pattern also seen in BMI across Australia and several European countries (Schousboe et al., 2003). The combination of higher levels of adiposity in females with substantial genetic variation in different patterns of fat distribution may contribute to greater variation in female versus

male BMI (Samaras et al., 1997). Additional data regarding more detailed body composition, morbidity, specific behavioral and dietary variables, and activity levels are needed to address potential underlying causes of patterns observed between male and female anthropometric variation in Bwa Mawego.

The individuals in this study range in age from birth to 27 years old throughout the 20-year data collection period (1997-2017). We modeled age, sex, and collection year simultaneously to capture secular trends independent of age or sex, and period effect beta coefficients show that height and weight have increased over these decades during which the population as a whole has gained access to imported and processed foods, piped water, electricity, and other resources such as internet and cell phones (Table 2.3). Averaged across age and sex, height has increased by approximately 5.8 cm (2.3 inches) and weight has increased by 5.9 kgs (13 lbs). Similar data from the Seychelles that span a nutritional transition show larger gains of 10-13 cm in height and 9-15 kgs in weight over a 50-year period comparing 15-year-old adolescents (Vidal et al. 2008). Secular trends in weight and BMI among U.S. children and adolescents show comparable increases in weight from 1960-2002 (+12-15 lbs), with relatively smaller gains in height (0.6-0.8 inches) to be expected in populations where stunting is less common (Ogden et al., 2004). Few individuals in Bwa Mawego fall into clinically defined overweight or obese categories at any timepoint in these longitudinal data (Table 2.2; Figure 2.3), and we do not find clear evidence of a population-wide increase in BMI in these younger generations (Table 2.3; Figure 2.6).

Age, sex, and secular trends account for the majority of variation in anthropometric phenotypes in Bwa Mawego, but far less in BMI than height or weight

(marginal R^2 in Table 2.3). Repeatabilities and heritabilities measure the proportions of phenotypic variation explained by within-individual and additive genetic variances that are residual to the variation explained by sex, age, and secular trends. Repeatability estimates show that aspects of an individual's identity, including both genetic and non-genetic factors such as behavior, are highly predictive of these anthropometrics as individuals age. All repeatabilities are greater than 75% (Table 2.3), leaving low residual variances unexplained in these repeated measures.

Heritability estimates for height (0.68), weight (0.64), and BMI (0.49) in Bwa Mawego are lower than many published estimates from twin studies (Elks et al., 2012; Silventoinen, 2003; Silventoinen et al., 2017), but well within the range of estimates from other types of family-based designs that are likely less inflated from common developmental environments than those shared by twins (Elks et al., 2012). We acknowledge that common environments may inflate our estimates of heritability slightly, however, flexible and fluctuating residence patterns in Bwa Mawego diffuse much of the household/spatial clustering known to influence anthropometric variation and heritability estimates in other populations (Heckerman et al., 2016; Saunders & Gulliford, 2006). Much of the variance in anthropometrics left unexplained by age, sex, and secular trends is attributed to additive genetic variance in this Caribbean population (Table 2.3; Figure 2.7), and further molecular research is needed to characterize specific genetic influences on variable anthropometric and metabolic health outcomes. This is particularly warranted in reference to body mass given that age, sex, and secular trends explain much less of the variation in BMI than in height or weight, bolstering the relative importance of genetic variation (Table 2.3).

Variation in BMI between populations is best explained by environmental, ecological, and behavioral factors, but most of the variation within populations appears to be explained by genetic variation (Bogardus, 2009). Despite estimating moderate to large heritabilities in a multitude of populations, geneticists have yet to account for most of this alleged genetic variation with specific variants, creating a problem of "missing" heritability. Diverse, small-scale populations that are under-represented in the current genetic literature may be valuable resources for discerning how biological, cultural, and environmental factors intersect to shape anthropometric variation and health on a more inclusive, global scale (Popejoy & Fullerton, 2016). Substantial contributions of additive genetic variance to anthropometric variation in this Caribbean population of mixed ancestry (Morena-Estrada et al., 2013) warrant further investigation, especially given the large amount of variation in body mass between individuals and the lack of population-wide secular trends throughout this transitional period (Table 2.3; Figure 2.3; Figure 2.6).

We have assessed the impacts of age, sex, secular trends, within-individual variance, and additive genetic variance on phenotypic variation in height, weight, and BMI in a Caribbean community that has recently transitioned to include more Western foods and technologies into traditional horticultural diets and subsistence practices. Anthropometric heritabilities are moderate in this population and body mass varies considerably between individuals, but metabolic health correlates of anthropometric variation remain unknown at this time. Additional data regarding specific behavioral, dietary, environmental, and genetic factors will enhance our understandings of anthropometric variation and health in the future.

Acknowledgements

We acknowledge funding support for this longitudinal data collection from the National Science Foundation (BCS-SBE 0640442, SBR-0136023, SBR-9205373, and BNS-8920569) and the University of Missouri Research Board. Many thanks to Robert Quinlan and Shane MacFarlane for providing the compiled genealogy, and to James Bain for compiling much of the growth database.

References

Barlow, S.E. (2007). Expert Committee Recommendations Regarding the Prevention, Assessment, and Treatment of Child and Adolescent Overweight and Obesity: Summary Report. *Pediatrics*, 120 (Supplement 4), S164–S192.

Bogardus, C. (2009). Missing Heritability and GWAS Utility. *Obesity*, 17, 209–210.

Boyne, M.S. (2009). Diabetes in the Caribbean: Trouble in paradise. *Insulin*, 4(2), 94–105.

Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear Smoothers and Additive Models. *The Annals of Statistics*, 17(2), 453–510.

Cherry, C.O., Serieux, E., Didier, M., Nuttal, M.E., & Schuster, R.J. (2014). Prevalence of Risk Factors for the Metabolic Syndrome in the Middle Income Caribbean Nation of St. Lucia. *Advances in Preventive Medicine*.

Carroll, J.F., Chiapa, A.L., Rodriquez, M., Phelps, D.R., Cardarelli, K.M., Vishwanatha, J.K., Cardarelli, R. (2008). Visceral Fat, Waist Circumference, and BMI: Impact of Race/ethnicity. *Obesity*, 16(3), 600–607.

Cole, T.J. (2003). The secular trend in human physical growth: a biological view. *Economics & Human Biology*, 1(2), 161–168.

Cole, T.J. (2015). Sitar: Super Imposition by Translation and Rotation Growth Curve Analysis. R package version 1.0.2. http://CRAN.R-project.org/package=sitar

de Onis, M., Blössner, M., Borghi, E., Frongillo, E.A., & Morris, R. (2004). Estimates of global prevalence of childhood underweight in 1990 and 2015. *JAMA*, 291(21), 2600–2606.

de Onis, M., Onyango, A.W., Borghi, E., Siyam, A., Nishida, C., & Siekmann, J. (2007). Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization*, 85, 660–667.

Decker, S.A., & Flinn, M.V. (2011). Parenting Styles and Gender-Linked Drinking Behaviors in Dominica. *Ethos*, 39(2), 189–210.

Dey, D.K., Rothenberg, E., Sundh, V., Bosaeus, I., & Steen, B. (1999). Height and body weight in the elderly. I. A 25-year longitudinal study of a population aged 70 to 95 years. *European Journal of Clinical Nutrition*, 53(12), 905.

Dohm, M.R. (2002). Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology*, *16*(2), 273–280.

Dubois, L., Ohm Kyvik, K., Girard, M., Tatone-Tokuda, F., Perusse, D., Hjelmborg, J., & Martin, N.J. (2012). Genetic and environmental contributions to weight, height, and BMI form birth to 19 years of age: An international study of over 12,000 twin pairs. *PLoS One*, 7(2), e30153.

Elks, C.E., den Hoed, M., Zhao, J.H., Sharp, S.J., Wareham, N.J., Loos, R.J.F., & Ong, K.K. (2012). Variability in the Heritability of Body Mass Index: A Systematic Review and Meta-Regression. *Frontiers in Endocrinology*, 3.

Godfrey, K.M., Gluckman, P.D., & Hanson, M.A. (2010). Developmental origins of metabolic disease: life course and intergenerational perspectives. *Trends in Endocrinology & Metabolism*, 21(4), 199–205.

Gravlee, C.C., Bernard, H.R., & Leonard, W.R. (2003). Boas's Changes in Bodily Form: The Immigrant Study, Cranial Plasticity, and Boas's Physical Anthropology. *American Anthropologist,* 105(2), 326–332.

Hadfield, J.D. (2010). Mcmc methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.

Hall, D.M.B., & Cole, T.J. (2006). What use is the BMI? *Archives of Disease in Childhood,* 91(4), 283–286.

Heckerman, D., Gurdasani, D., Kadie, C., Pomilla, C., Carstensen, T., Martin, H., Sandhu, M. S. (2016). Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proceedings of the National Academy of Sciences*, 113(27), 7377–7382.

Hoffmann, A.A., & Merila, J. (1999). Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution,* 14, 96–101.

Falconer, D.S. (1960). *Introduction to Quantitative Genetics*. New York: Ronald Press Co.

Falconer, D.S. & Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edition. Prentice Hall: London.

Flinn, M.V., Duncan, C.M., Ponzi, D., Quinlan, R.J., Decker, S.A., & Leone, D.V. (2012). Hormones in the Wild: Monitoring the Endocrinology of Family Relationships. *Parenting*, 12(2–3), 124–133.

Kruuk, L.E.B. (2004). Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359. 873–890.

Lewontin, R.C. (1974). The Analysis of variance and the analysis of causes. *American Journal of Human Genetics*, 26:400–11.

Livshits, G., Malkin, I., Williams, F.M.K., Hart, D.J., Hakim, A., & Spector, T.D. (2012). Longitudinal study of variation in body mass index in middle-aged UK females. *Age*, 34(5). 1285–1294.

Luke, A., Guo, X., Adeyemo, A.A., Wilks, R., Forrester, T., Lowe Jr, W., & Cooper, R.S. (2001). Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *International Journal of Obesity*, 25(7), 1034–1041.

McElreath, R. (2015). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press.

Mintz S.W. (1985). *Sweetness and Power*: *the place of sugar in modern history*. New York: Penguin Books.

Monteiro, C.A., Conde, W.L., & Popkin, B.M. (2004). The Burden of Disease from Undernutrition and Overnutrition in Countries Undergoing Rapid Nutrition Transition: A View from Brazil. *American Journal of Public Health*, 94(3), 433–434.

Moore, L.L., Gao, D., Bradlee, M.L., Cupples, L.A., Sundarajan-Ramamurti, A., Proctor, M.H., Ellison, R.C. (2003). Does early physical activity predict body fat change throughout childhood? *Preventive Medicine*, 37(1), 10–17.
Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., & Bustamante, C.D. (2013). Reconstructing the Population Genetic History of the Caribbean. arXiv e-print No. 1306.0558

Morrissey, M. (2018). pedantics: Functions to Facilitate Power and Sensitivity Analyses for Genetic Studies of Natural Populations. R package version 1.7.

Nader P.R., O'Brien M., Houts R., Bradley R., Belsky J., Crosnoe R., & Susman E.J. (2006). Identifying Risk for Obesity in Early Childhood. *Pediatrics*, 118, e594–e601.

Nakagawa, S., & Scheilzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed effects models. *Methods in Ecology and Evolution*, 4, 133–142.

Nan, C., Guo, B., Warner, C., Fowler, T., Barrett, T., Boomsma, D., & Zeegers, M. (2012). Heritability of body mass index in preadolescence, young adulthood and late adulthood. *European Journal of Epidemiology*, 27(4), 247–253.

Ogden, C.L., Fryar, C.D., Carroll, M.D., & Flegal, K.M. (2004). Mean body weight, height, and body mass index, United States 1960-2002. *Advance Data*, (347), 1–17.

Prista, A., Maia, J.A.R., Damasceno, A., & Beunen, G. (2003). Anthropometric indicators of nutritional status: implications for fitness, activity, and health in school-age children and adolescents from Maputo, Mozambique. *The American Journal of Clinical Nutrition*, 77(4), 952–959.

Popejoy, A.B., & Fullerton, S.M. (2016). Genomics is failing on diversity. *Nature News*, 538(7624), 161.

Popkin, B.M., Adair, L.S., & Ng, S.W. (2012). NOW AND THEN: The Global Nutrition Transition: The Pandemic of Obesity in Developing Countries. *Nutrition Reviews*, 70(1), 3–21.

Quinlan, M.B. (2004). From the Bush: The Front Line of Health Care in a Caribbean Village. Belmont, CA: Wadsworth.

Quinlan, R.J. (2006). Gender and Risk in a Matrifocal Caribbean Community: A View from Behavioral Ecology. *American Anthropologist*, 108(3), 464–479.

Quinlan, R.J., & Hagen, E.H. (2008). New Genealogy: It's Not Just for Kinship Anymore. *Field Methods*.

Quinlan, R.J., & Flinn, M. V. (2003). Intergenerational Transmission of Conjugal Stability in a Caribbean Community. *Journal of Comparative Family Studies*, 34(4), 569–583.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rueda-Clausen C.F., Silva F.A. & López-Jaramillo P. (2008). Epidemic of overweight and obesity in Latin America and the Caribbean. *International Journal of Cardiology*, 125(1), 111–112.

Samaras, K., Spector, T.D., Nguyen, T.V., Baan, K., Campbell, L.V., & Kelly, P.J. (1997). Independent genetic factors determine the amount and distribution of fat in women after the menopause. *The Journal of Clinical Endocrinology and Metabolism,* 82(3), 781–785.

Saunders, C.L., & Gulliford, M.C. (2006). Heritabilities and shared environmental effects were estimated from household clustering in national health survey data. *Journal of Clinical Epidemiology,* 59(11), 1191–1198.

Schousboe, K., Willemsen, G., Kyvik, K.O., Mortensen, J., Boomsma, D.I., Cornes, B.K., Harris, J.R. (2003). Sex Differences in Heritability of BMI: A Comparative Study of Results from Twin Studies in Eight Countries. *Twin Research and Human Genetics*, 6(05), 409–421.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, 35(2), 263–285.

Silventoinen, K., Jelenkovic, A., Sund, R., Yokoyama, Y., Hur, Y.-M., Cozen, W., Kaprio, J. (2017). Differences in genetic and environmental variation in adult BMI by sex, age, time period, and region: an individual-based pooled analysis of 40 twin cohorts. *The American Journal of Clinical Nutrition*, 106(2), 457–466.

Starkweather, K.E., & Keith, M.H. (2018). Estimating impacts of the nuclear family and heritability of nutritional outcomes in a boat-dwelling community. *American Journal of Human Biology*, 30(3), e23105.

Stinson, S. (1985). Sex differences in environmental sensitivity during growth and development. *American Journal of Physical Anthropology*, 28(S6), 123–147.

Urlacher, S.S., Blackwell, A.D., Liebert, M.A., Madimenos, F.C., Cepon-Robins, T.J., Gildner, T.E., Sugiyama, L.S. (2016). Physical growth of the Shuar: Height, Weight, and BMI references for an indigenous amazonian population. *American Journal of Human Biology*, 28(1), 16–30.

Marques-Vidal, P., Madeleine, G., Romain, S., Gabriel, A., & Bovet, P. (2008). Secular trends in height and weight among children and adolescents of the Seychelles, 1956–2006. *BMC Public Health*, 8, 166.

Vandemark, G.J., Kelly, M., & Eckhardt, R.B. (1985). A PASCAL program that calculates heritability estimates using weighted linear regressions. *Journal of Heredity*, 76(5), 400–401.

Visscher, P.M., Hill, W.G., & Wray, N.R. (2008). Heritability in the genomics era – concepts and misconceptions. *Nature Reviews Genetics*, 9(4), 255–266.

Vitzthum, V.J. (2003). A number no greater than the sum of its parts: the use and abuse of heritability. *Human Biology*, 75(4), 539–558.

Walker, R., Gurven, M., Hill, K., Migliano, A., Chagnon, N., Souza, R.D.,Yamauchi, T. (2006). Growth rates and life histories in twenty-two small-scale societies. *American Journal of Human Biology*, 18(3), 295–311.

Wells, J.C.K. (2012). Ecogeographical associations between climate and human body composition: Analyses based on anthropometry and skinfolds. *American Journal of Physical Anthropology*, 147(2), 169–186.

Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A. & Nussey, D.H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, 79(1), 13–26.

World Health Organization. (1995). Physical status: The use of and interpretation of anthropometry. Report of a WHO Expert Committee.

World Health Organization. (2010). Nutrition Landscape Information System (NLIS) country profile indicators: interpretation guide.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H. & Visscher, P.M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, 47(10), 1114–1120

Zhao, Y., Staudenmayer, J., Coull, B.A., & Wand, M.P. (2006). General Design Bayesian Generalized Linear Mixed Models. *Statistical Science*, 21(1), 35–51.

CHAPTER THREE

Genetic Ancestry and Population Structure

Abstract

The Caribbean is a genetically diverse region with heterogeneous admixture compositions that reflect local island ecologies, migrations, colonialist forces, and demographic histories. The Commonwealth of Dominica is an exceptionally mountainous island in the Lesser Antilles historically known to have unique pockets of ancestry and demographic structure. Single nucleotide polymorphism data from 159 people in a localized horticultural community provide insights on genetic ancestry and population structure in Dominica. We detect four distinct family clusters using `fastSTRUCTURE`, and there is clear evidence of admixture between African, European, and indigenous Amerindian ancestries that occurred approximately 130-150 years ago shown via phylogenetic methods in `TreeMix`, correlated linkage disequilibrium decay in `ALDER`, and visualizing Dominica samples and Human Genome Diversity Panel (HGDP) genotypes in Principal Component space. Our results are consistent with other genetic evidence that shows substantially higher proportions of indigenous ancestry and lower proportions of African ancestry in Dominica compared to variation sampled on other Caribbean islands. We detected significantly correlated linkage disequilibrium with Karitiana and Surui HGDP samples but not with other Amerindian groups, indicating that more deeply diverged indigenous lineages are likely present in Dominica that may not be widely represented elsewhere.

Introduction

The Caribbean is a diverse region where migration, local island ecologies, and colonialist forces have uniquely shaped demographic patterns and population structures (Belbin at el., 2018; Benn-Torres et al., 2008; Fitzpatrick and Keegan, 2007). The Commonwealth of Dominica is a mountainous island nation in the Lesser Antilles where exceptionally steep terrain is historically known to have provided refuge for people of African and indigenous ancestries fleeing enslavement by European colonists (Beckles, 1992; Quinlan, 2004). Given the challenges inherent to navigating mountainous tropical landscapes, little is known about the relatively isolated rural communities in this region (Montenegro and Stephens, 2006). We characterize population structure and genetic ancestry in a horticultural community on the windward coast of Dominica using high-density single nucleotide polymorphism (SNP) data from 159 people.

The 2010 Census counted 72,862 people among Dominica's ten parishes, and St. David's Parish along the eastern coast reports the highest proportions of indigenous Kalinago ancestry among its 6,043 residents (Pan American Health Organization, 2012). Historical documents and archaeological evidence indicate that at least two distinct waves of migration from South America brought the earliest human inhabitants to the island. Taino people of the South American Arawak group arrived around 400 CE, and Island Caribs, also known as Kalinago, followed around 1,000 CE (Beckles, 1992). An earlier population of Ortoiroid hunter-gatherers may have migrated from South America prior to the Taino as early as 3,000 BCE (Honychurch, 1995). Taino and Kalinago groups are known to have joined forces against Spanish invaders following Christopher Columbus's contact at Dominica in 1493, and it is estimated that the Kalinago population

declined by as much as 90% between the late 15th and early 18th centuries as Spanish, British, and French conquests reached the Lesser Antilles (Beckles, 1992). Labor from enslaved Africans, indigenous Carib groups, and European indentured servants in the 17th and 18th centuries enabled a mix of French and British plantations to produce coffee and sugar, respectively, until approximately 14,000 enslaved people were legally emancipated in Dominica in 1834 (Beckles, 1992; Honychurch, 1995).

Bwa Mawego is a rural horticultural community in Dominica located on the island's steep eastern coast, south of the indigenous Kalinago reserve (Quinlan, 2004). This village is one of the most remote on the island and is thought to have been populated by newly emancipated people who settled in the exceptionally steep windward landscape during the mid-19th century. The majority of Bwa Mawego's residents (approximately 500 at any given timepoint) have been engaging in anthropological and psychological health research for the past 30 years (Flinn et al., 1999; Macfarlan et al., 2012; Quinlan, 2004). Cardiometabolic health and related chronic conditions are growing local concerns in Bwa Mawego. Much of the variation observed in longitudinal health traits is explained by genetic variation in this population (Keith et al, 2019), yet genetic variation in this region has yet to be explored in detail. We aim to characterize population structure and genetic ancestry in order to explore how recent admixture has shaped this Caribbean community using high-density genotype data from more than 30% of Bwa Mawego's residents.

Caribbean and Latin American groups are heterogeneous in their ancestral compositions, and indigenous components of these genomes are particularly variable (Belbin et al., 2018; Moreno-Estrada et al., 2013). People with recently mixed ancestries

are under-represented in genetic research (Bustamante et al., 2011; Popejoy and Fullerton, 2016), and relatively isolated communities are known to have otherwise rare genetic variants reach high frequencies, reflecting unique local histories and founder effects (Belbin et al., 2018; Hunley and Healy, 2011). As genetic variants become increasingly informative in managing complex diseases, including uniquely admixed genomes enhances our understanding of polygenic traits and mitigates European bias in genetic research (Pulit et al., 2010).

An analysis of admixture throughout the English-speaking Caribbean used a targeted set of ancestry informative markers (AIMs) and found significantly more indigenous and European ancestry in a sample of 37 people from Dominica relative to more African ancestry sampled on other islands, indicating that patterns of genetic admixture in Dominica are particularly unique (Benn-Torres et al., 2013). Our samples represent more than 30% of a rural horticultural community in Dominica and capture genome-wide variation with a high-density SNP array (Illumina Human OmniExpress). Genetic research that is inclusive to people from ancestrally heterogeneous populations, such as those in the Caribbean and Latin America, requires sampling and analyses at finer scales (Belbin et al., 2018). We describe population structure and admixture in a community that is both culturally and geographically defined in a unique region of the Caribbean.

Materials and Methods

We extracted DNA from buccal swabs to produce genotype data from 160 people in Bwa Mawego, Dominica. These data were collected during July-August 2017 following research approvals from both the University of Missouri Institutional Review

Board and the local village council in Dominica. All participants gave written informed consent prior to any data collection, and parental consent was also obtained for participants under the age of 18.

Isohelix buccal swabs were stabilized at room temperature using Dri-Capsules (Boca Scientific) during data collection, and samples were extracted with the Buccal-Prep Plus DNA Isolation Kit (Boca Scientific) and purified with the MinElute PCR Purification Kit (Qiagen). 160 samples were genotyped for 960,923 SNPs on the Human OmniExpress BeadChip (Illumina, 2018). This high-density array has genome-wide coverage and captures variants across diverse human populations defined in the 1,000 Genomes Project (Illumina, 2018).

We filtered SNP data using `PLINK v1.9` (Purcell et al., 2007) to remove SNPs with call rates < 0.9 or Hardy-Weinberg Equilibrium (HWE) p-values < $1x10^{-40}$ and individuals with call rates < 0.9. Filtering removed 1,181 SNPs due to low call rate, one SNP due to HWE p-value, and one individual due to low call rate. For population-comparison analyses, we used genotyped individuals from the Human Genome Diversity Panel (HGDP) (Cann et al., 2002). Individuals in the HGDP were genotyped on Illumina 650Y array. We filtered the HGDP data as above, removed populations with less than five individuals, then merged with the remaining 159 Dominica samples. The resulting dataset contained 1,078 total individuals (159 Dominica and 919 HGDP) genotyped at 468,721 SNPs shared across panels.

We inferred population structure and admixture proportions via *K*-means clustering in a variational Bayesian framework using `fastSTRUCTURE` (Raj et al., 2014). Allowing the number of clusters (*K*) to vary from 2-10, we assessed genetic

clustering within only the Dominica genotypes as well as among the Human Genome Diversity Panel (HGDP) with a subset of Dominica samples (Cann et al., 2002). Using `PLINK's '--rel-cutoff'` flag, we down-sampled individuals from Dominica to exclude close relatives ($r<0.025$, $n=22$). This subset of Dominican individuals was used for clustering with the HDGP dataset in `fastSTRUCTURE` to infer ancestry with less confounding family-based structure. In order to visualize potential sex-biased admixture, we also ran `fastSTRUCTURE` to compare clustering between autosomes and X chromosomes for 336 females from the down-sampled dataset.

We used a generalized linear mixed model (GLMM) to compare genetic clustering among all 159 genotyped individuals from Bwa Mawego, Dominica with an 11-generation village-wide pedigree that dates back to 1899 (Table S2.1). We used the cluster affinities for the *K* number of clusters that had the highest likelihood from `fastSTRUCTURE` as the outcome variable in this GLMM. Using the `MCMCglmm` package in `R v.3.6.3`, we modeled the pedigree-derived kinship matrix as a random effect to assess the extent to which genetic population structure in Bwa Mawego is explained by close family relatedness (Hadfield, 2010; R Core Team, 2019).

We ran a principal component analysis (PCA) in `PLINK` with the HGDP SNPs and projected all 159 samples from Dominica onto the HGDP space. This enabled visualization of the Dominican genotypes with globally diverse samples while preventing our relatively large Caribbean sample from disproportionately influencing the principal components that reflect global genetic variation more broadly.

We used *f* statistics to test for admixture in Bwa Mawego, Dominica and visualized historical relationships between these Caribbean samples and populations in

the HGDP dataset using the '`threepop`' and '`fourpop`' programs implemented in `TreeMix` (Pickrell and Pritchard, 2012). HGDP reference populations serve as proxies for actual ancestral populations from which we expect to detect admixture. f_3 statistics test the phylogenetic structure underlying allele frequencies among three different populations (Peter, 2016; Reich et al., 2009), operating from a non-admixed null hypothesis that variation in allele frequencies follows a tree-like process of population differentiation over time with positive branch lengths. f_4 statistics test the tree-like structure among four populations, allowing for one internal branch that will have a length of zero among populations with no detectable admixture (Peter, 2016; Reich et al., 2009). Using f_4 ratio estimation, f_4 statistics can be used to estimate ancestry contributions from two diverged populations in an admixed population of interest (Patterson et al., 2012).

In f_4 ratio estimation, the mixing parameter is calculated as $\frac{f4\ (A,O;X,C)}{f4\ (A,O;B,C)}$ where *B* and *C* are populations hypothesized to have formed *X*, *A* is a sister population to *B*, and *O* is an outgroup to *A*, *B*, *C*, and *X*. is interpreted as an estimate of the relative contribution of population *B* to *X*, while the relative contribution of *C* (β) is 1-α. We estimated f_4 admixture ratios using four different combinations of African, European, and Amerindian HGDP populations informed by initial f_3 results. Neither f_3 nor f_4 statistics directly test for admixture in a fourth population from three divergent source populations as we expect to find in Dominica, therefore, we interpret these phylogenetic tests with context from admixture analyses.

We used `ALDER` to date admixture events and infer minimum mixture proportions in Bwa Mawego, Dominica by assessing correlations of linkage disequilibrium decay with HGDP reference populations (Loh et al., 2013). Recombination events increasingly

dissociate alleles from one another each generation with a likelihood that increases with genetic distance along each chromosome, thus detailed evolutionary relationships can be inferred between admixed and reference populations based on the decaying correlation of linkage disequilibrium across increasingly long chromosome tracts (Moorjani et al., 2011). We ran `ALDER` with pairs of HGDP reference populations, and also with individual reference populations one at a time, to estimate the timing of admixture events and the mixture proportions among 159 people in rural Dominica. Together, these clustering, dimension reduction, phylogenetic, and linkage disequilibrium analyses characterize population structure in a localized horticultural community and reveal historical admixture in a unique area of the Caribbean.

Results

We assessed population structure in Bwa Mawego, Dominica from 468,721 SNPs in a sample of 159 people using `fastSTRUCTURE` (Raj et al., 2014). The two values of *K* reported from `fastSTRUCTURE` indicate the number of clusters that maximize the marginal likelihood of observed genetic variation (Raj et al., 2014), and the lower bound estimate of K=4 indicates that we are most likely to observe this population-wide variation from four well-defined genetic groups (Figure 3.1). The upper bound estimate of K=9 accounts for additional weaker population structure in Bwa Mawego. We utilize the well-defined K=4 clusters in subsequent analyses. GLMM results that compared these four cluster affinities with an 11-generation population-wide pedigree (Figure S2.1; Table S2.1) show that these four groups largely reflect family relatedness. Pedigree-based relatedness

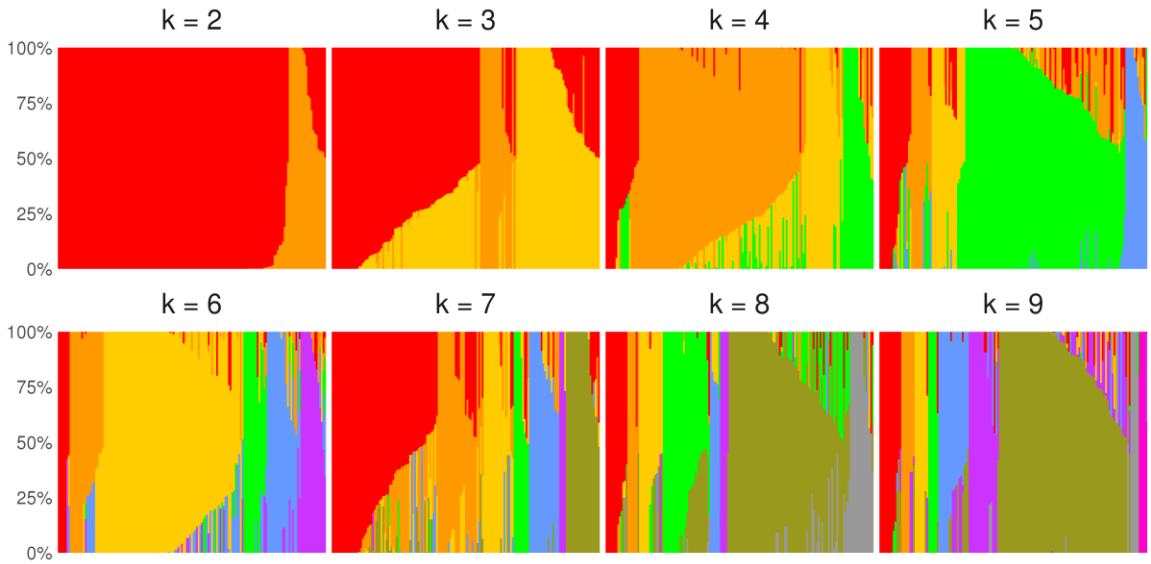

Figure 3.1. Genetic structure in Bwa Mawego, inferred from `fastSTRUCTURE`. Each color represents a cluster (for *K* number of clusters in each model), and each individual (n=159) is a vertical line with colored segments that correspond to the proportion of cluster similarity.

explains approximately 99% of red cluster affinity, 77% of orange affinity, 72% of yellow affinity, and 98% of green cluster affinity (for K=4) estimated from GLMM variance component ratios.

We inferred ancestry clusters in Bwa Mawego using a subset of 22 individuals that excludes close relatives ($r<0.025$) in reference to data from 919 people across 53 populations in the HGDP dataset (Cann et al., 2002). For the lower bound of strong genetic structure, `fastSTRUCTURE` indicates that six clusters maximize the marginal likelihood across the HDGP data and 22 Dominica samples, and seven clusters maximize the amount of variation explained when accounting for additional weaker structure (Figure 3.2). Bwa Mawego samples share ancestry with African, European, and Amerindian populations in substantial proportions, showing clear evidence of admixture from three genetically distinct ancestries.

K=5 maximizes the marginal likelihood of ancestral variation among only females sampled from Dominica and the HGDP. Comparing cluster affinities along autosomes to X chromosomes shows 10-13% less European admixture along X chromosomes in Dominica than along autosomes (Figure 3.3).

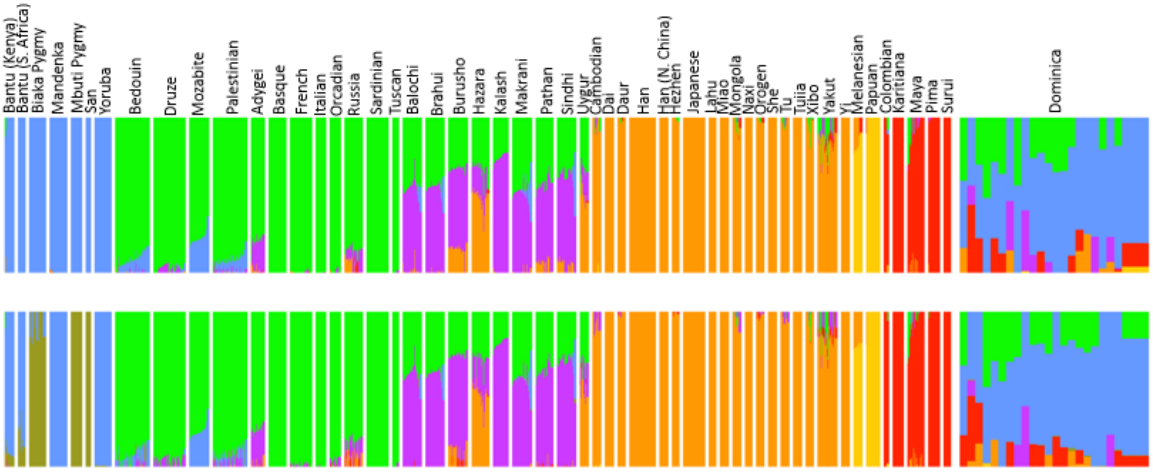

Figure 3.2. Ancestry proportions inferred from `fastSTRUCTURE` among 919 samples from the HGDP and a subset of 22 unrelated Dominica samples for *K*=6 (top panel) and *K*=7 (bottom panel).

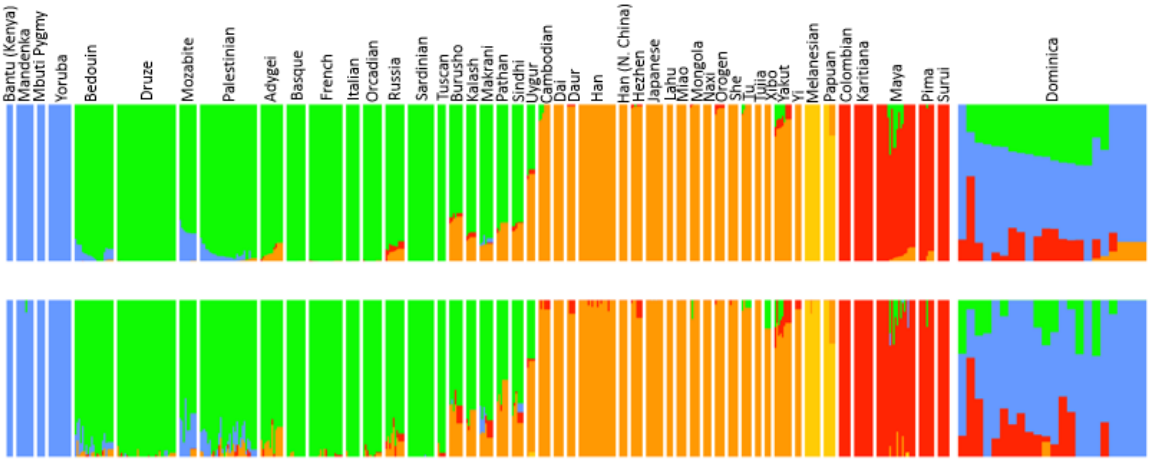

Figure 3.3. Ancestry proportions of autosomes (top panel) and X chromosomes (bottom panel) among females, inferred from `fastSTRUCTURE` (*K*=5).

We derived the first two principal components in the HGDP reference dataset in `PLINK` and then mapped the loadings of all 159 genotypes from Dominica onto that two-

dimensional space (Figure 3.4). As shown consistently in other HGDP analyses, the first principal component captures increased genetic variation in African populations relative to non-African groups, and the second differentiates European and Asian clusters (Lawson et al., 2012; Li et al., 2008). Samples from Dominica form a diffuse cluster spread along this African-European vector and also show varying similarity to Amerindian HGDP samples along PC2 (Figure 3.4).

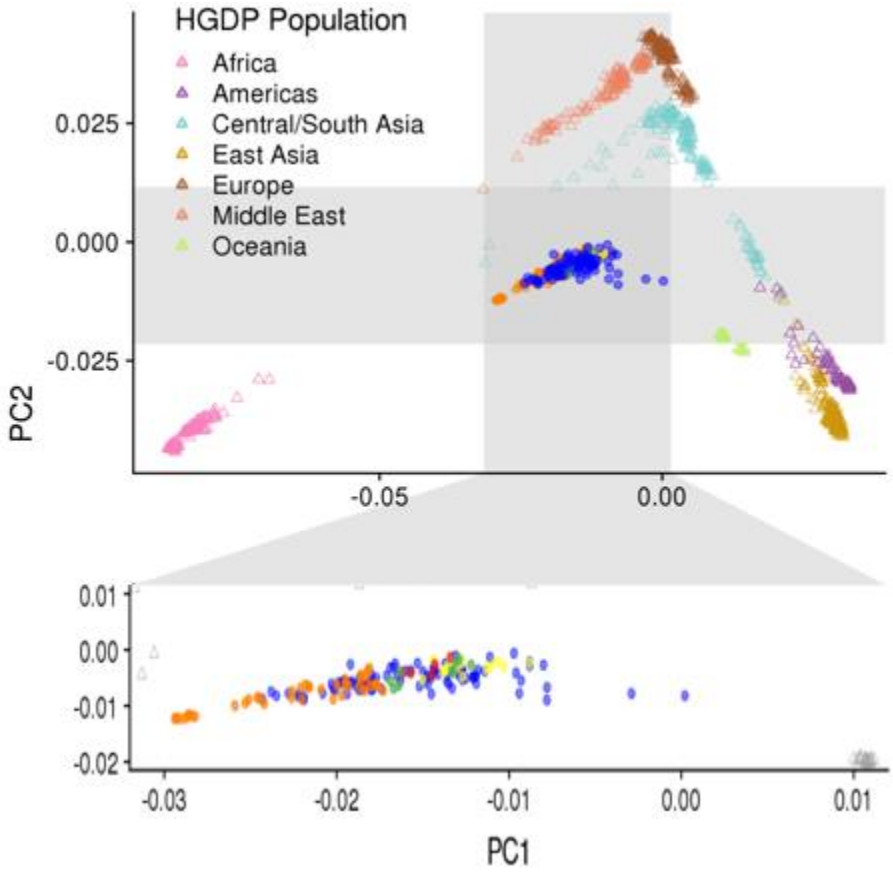

Figure 3.4. Dominica samples (n=159) projected onto HGDP PCA axes (PC1= first HGDP principal component, PC2=second HGDP principal component). HGDP loadings are plotted as triangles, and Dominica loadings are plotted as closed circles with colors that correspond to the four clusters identified in `fastSTRUCTURE` (Figure 1). Individuals were assigned a matching color (red, orange, yellow, or green) if one of the four `fastSTRUCTURE` clusters accounted for more than 90% of their genetic variation, and un-assigned individuals who do not meet this cluster affinity threshold are plotted in blue.

Phylogenetic inferences from `TreeMix` indicate that individuals in Bwa Mawego, Dominica share the most ancestry with African populations and more similarity

with individuals sampled from Yoruba and Mandenka populations than with San groups (Table 3.1). *f3* statistics show the most significant negative branch lengths between Dominica and Yoruba, French, and Karitiana samples (Table 3.1). Strong genetic drift and founder effects can mask signals of admixture captured by this statistic (Patterson et al., 2012), yet we detect highly significant negative branch lengths in these data. We detect significant admixture from African/European and African/Amerindian source pairs but not from European/Amerindian pairs (Table 3.1). f_4 admixture ratios estimate a larger contribution to ancestry in Bwa Mawego from African populations than European populations, and a larger contribution from European populations than Amerindian populations (Table 3.2).

We ran two-reference admixture models in `ALDER` for a subset of African, European, and Amerindian HGDP populations in relation to all 159 genotypes from Dominica and report date estimates from the reference pairs with significant LD correlations (Table 3.3). Correlated LD begins to significantly decay beginning at lengths around 1.00 centimorgan. With a generation length of 25 years, two-reference weighted LD curves indicate that admixture occurred between approximately 130-150 years ago in rural Dominica, with slightly more recent date estimates from European and Amerindian ancestries (Table 3.3). Mixture proportions from single-reference models in `ALDER` support our f_4 admixture ratio results (Table 3.2), indicating that at least 40% of genetic ancestry in Bwa Mawego is African, more than 20% is European, and at least 6% is shared with indigenous Amerindian ancestors (Table 3.4) captured by the HGDP data (Cann et al., 2002).

Table 3.1. f_3 statistics from `TreeMix`. Negative values indicate non-phylogenetic relationships as evidence of admixture and positive values represent branch lengths resulting from phylogenetic relationships.

Admixed	Ref A	Ref B	f_3	Standard error	Z-score
Dominica	Yoruba	French	-0.008	0.00011	-72.44
Dominica	Yoruba	Orcadian	-0.008	0.00012	-69.44
Dominica	Yoruba	Pima	-0.010	0.00015	-65.91
Dominica	Karitiana	Yoruba	-0.011	0.00017	-64.52
Dominica	Yoruba	Surui	-0.011	0.00017	-63.71
Dominica	Mandenka	French	-0.007	0.00011	-63.05
Dominica	Mandenka	Orcadian	-0.007	0.00012	-61.77
Dominica	Mandenka	Karitiana	-0.010	0.00017	-60.67
Dominica	Mandenka	Pima	-0.009	0.00016	-60.30
Dominica	Mandenka	Surui	-0.010	0.00017	-58.94
Dominica	San	Orcadian	-0.007	0.00016	-40.31
Dominica	San	French	-0.006	0.00016	-40.04
Dominica	San	Surui	-0.009	0.00024	-38.35
Dominica	San	Karitiana	-0.009	0.00024	-37.56
Dominica	San	Pima	-0.008	0.00022	-37.25
Dominica	Surui	French	0.014	0.00030	45.67
Dominica	Surui	Orcadian	0.014	0.00030	46.52
Dominica	Karitiana	Orcadian	0.014	0.00028	50.77
Dominica	Karitiana	French	0.014	0.00026	52.07
Dominica	Orcadian	Pima	0.015	0.00027	55.44
Dominica	French	Pima	0.014	0.00025	55.81
Dominica	San	Yoruba	0.011	0.00014	76.83
Dominica	Mandenka	San	0.011	0.00014	77.54
Dominica	French	Orcadian	0.026	0.00027	97.47
Dominica	Mandenka	Yoruba	0.012	0.00012	101.39
Dominica	Surui	Pima	0.047	0.00043	107.80
Dominica	Karitiana	Pima	0.047	0.00043	109.39
Dominica	Karitiana	Surui	0.053	0.00048	110.86

Table 3.2. f_4 ratio estimates of pairwise mixture proportions.

Population	Position	Estimate	f_4 test
French	B	0.370	(San,Basque;Yoruba,Dominica)/(San,Basque;Yoruba,French)
Yoruba	C	0.630	(San,Basque;Yoruba,Dominica)/(San,Basque;Yoruba,French)
Orcadian	B	0.370	(San,Basque;Yoruba,Dominica)/(San,Basque;Yoruba,Orcadian)
Yoruba	C	0.630	(San,Basque;Yoruba,Dominica)/(San,Basque;Yoruba,Orcadian)
Karitiana	B	0.261	(San,Surui;Yoruba,Dominica)/(San,Surui;Karitiana,Yoruba)
Yoruba	C	0.739	(San,Surui;Yoruba,Dominica)/(San,Surui;Karitiana,Yoruba)
Surui	B	0.260	(San,Karitiana;Yoruba,Dominica)/(San,Karitiana;Yoruba,Surui)
Yoruba	C	0.740	(San,Karitiana;Yoruba,Dominica)/(San,Karitiana;Yoruba,Surui)

Table 3.3. Dates of admixture in Dominica computed in ALDER from pairs of HGDP reference populations. Date estimates are in generations, and genetic distances (d) are the lengths (in centimorgans (cM)) at which weighted LD correlations significantly decay.

Ref 1	Ref 2	d (cM)	Weighted LD amplitude	Date estimate	z-score
Mandenka	Orcadian	1.20	0.00086 ± 0.00003	5.89 ± 0.22	27.15
Yoruba	Orcadian	1.60	0.00089 ± 0.00003	5.81 ± 0.22	26.75
Mandenka	French	1.80	0.00083 ± 0.00003	5.81 ± 0.23	25.66
Yoruba	French	1.80	0.00088 ± 0.00003	5.80 ± 0.23	25.45
Mandenka	Karitiana	1.20	0.00124 ± 0.00004	5.49 ± 0.25	21.89
Yoruba	Karitiana	1.60	0.00128 ± 0.00005	5.44 ± 0.25	21.43
Mandenka	Surui	1.20	0.00124 ± 0.00004	5.47 ± 0.26	21.32
Yoruba	Surui	1.60	0.00128 ± 0.00005	5.41 ± 0.26	21.14
Surui	Orcadian	1.00	0.00035 ± 0.00002	5.40 ± 0.30	17.92
Karitiana	Orcadian	1.10	0.00033 ± 0.00001	5.39 ± 0.32	16.64
French	Surui	1.80	0.00035 ± 0.00002	5.26 ± 0.33	16.00
French	Karitiana	1.80	0.00033 ± 0.00001	5.26 ± 0.34	15.29

Table 3.4. Dates of admixture in Dominica computed in ALDER from single HGDP reference populations. Date estimates are in generations, and mixture proportions are lower bound estimates. Genetic distances (d) are the lengths (in centimorgans (cM)) at which weighted LD correlations significantly decay.

Ref pop	d (cM)	Weighted LD amplitude	Date estimate	z-score	Mixture %
Yoruba	1.40	0.00023 ± 0.00001	5.42 ± 0.25	21.57	40.7 ± 0.8
Mandenka	1.00	0.00021 ± 0.00001	5.43 ± 0.28	19.35	34.5 ± 0.8
French	1.60	0.00033 ± 0.00001	6.32 ± 0.25	24.84	20.4 ± 0.5
Orcadian	0.80	0.00034 ± 0.00001	6.42 ± 0.26	25.09	19.4 ± 0.4
Karitiana	0.90	0.00070 ± 0.00001	5.62 ± 0.34	16.63	6.7 ± 0.3

Discussion

We found signals of admixture between African, European, and indigenous Amerindian ancestries that occurred approximately 130-150 years ago in rural Dominica (Figure 3.2; Tables 3.1-3.4), with the largest proportion of African ancestry and smallest proportion of indigenous ancestry (Tables 3.2 and 3.4). Oral history in Bwa Mawego indicates that a Canadian family founded the community prior to 1840 when it first appeared on historic maps (Quinlan, 2004), and the mid-late 19th century admixture date estimates that we detected indicate that the community quickly grew to include people of African, European, and indigenous ancestries in the decades between its founding and the 20th century.

The current population of Bwa Mawego has mostly African ancestry, a substantial proportion of European ancestry, and at least 6% of these genotype data correlate with indigenous Amerindian ancestry (Figure 3.2; Table 3.4). Bwa Mawego is geographically less than ten kilometers from the only indigenous reservation in the Caribbean (Kalinago), yet ALDER shows consistent decay curves across relatively short lengths of *d* (<2.0 cM) that indicate admixture was localized to a specific time period (Loh et al., 2013) an average of 5-6 generations ago (Tables 6-7). Continuous admixture is characterized by LD that extends over longer (>10 cM) chromosome tracts (Pfaff et al., 2001) and tends to produce more varied decay parameters in ALDER (Loh et al., 2013). Mixture proportion estimates from ALDER are lower bounds, and our ability to detect indigenous ancestry in these admixed genotypes depends on how similar surviving indigenous lineages in the Lesser Antilles are to those sampled in the HGDP Amerindian

reference groups, which are proxies for ancestral source populations (Montinaro et al., 2015).

Some admixture introgression in Bwa Mawego appears to be sex-biased (Figure 3.3), and 10-13% less European ancestry along X chromosomes compared with autosomes suggests that there may have been a higher proportion of European males than females among the community's founders. This is consistent with historical accounts and genetic data across the English-speaking Caribbean, and there is prior evidence for relatively higher proportions of non-African male admixture in Dominica based on Y chromosome short tandem repeats (STRs) (Benn-Torres et al., 2007). The small number (5-6) of generations (Tables 3.3-3.4) since admixture occurred in Bwa Mawego's founding generations lead us to interpret our X chromosome comparisons with caution, since calculated mixture fractions from males and females can oscillate for up to 5-10 generations as they approach their mean proportion limit in an admixed population (Goldberg and Rosenberg, 2015).

Consistent with other admixture analyses across the Americas and Caribbean, the closest HGDP match to the African ancestry in Bwa Mawego is Yoruba (Tables 3.1 and 3.4) (Montinaro et al., 2015). We identified significantly correlated LD between Dominica and Yoruba, Mandenka, and Bantu South Africa groups, but not with Bantu Kenya (Table 3.3). We also detected correlated LD with indigenous Karitiana and Surui groups, but not with Pima. Karitiana and Surui are Amazonian groups historically known to have been isolated from other indigenous and non-indigenous groups, and they share a small portion of genetic ancestry with indigenous groups in Australia and Oceania that is not found in other Amerindian populations (Skoglund et al., 2015). Although at least 40%

of the ancestry in Bwa Mawego is African, this estimate is significantly lower than African mixture components in other Caribbean populations and is consistent with another ancestry analysis that includes samples from Dominica (Benn-Torres et al., 2013). These results indicate that Dominica's admixture composition is especially unique and includes significantly more indigenous ancestry than other Caribbean groups (Benn-Torres et al., 2013), with lineages that may not be widely represented among other Caribbean or Amerindian populations.

We found strong genetic structure within Bwa Mawego, Dominica, and the four distinct clusters in this community appear to reflect relatedness and family-based similarity more so than differing admixture proportions between clusters (Figure 3.4). Despite some evidence that Dominica has less genetic diversity than other Caribbean islands that could potentially reflect founder effects and confound demographic signals (Benn-Torres et al., 2007), we were able to detect clear admixture signals and family-based structure to form a detailed analysis of structured SNP variation (Figures 3.1-3.2; Tables 3.3-3.4). The ability to distinguish arbitrary and admixed structure impacts our ability to characterize linkage signals and map traits in heterogeneous populations (Belbin et al., 2018; Pfaff et al., 2001). Our results highlight the need to assess detailed demographic histories in diverse populations since we detect clear, localized admixture signals as well as structured clusters based on demography following the initial admixture in Bwa Mawego (Figures 3.1-3.2; Tables 3.3-3.4).

As modern globalization continues to transform patterns of migration and genetic variation, admixture is becoming more common at the individual level such that it creates substantial genetic heterogeneity among people sampled from cosmopolitan areas in

study designs that do not account for genetic structure or sub-structure based on ancestry and demography (Cooper et al., 2008). Genetic variation is increasingly informing our understandings of human biological variation and health traits, and admixed populations enhance our ability to map complex, polygenic traits more precisely when structure is modeled explicitly. Mapping causal genetic variants depends on our ability to isolate them from surrounding LD, and admixed populations often have heterogeneous distributions of ancestral haplotypes that enhance our ability to detect trait associations and localize genetic signals (Cooper et al., 2008; Patterson et al., 2010).

While we detect clear admixture signals in Bwa Mawego, Dominica, admixture compositions vary among individuals (Figure 3.2) and do not cluster neatly according to the four-group structure that is specific to this rural community (Figures 3.2 and 3.4). Unique genetic ancestry and haplotype structure in combination with longitudinal health data in this localized horticultural community (Keith et al., 2019) may present unique opportunities to identify biologically significant genetic variants through admixture mapping and other analyses that utilize population structure to inform gene-trait, gene-gene, and gene-environment associations (Patterson et al., 2010). Additionally, the shared local ecology reduces environmental heterogeneity in this small-scale community compared to variation in lifestyles, resource access, etc. represented among participants in most genotype-phenotype studies (Kulminski et al., 2016). The combination of relatively low environmental heterogeneity, diverse admixture compositions, and clearly defined population structure indicate that this culturally and geographically defined community presents unique opportunities for admixture mapping, epigenetic exploration, and association analyses.

Acknowledgements

We thank the participants who provided genetic samples for these analyses, the community of Bwa Mawego as a whole, and our two local research assistants Nethilia Durand and Nakisha Wiltshire. Many thanks to Kristen Taylor and Alex Stuckel for providing lab space and equipment for DNA extractions and sample processing. This research was funded by the University of Missouri Research Board.

References

Beckles H.M. (1992). Kalinago (Carib) resistance to European colonization of the Caribbean. *Caribbean Quarterly*. 1992: 77–94.

Belbin, G.M., Nieves-Colón, M.A., Kenny, E.E., Moreno-Estrada, A., & Gignoux, C.R. (2018). Genetic diversity in populations across Latin America: implications for population and medical genetic studies. *Current Opinion in Genetics & Development*, 53, 98–104. https://doi.org/10.1016/j.gde.2018.07.006

Benn-Torres, J., Bonilla, C., Robbins, C.M., Waterman, L., Moses, T.Y., Hernandez, W., & Kittles, R.A. (2008). Admixture and Population Stratification in African Caribbean Populations. *Annals of Human Genetics*, 72(1), 90–98. https://doi.org/10.1111/j.1469-1809.2007.00398.x

Benn-Torres, J., Kittles, R.A., & Stone, A.C. (2007). Mitochondrial and Y Chromosome Diversity in the English-Speaking Caribbean. *Annals of Human Genetics*, *71*(6), 782–790. https://doi.org/10.1111/j.1469-1809.2007.00380.x

Benn-Torres, J., Stone, A.C., & Kittles, R. (2013). An anthropological genetic perspective on creolization in the anglophone Caribbean. *American Journal of Physical Anthropology*, 151(1), 135–143. https://doi.org/10.1002/ajpa.2226

Boca Scientific. Buccal Swabs/DNA Extraction. www.bocascientific.com. Accessed February 2020.

Bustamante, C.D., Burchard, E.G., & De La Vega, F.M. (2011). Genomics for the world. *Nature*, 475(7355), 163–165. https://doi.org/10.1038/475163a

Cann, H.M., Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., & Cavalli-Sforza, L.L. (2002). A Human Genome Diversity Cell Line Panel. *Science*, 296(5566), 261–262. https://doi.org/10.1126/science.296.5566.261b

Cooper, R.S., Tayo, B., & Zhu, X. (2008). Genome-wide association studies: implications for multiethnic samples. *Human Molecular Genetics*, 17(R2), R151–R155. https://doi.org/10.1093/hmg/ddn263

Fitzpatrick, S.M., & Keegan, W.F. (2007). Human impacts and adaptations in the Caribbean Islands: an historical ecology approach. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, 98(1), 29–45. https://doi.org/10.1017/S1755691007000096

Flinn, M.V., Leone, D.V., & Quinlan, R.J. (1999). Growth and Fluctuating Asymmetry of Stepchildren. *Evolution and Human Behavior*, 20(6), 465-479.

Goldberg, A. & Rosenberg, N.A. (2015). Beyond 2/3 and 1/3: The Complex Signatures of Sex-Biased Admixture on the X Chromosome. *Genetics*, 201, 263-279. doi:10.1534/genetics.115.178509

Hadfield, J.D. (2010). Mcmc methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.

Honychurch, L. (1995). The Dominica Story: A History of the Island. Macmillan.

Hunley, K., & Healy, M. (2011). The impact of founder effects, gene flow, and European admixture on native American genetic diversity. *American Journal of Physical Anthropology*, 146(4), 530–538. https://doi.org/10.1002/ajpa.21506

Kulminski, A.M., Loika, Y., Culminskaya, I., Arbeev, K.G., Ukraintseva, S.V., Stallard, E., & Yashin, A.I. (2016). Explicating heterogeneity of complex traits has strong potential for improving GWAS efficiency. *Scientific Reports*, 6(1), 35390. https://doi.org/10.1038/srep35390

Illumina. (2018). OmniExpress-24 v1.3 BeadChip Datasheet. www.illumina.com. Accessed February 2020.

Lawson, D.J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLOS Genetics*, 8(1), e1002453. https://doi.org/10.1371/journal.pgen.1002453

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., & Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* (New York, N.Y.), 319(5866), 1100–1104. https://doi.org/10.1126/science.1153717

Macfarlan, S.J., Remiker, M., & Quinlan, R. (2012). Competitive Altruism Explains Labor Exchange Variation in a Dominican Community. *Current Anthropology*, 53(1), 118–124. https://doi.org/10.1086/663700

Montenegro, R.A., & Stephens, C. (2006). Indigenous health in Latin America and the Caribbean. *The Lancet*, 367(9525), 1859–1869. https://doi.org/10.1016/S0140-6736(06)68808-9

Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., & Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. *Nature Communications*, 6(1), 1–7. https://doi.org/10.1038/ncomms7596

Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., & Atzmon, G. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genetics* 7: e1001373.

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J.K., Gignoux, C.R., & Bustamante, C.D. (2013). Reconstructing the Population Genetic History of the Caribbean. *PLOS Genetics*, 9(11), e1003925. https://doi.org/10.1371/journal.pgen.1003925

Pan American Health Organization. (2012). Dominica. Health in the Americas, 2012 Edition: Country Volume.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, 192(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Petersen, D.C., van der Ross, R.E., Sudoyo, H., Glashoff, R.H., Marzuki, S., & Hayes, V.M. (2010). Genetic structure of a unique admixed population: implications for medical research. *Human Molecular Genetics*, 19(3), 411–419. https://doi.org/10.1093/hmg/ddp505

Peter, B.M. (2016). Admixture, Population Structure, and F-Statistics. *Genetics,* 202(4), 1485–1501. https://doi.org/10.1534/genetics.115.183913

Pfaff, C L., Parra, E.J., Bonilla, C., Hiester, K., McKeigue, P.M., Kamboh, M.I., & Shriver, M.D. (2001). Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *American Journal of Human Genetics*, 68(1), 198–207. https://doi.org/10.1086/316935

Pickrell, J.K. & Pritchard, J.K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* 8(11): e1002967. https://doi.org/10.1371/journal.pgen.1002967

Popejoy, A.B., & Fullerton, S.M. (2016). Genomics is failing on diversity. *Nature News*, 538(7624), 161. https://doi.org/10.1038/538161a

Pulit, S.L., Voight, B.F., & de Bakker, P.I.W. (2010). Multiethnic Genetic Association Studies Improve Power for Locus Discovery. *PLoS ONE*, 5(9). https://doi.org/10.1371/journal.pone.0012600

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.

Qiagen. Products: DNA Clean Up. www.qiagen.com. Accessed February 2020.

Quinlan, M. (2004). From the Bush: The Front Line of Health Care in a Caribbean Village. Wadsworth: Belmont, CA.

Raj, A., Stephens, M., & Pritchard, J.K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, 197(2), 573–589. https://doi.org/10.1534/genetics.114.164350

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Reich, D., K. Thangaraj, N. Patterson, A.L. Price, and Singh, L. (2009) Reconstructing Indian population history. *Nature* 461: 489–494.

Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, 525(7567), 104–108. https://doi.org/10.1038/nature14895

CHAPTER FOUR

Genome-Wide Association (GWA) Tests of Cardiometabolic Traits

Abstract

Increasing diverse representation in genome-wide association studies (GWAS) enhances our understanding of genotype-phenotype associations and mitigates strong European sample bias in genetic research. Bwa Mawego is a small-scale horticultural community in Dominica with increasing cardiometabolic health concerns. We tested genotype-phenotype associations between cardiometabolic phenotypes and 371,835 single nucleotide polymorphisms (SNPs) among 159 people in this Caribbean community of mixed continental ancestry. After assessing family relatedness and ancestral population structure to obtain kinship coefficients and structural principal components, we ran mixed model single-variant association tests for anthropometric traits, blood pressure, and glucose levels that controlled for confounding variation from sex, age, population structure, and relatedness as covariates. We detect a potentially significant association between body mass index (BMI) and 2 SNPs downstream of the *ANK3* gene on chromosome 10. Further research is needed to investigate any functional relationship between *ANK3* and body mass, and there is unique potential to explore complex gene-environment-phenotype landscapes in Bwa Mawego, Dominica.

Introduction

The genetic components of most health phenotypes remain largely uncharacterized (Manolio et al., 2009), and genome-wide association studies (GWAS) remain a primary tool for detecting novel associations to incrementally increase the proportions of variation explained and enhance our functional understandings of complex traits and diseases (Visscher et al., 2017). GWAS have limited ability to detect associations given that they rely on common genetic variants in linkage disequilibrium (LD) with causal features, are sensitive to genetic architecture (allele frequency and variant effect size distributions), and must account for confounding variables (ancestral variation, population structure, phenotypic covariates, etc.) in order to detect associations between genetic loci and target phenotypes. These features are population- and sample-specific (Cooper et al., 2008). Strong European bias in GWAS samples thus far leaves much genetic variation unexplored and under-represented, potentially skewing our understandings of genetic risk and trait variation (Bustamante et al., 2011). We test genotype-phenotype associations for cardiometabolic traits using 371,835 single nucleotide polymorphisms (SNPs) in a small-scale Caribbean community of mixed continental ancestry.

Bwa Mawego is a rural horticultural community of approximately 500 people in the Commonwealth of Dominica that has engaged in longitudinal health research for several decades (Flinn et al., 1999; Macfarlan et al., 2012; Quinlan, 2004). Cardiometabolic phenotypes (including type II diabetes and hypertension) are primary local health concerns, and we measured anthropometrics (height, weight, arm circumference, triceps skinfold), blood pressure, and glucose levels among a majority of

the community's residents in 2017, providing point of care information to participants. Longitudinal anthropometric data indicate that despite increasing market integration (including the introduction of processed foods) over the past two decades, secular trends do not predict variation in body mass index (BMI) in Bwa Mawego, and a substantial proportion of observed BMI variation is explained by shared genetic variation (h^2=0.49) (Keith et al., 2019). We further explore the genetic components of phenotypic variation in this study using SNP data to test single variant-trait associations.

Bwa Mawego was established in the mid-19th century in the steepest, harshest terrain along Dominica's windward coast (Quinlan, 2004). Dominica has uniquely mixed ancestry that includes higher proportions of Amerindian ancestry relative to other Caribbean islands (Benn-Torres et al., 2013). Bwa Mawego has minimum proportions of 40% African, 20% European, and 6% Amerindian ancestry with LD patterns that indicate representation of deeply diverged indigenous South American lineages found among Karitiana and Surui peoples (Keith et al., in prep). Genotype-phenotype tests may identify novel associations in Bwa Mawego given its unique representations of multiple ancestries and enriched family lineages. Furthermore, environmental sources of variation are relatively reduced among this small-scale horticultural population compared with the variation in lifestyles and environments represented among participants in most GWAS designs (Kulminski et al., 2016). Reduced non-genetic heterogeneity in Bwa Mawego may enhance our ability to detect meaningful genotype-phenotype associations.

In order to mitigate strong European bias and increase diverse representation in genetic research, we must explicitly model population structure and other confounding sources of heterogeneity that impact the ability to detect variant-trait associations. This is

essential among people with recently-mixed ancestries, including a large proportion of the world's population in Latin America and the Caribbean where recent historical patterns of admixture vary regionally (Belbin et al., 2012). The population structure of Bwa Mawego, Dominica is well-defined (Keith et al., in prep), and we test associations between autosomal SNPs and cardiometabolic phenotypes while controlling for confounding impacts of age, sex, population structure, and family relatedness in this small-scale Caribbean community.

Methods

Genotype and phenotype data were collected during July-August 2017 following research approvals from both the University of Missouri Institutional Review Board and the local village council in Dominica. All participants gave written informed consent prior to any data collection, and parental consent was also obtained for participants under the age of 18. Height was measured with a stadiometer on a flat surface, weight was measured using a digital scale, and body mass index (BMI) was calculated with the standard equation $[weight(kg)/height(m)^2]$. Tricep skinfolds were measured with manual calipers and upper arm circumference was measured with a tape measure. Blood pressure was measured using a digital arm cuff after participants had been seated with feet flat on the ground and backs supported for a minimum of 10 minutes. Participants were surveyed about their known diabetic status. We tested glucose concentrations in urine samples using Bayer Diastix Glucose Reagent Strips for Urinalysis by dipping test strips into each sample at the time of collection and recording the strip color after 30 seconds.

DNA samples were collected using Isohelix buccal swabs that were stabilized at room temperature with Dri-Capsules (Boca Scientific). Samples were extracted with the Buccal-Prep Plus DNA Isolation Kit (Boca Scientific) and purified with the MinElute PCR Purification Kit (Qiagen). 160 samples were genotyped for 960,923 single nucleotide polymorphisms (SNPs) on the Human OmniExpress BeadChip (Illumina, 2018). This high-density array has genome-wide coverage and captures variants across diverse human populations defined in the 1,000 Genomes Project (Illumina, 2018).

SNP data were filtered using the `SNPRelate` package in `R v.4.0.0` (Zheng et al., 2012; R Core Team, 2020). We excluded non-autosomal and monomorphic SNPs and removed SNPs with call rates < 0.90, minor allele frequencies (maf) <0.10, or Hardy-Weinberg Equilibrium p-values $<1x10^{-6}$ as well as individuals with call rates <0.90 (Marees et al., 2018). 24,306 SNPs were non-autosomal, 185,415 were monomorphic, and we removed 1,181 SNPs for low call rates, 185,383 SNPs for maf<0.10, 25,446 SNPs for low Hardy-Weinberg p-values, and one individual for low call rate. The 159 individuals retained in these analyses range from 5-88 years old and include 88 males and 71 females.

After the initial filtering, we pruned SNPs for linkage disequilibrium (LD) <0.2 in order to obtain a set of independent markers for measuring relatedness and population structure (Marees et al., 2018). We used this independent set of 31,968 SNPs to measure identity by descent (IBD) with the `KING` moment estimator method in the `SNPRelate` package (Zheng et al., 2012). The kinship coefficient matrix derived from this IBD analysis was then included in a Principal Component Analysis (PCA) in order to adjust for family relatedness when analyzing ancestral population structure using the `PC-AiR`

function in GENESIS (Gogarten et al., 2019). We used the recommended kinship threshold of 0.022 to partition the dataset into related and unrelated subsets. PC-AiR used the unrelated subset of 32 individuals for a traditional PCA that returned 32 eigenvectors to then predict PC values for the related subset of 127 people. We retained the first 2 PCs from this PCA in our subsequent analyses upon visualization of the variable loadings and scree plot (Figure S4.1) in order to account for population structure in this community of recently mixed ancestry (Keith et al., in prep). We then re-calculated IBD probabilities and kinship coefficients with GENESIS using the PC-Relate function that included the first 2 PCs from PC-AiR as controls. This approach enabled us to distinguish patterns based on ancestral population structure (PCs) from those resulting from recent family relatedness (IBD), which are both known components of population-wide genetic variation in Bwa Mawego (Keith et al., in prep).

We ran SNP-phenotype association tests with GENESIS by first fitting null mixed models to control for phenotypic covariates (age and sex), ancestral population structure (PC1 and PC2), and family relatedness (genetic relationship matrix (GRM) of pairwise kinship coefficients). We fit linear mixed models for BMI, weight, height, mean upper arm circumference (MUAC), triceps skinfold, systolic blood pressure, and diastolic blood pressure as continuous outcomes (Table 4.1) which all included sex, age, and the first two Principal Component vectors from PC-AiR as fixed effect covariates as well as the GRM from PC-Relate as a random effect covariate. These covariates control for the impacts of age, sex, population structure, and family-based similarity on each phenotype outcome. We fit generalized linear mixed models with the binomial distribution for two binary phenotypes: known diabetic status (1/0=participant has/has

not been diagnosed as diabetic) and positive glucose test (1/0=urine glucose test had color change/had no color change) (Table 4.1). The same fixed effect (age, sex, PC1, PC2) and random effect (GRM) covariates were included in these generalized mixed models as well.

Table 4.1. Cardiometabolic phenotype statistics (n=159; 88 males, 71 females).

	Min	Q1	Median	Mean	Q3	Max
Age (yrs)	5.80	15.40	41.50	38.94	55.45	88.20
Height (in)	43.81	60.89	64.53	62.61	66.95	72.09
Weight (lbs)	44.20	111.20	137.80	136.20	164.10	323.00
BMI	14.80	19.10	22.40	23.86	27.75	55.00
MUAC (in)	6.20	9.10	10.60	10.53	12.10	16.70
Tri skinfold (mm)	3.00	5.00	7.00	10.31	14.00	35.00
Systolic Blood Pressure	90.00	113.00	126.00	130.10	145.80	178.00
Diastolic Blood Pressure	58.00	69.00	78.00	79.59	87.75	117.00
	Yes	**No**				
Glucose test	7	152				
Diabetic	9	150				

Using our initially filtered set of SNPs, we applied an LD threshold of 0.8 to exclude SNPs in high LD from subsequent association analyses (Carlson et al., 2004). 371,835 SNPs passed this LD threshold and the previous quality-control filtering, and 159 individuals (88 males and 71 and females, ages 5-88 years) are included in all models. We used the `assocTestSingle` function in `GENESIS` to run single variant association tests between the pruned set of 371,835 SNPs and the phenotype outcomes detailed above (Table 4.1). We ran score tests for all continuous outcomes and applied saddle point approximation (SPA) to the score test statistics for the binary outcomes with a p-value threshold to recalculate the statistic at 0.05 (Zhou et al., 2018). SPA reduces type I errors, which become increasingly problematic as case-control ratios become more unbalanced in binary trait outcomes.

Results

The null mixed model results contain beta coefficient estimates for the fixed effect covariates in each model (sex, age, PC1, and PC2) (Table S4.1). Height, BMI, and triceps skinfolds are significantly associated with sex such that males are generally taller, and females are more likely to have higher BMI and thicker triceps skinfolds. All modeled phenotypes positively associate with age except for triceps skinfolds. PC1 significantly associates with height, weight, and diastolic blood pressure, and PC2 significantly associates with systolic blood pressure and elevated glucose (Table S4.1). Fitting the genetic relationship matrix (GRM) of pairwise kinship coefficients as a random effect in the null models produced variance component estimates of heritability for all continuous phenotypes (Table 4.2). These point estimates range from 0.0001 (triceps skinfold) to 0.341 (systolic blood pressure), representing the proportions of observed phenotypic variation explained by shared alleles in Bwa Mawego while controlling for the impacts of sex, age, and population structure.

Table 4.2. Heritability estimates from cardiometabolic null linear mixed models. Sex, age, PC1, and PC2 were included as fixed effects in all models.

	h^2	confidence interval (95%)
BMI	0.108	(-0.218, 0.435)
Weight	0.132	(-0.170, 0.435)
Height	0.095	(-0.218, 0.408)
MUAC	0.078	(-0.204, 0.359)
Triceps Skinfold	0.0001	(-0.293, 0.294)
Systolic Blood Pressure	0.341	(-0.048, 0.731)
Diastolic Blood Pressure	0.102	(-0.238, 0.442)

Mixed model association tests calculated Rao scores to test the associations between 371,835 SNPs with each modeled phenotype (while controlling for sex, age,

population structure, and relatedness), and we report score test statistics for SNPs with the 5 smallest p-values for each trait (Table 4.3; Table S4.2). Quantile-Quantile (Q-Q) plots show the distributions of p-values (plotted with $-\log_{10}$ transformations to enhance visualization of small values) for each phenotype (Figures 4.1, S4.2-S4.9). SNPs rs10994198 and rs2393599 deviate significantly from their expected BMI score p-values and from the observed score value trend below them (Figure 4.1; Table 4.3). These two SNPs are within 1,000 base pairs of one another on Chromosome 10.

Table 4.3. BMI association score test statistics for SNPs with the 5 smallest p-values.

variant.id	chr	pos	Freq	MAC	Score	Score.SE	Score.Stat	Score.pval	Est	Est.SE	PVE
rs11761744	7	153374029	0.101	32	4.236	0.810	5.228	1.71E-07	6.452	1.234	0.177
rs10994198	10	60096277	0.286	91	6.809	1.308	5.206	1.93E-07	3.980	0.765	0.176
rs2393599	10	60097103	0.110	35	4.690	0.906	5.178	2.24E-07	5.717	1.104	0.174
rs12200377	6	132607877	0.358	113	6.068	1.293	4.692	2.71E-06	3.627	0.773	0.143
rs2487031	9	104941947	0.195	62	4.554	1.002	4.543	5.55E-06	4.532	0.998	0.134

freq=minor allele frequency, MAC=minor allele count, Score.SE=Score standard error, Score.Stat=Score z-test statistic, Est=effect size estimate per copy of minor allele, Est.SE=effect size standard error, PVE=proportion of phenotype variance explained

Manhattan plots show the distributions of score p-values (with $-\log_{10}$ transformations) by chromosome and position (Figures 4.2, S4.10-S4.17). We applied a Bonferroni correction at the 0.05 significance level for 371,835 SNP tests to obtain a p-value threshold of 1.34×10^{-7}, recognizing that this cutoff is conservative given weak to moderate levels of LD between many non-independent SNPs included below our 0.8 LD threshold. The BMI Manhattan plot shows 3 SNPs (2 overlapping) just below the 0.05 p-value threshold with noticeably smaller values than the rest (Figure 4.2; Table 4.3). These 3 high points on the Manhattan plot include rs10994198 and rs2393599 in close proximity on Chromosome 10 which have observed score p-values significantly beyond the expected distribution (Figure 4.1).

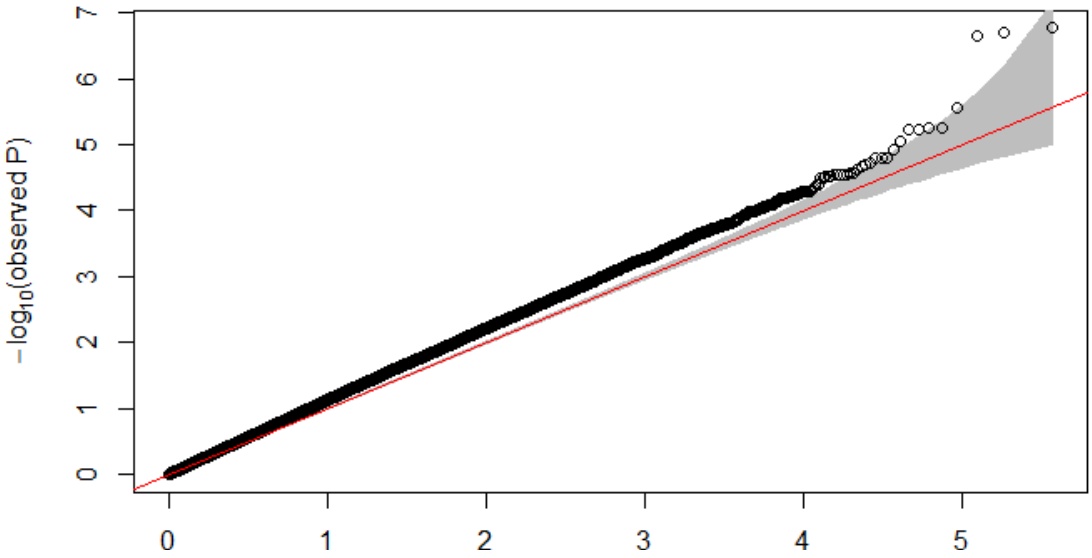

Figure 4.1. BMI Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

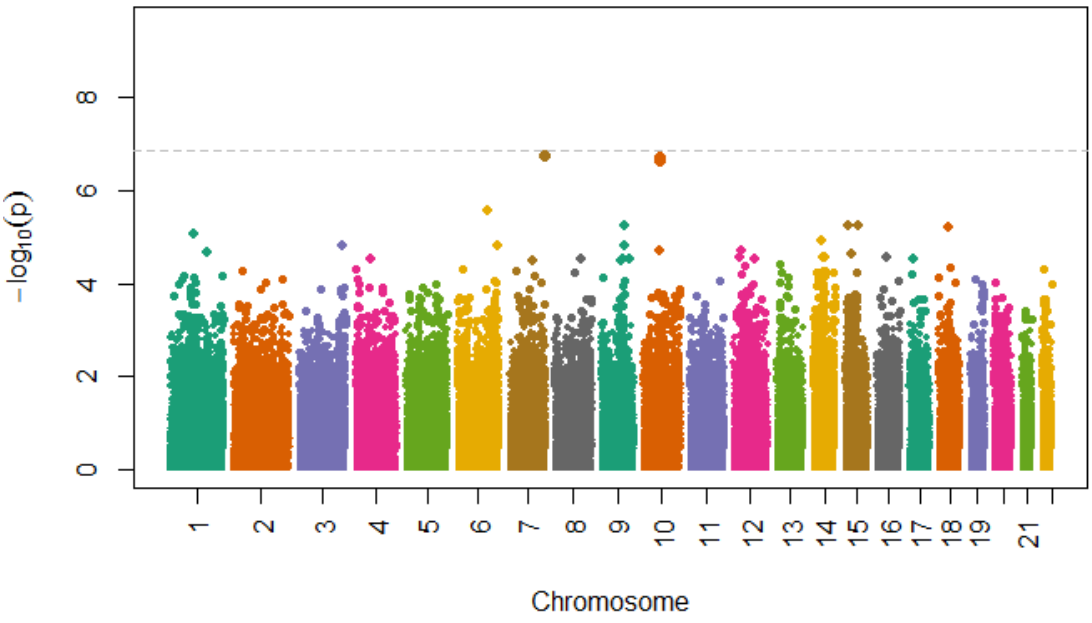

Figure 4.2. BMI Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

Discussion

Our null mixed model results indicate that many cardiometabolic phenotypes in Bwa Mawego vary with population structure shown via PC1 and PC2 beta coefficients (Table S2). This Caribbean horticultural community has minimum estimates of 40% African ancestry, 20% European ancestry, and 6% Amerindian ancestry that contribute to its genetic structure (Keith et al., in prep), and body composition and cardiovascular traits are known to vary by ancestry both phenotypically and genetically, particularly in Latin and Caribbean populations (Noel et al., 2017; Muñoz et al. 2016). These associations further highlight the need to account for population structure and ancestry when testing genotype-phenotype associations.

Heritability estimates from null linear mixed models suggest that kinship coefficients explain relatively low amounts of phenotypic variation observed in these cardiometabolic traits (Table 4.2). All confidence intervals span zero, and heritability point estimates for cross-sectional height (0.095), weight (0.132), and BMI (0.108) are substantially lower than pedigree-derived estimates for the same population calculated previously in a Bayesian framework from longitudinal data (h^2_{height}=0.683, h^2_{weight}=0.640, h^2_{BMI}=0.487) (Keith et al., 2019). Bayesian longitudinal estimates included sex, age, and data collection year as covariates but had no measure of ancestry or population structure; one potential explanation for some discrepancy between estimates is inflation in previously calculated heritabilities due to unmodeled ancestry or population structure.

SNPs rs10994198 and rs2393599 that appear to associate with BMI variation in Bwa Mawego (Figures 4.1-4.2) are located on chromosome 10 at positions 60096277 and

60097103, respectively (Table 4.3). These intron variants are downstream of the protein-coding *ANK3* gene, both overlap eight transcripts, and rs2393599 overlaps a regulatory enhancer for *ANK3* (Yates et al., 2020). rs2393599 has global allele frequencies of 0.88 (G) and 0.12 (A), and frequencies of 0.81 (G) and 0.19 (A) in African populations, 0.93 (G) and 0.07 (A) in European populations, 0.95 (G) and 0.05 (A) in Amerindian populations, and 0.89 (G) and 0.11 (A) in Bwa Mawego, Dominica (Yates et al., 2020; Table 3).

The *ANK3* gene is located at 10q21.2 (Chromosome 10: 60026298-60733490) and encodes the Ankyrin-G protein involved in sodium-ion channel function and neuronal development and stability (Kordeli et al., 1995; Rasband et al., 1999; Yates et al., 2020). Initially identified in nodes of Ranvier, there are multiple known isoforms of Ankyrin-G that result from alternative splicing and are expressed differentially in a variety of tissues including brain, skeletal muscle, kidney, and heart (Peters et al., 1995; Yamankurt et al., 2015). *ANK3* has reported associations with bipolar disorder, schizophrenia, BMI, height, bone mineral density, blood pressure, and a variety of neurological and psychiatric conditions (Cook et al., 2018). *ANK3* has a particularly robust association with bipolar disorder, and the consistently significant SNPs for this phenotype are also in regulatory regions (MacKenzie et al., 2013; Rueckert et al., 2013). *ANK3* has been identified as an informative marker of early-life stress as a gene whose expression and methylation patterns are consistently modified by perinatal stress, establishing epigenetic evidence for regulatory importance of *ANK3* (Luoni et al., 2016).

In addition to our SNP-BMI findings, 3 large European studies have detected significant associations between BMI and *ANK3* variants (Kichaev et al., 2018; Pulit et

al., 2019; Zhu et al., 2019), but any functional relationships between *ANK3*/Ankyrin-G and cardiometabolic traits are currently unknown. In addition to the direct association between *ANK3* and bipolar disorder, there is a well-established phenotypic association between bipolar disorder and BMI (McElroy and Keck, 2012; Sicras et al., 2012). SNP rs12772424, an intron of the TCF7L2 transcription factor on chromosome 10, has been functionally identified as a BMI-dependent mediator of psychiatric risk as well as a strong genetic risk variant for type II diabetes (Winham et al., 2014; Cuellar-Barboza et al., 2016). These associations suggest that regulatory variants may reveal functional pleiotropic relationships between health phenotypes, and additional research is needed to further investigate regulation, expression, and epigenetic modification of the *ANK3* region.

The smooth and uniform observed p-value trends in our Q-Q plots indicate that the mixed association models adequately accommodate covariates for continuous phenotypes, but the null expected p-value distribution tracks less well for height and weight, perhaps due to the influence of additional un-modeled covariates (Figures 4.1, S4.2-S4.9). Unbalanced case-control ratios present computational challenges in this mixed association model framework for binary phenotypes, and saddle point approximation does not correct for these adequately in our related sample of 159 (Figures S4.10-S4.17) (Zhou et al., 2018). More precise, continuous glucose phenotypes are needed to investigate genetic components of diabetic outcomes in Bwa Mawego in more detail. However, we are able to clearly visualize potential associations for continuous traits with this small, localized sample of mixed ancestry when population structure and relatedness are accommodated (Figures 4.1-4.2).

As genetic data increasingly inform our understandings of human health, more inclusive research is needed in populations of non-European and mixed ancestries (Pulit et al., 2010; Petrovski and Goldstein, 2016). We have identified a potential genetic region of interest associated with BMI variation in a horticultural Caribbean community that may indicate a role of *ANK3* regulation in metabolic phenotypes. Regulatory variants are increasingly recognized for their functional roles in GWAS, also highlighting the importance of characterizing epigenetic variation and other influences on gene expression (Gallagher and Chen-Plotkin, 2018). Given the context-specific properties of environmental and inter-generational impacts on epigenomic variation (Carja et al., 2017), there is unique potential to explore complex gene-environment-phenotype landscapes in Bwa Mawego with its small-scale community structure and longitudinal research engagement (Keith et al., 2019). Admixture mapping analyses in this Caribbean population may also yield additional association insights given their increased statistical power over traditional GWAS and potential to detect rare variants (Qin et al., 2019).

Acknowledgements

We thank the community of Bwa Mawego for facilitating this research and Nethilia Durand and Nakisha Wiltshire for their data collection assistance. Many thanks to each of the research participants and to Kristen Taylor and Alex Stuckel for providing lab space and equipment for DNA extractions and sample processing. This research was funded by the University of Missouri Research Board.

References

Belbin, G.M., Nieves-Colón, M.A., Kenny, E.E., Moreno-Estrada, A., & Gignoux, C.R. (2018). Genetic diversity in populations across Latin America: implications for population and medical genetic studies. *Current Opinion in Genetics & Development*, 53, 98–104. https://doi.org/10.1016/j.gde.2018.07.006

Benn-Torres, J., Stone, A.C., & Kittles, R. (2013). An anthropological genetic perspective on creolization in the anglophone Caribbean. *American Journal of Physical Anthropology*, 151(1), 135–143. https://doi.org/10.1002/ajpa.22261

Boca Scientific. Buccal Swabs/DNA Extraction. www.bocascientific.com. Accessed February 2020.

Bustamante, C.D., De La Vega, F.M., & Burchard, E.G. (2011). Genomics for the world. *Nature*, 475(7355), 163–165. https://doi.org/10.1038/475163a

Carja, O., MacIsaac, J.L., Mah, S.M., Henn, B.M., Kobor, M.S., Feldman, M.W., & Fraser, H.B. (2017). Worldwide patterns of human epigenetic variation. *Nature Ecology & Evolution*, 1(10), 1577–1583. https://doi.org/10.1038/s41559-017-0299-z

Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L., & Nickerson, D.A. (2004). Selecting a Maximally Informative Set of Single-Nucleotide Polymorphisms for Association Analyses Using Linkage Disequilibrium. *The American Journal of Human Genetics*, 74(1), 106–120. https://doi.org/10.1086/381000

Cook, C.E., Bergman, M.T., Cochrane, G., Apweiler, R., & Birney, E. (2018). The European Bioinformatics Institute in 2017: data coordination and integration. *Nucleic Acids Research*, 46 (Database issue), D21–D29. https://doi.org/10.1093/nar/gkx1154

Cooper, R.S., Tayo, B., & Zhu, X. (2008). Genome-wide association studies: implications for multiethnic samples. *Human Molecular Genetics*, 17(R2), R151–R155. https://doi.org/10.1093/hmg/ddn263

Cuellar-Barboza, A.B., Winham, S.J., McElroy, S.L., Geske, J.R., Jenkins, G. D., Colby, C.L., & Biernacka, J.M. (2016). Accumulating evidence for a role of TCF7L2 variants in bipolar disorder with elevated body mass index. *Bipolar Disorders*, 18(2), 124–135. https://doi.org/10.1111/bdi.12368

Flinn, M.V., Leone, D.V., & Quinlan, R.J. (1999). Growth and Fluctuating Asymmetry of Stepchildren. *Evolution and Human Behavior*, 20(6), 465-479.

Gallagher, M.D., & Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. *American Journal of Human Genetics*, 102(5), 717–730. https://doi.org/10.1016/j.ajhg.2018.04.002

Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., & Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*. doi:10.1093/bioinformatics/btz567.

Illumina. (2018). OmniExpress-24 v1.3 BeadChip Datasheet. www.illumina.com. Accessed February 2020.

Keith, M.H., Durbin, H.J., Rowan, T.N., Blomquist, G.E., Flinn, M.V., Taylor, J.F., & Decker, J.E. (in prep). Genetic ancestry and population structure of a rural Caribbean community.

Keith, M.H., Blomquist, G.E., & Flinn, M.V. (2019). Anthropometric heritability and child growth in a Caribbean village: A quantitative genetic analysis of longitudinal height, weight, and body mass index in Bwa Mawego, Dominica. *American Journal of Physical Anthropology*, 170(3), 393–403. https://doi.org/10.1002/ajpa.23924

Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., & Price, A.L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *American Journal of Human Genetics*, 104(1), 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008

Kordeli, E., Lambert, S., & Bennett, V. (1995). Ankyrin G: A new Ankyrin gene with neural-specific isoforms localized at the axonal initial segment and node of Ranvier. *Journal of Biological Chemistry*, 270(5), 2352–2359. https://doi.org/10.1074/jbc.270.5.2352

Kulminski, A.M., Loika, Y., Culminskaya, I., Arbeev, K.G., Ukraintseva, S.V., Stallard, E., & Yashin, A.I. (2016). Explicating heterogeneity of complex traits has strong potential for improving GWAS efficiency. *Scientific Reports*, 6(1), 35390. https://doi.org/10.1038/srep35390

Luoni, A., Massart, R., Nieratschker, V., Nemoda, Z., Blasi, G., Gilles, M., & Riva, M.A. (2016). Ankyrin-3 as a molecular marker of early-life stress and vulnerability to psychiatric disorders. *Translational Psychiatry*, 6(11), e943–e943. https://doi.org/10.1038/tp.2016.211

Macfarlan, S.J., Remiker, M., & Quinlan, R. (2012). Competitive Altruism Explains Labor Exchange Variation in a Dominican Community. *Current Anthropology*, 53(1), 118–124. https://doi.org/10.1086/663700

MacKenzie, A., Hing, B., & Davidson, S. (2013). Exploring the effects of polymorphisms on cis-regulatory signal transduction response. *Trends in Molecular Medicine*, 19(2), 99–107. https://doi.org/10.1016/j.molmed.2012.11.003

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., & Visscher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753. https://doi.org/10.1038/nature08494

Marees, A.T., Kluiver, H. de, Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, 27(2), e1608. https://doi.org/10.1002/mpr.1608

McElroy, S.L., & Keck, P.E. (2012). Obesity in Bipolar Disorder: An Overview. *Current Psychiatry Reports*, 14(6), 650–658. https://doi.org/10.1007/s11920-012-0313-8

Muñoz, A.M., Velásquez, C.M., & Bedoya, G. (2016). Cardio-metabolic parameters are associated with genetic admixture estimates in a pediatric population from Colombia. *BMC Genetics*, 17. https://doi.org/10.1186/s12863-016-0402-5

Noel, S.E., Arevalo, S., Smith, C E., Lai, C.Q., Dawson-Hughes, B., Ordovas, J.M., & Tucker, K.L. (2017). Genetic admixture and body composition in Puerto Rican adults from the Boston Puerto Rican Osteoporosis Study. *Journal of Bone and Mineral Metabolism*, 35(4), 448–455. https://doi.org/10.1007/s00774-016-0775-6

Peters, L.L., John, K.M., Lu, F.M., Eicher, E.M., Higgins, A., & Yialamas, M. (1995). Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene family and the major ankyrin in kidney, is expressed in alternatively spliced forms, including forms that lack the repeat domain. *Journal of Cellular Biology*, 130 (2), pp. 313-330.

Petrovski, S., & Goldstein, D.B. (2016). Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biology*, 17. https://doi.org/10.1186/s13059-016-1016-y

Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., & Lindgren, C. M. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics*, 28(1), 166–174. https://doi.org/10.1093/hmg/ddy327

Pulit, S.L., Voight, B.F., & de Bakker, P.I.W. (2010). Multiethnic Genetic Association Studies Improve Power for Locus Discovery. *PLoS ONE*, 5(9). https://doi.org/10.1371/journal.pone.0012600

Qiagen. Products: DNA Clean Up. www.qiagen.com. Accessed February 2020.

Qin, H., Zhao, J., & Zhu, X. (2019). Identifying Rare Variant Associations in Admixed Populations. *Scientific Reports*, 9(1), 5458. https://doi.org/10.1038/s41598-019-41845-3

Quinlan, M. (2004). From the Bush: The Front Line of Health Care in a Caribbean Village. Wadsworth: Belmont, CA.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rasband, M.N., Peles, E., Trimmer, J.S., Levinson, S.R., Lux, S.E., & Shrager, P. (1999). Dependence of nodal sodium channel clustering on paranodal axoglial contact in the developing CNS. *Journal of Neuroscience*, 19 (17), pp. 7516-7528

Rueckert, E.H., Barker, D., Ruderfer, D., Bergen, S.E., O'Dushlaine, C., Luce, C.J., & Sklar, P. (2013). Cis-Acting Regulation of Brain-Specific ANK3 Gene Expression by a Genetic Variant Associated with Bipolar Disorder. *Molecular Psychiatry,* 18(8), 922–929. https://doi.org/10.1038/mp.2012.104

Sicras, A., Rejas, J., Navarro, R., Serrat, J., & Blanca, M. (2008). Metabolic syndrome in bipolar disorder: a cross-sectional assessment of a Health Management Organization database. *Bipolar Disorders*, 10(5), 607–616. https://doi.org/10.1111/j.1399-5618.2008.00599.x

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 101(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Winham, S.J., Cuellar-Barboza, A.B., Oliveros, A., McElroy, S.L., Crow, S., Colby, C., & Biernacka, J.M. (2014). Genome-wide association study of bipolar disorder accounting for effect of body mass index identifies a new risk allele in TCF7L2. *Molecular Psychiatry*, 19(9), 1010–1016. https://doi.org/10.1038/mp.2013.159

Yamankurt, G., Wu, H.C., McCarthy, M., & Cunha, S.R. (2015). Exon Organization and Novel Alternative Splicing of Ank3 in Mouse Heart. *PLoS ONE*, 10(5). https://doi.org/10.1371/journal.pone.0128177

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., & Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, 48(D1), D682–D688. https://doi.org/10.1093/nar/gkz966

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., & Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., & Lee, S. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, 50(9), 1335–1341. https://doi.org/10.1038/s41588-018-0184-y

Zhu, Z., Guo, Y., Shi, H., Liu, C.L., Panganiban, R.A., Chung, W., & Liang, L. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *The Journal of Allergy and Clinical Immunology*, 145(2), 537–549. https://doi.org/10.1016/j.jaci.2019.09.035

CHAPTER FIVE

Summary of Key Findings

The horticultural community of Bwa Mawego, Dominica has unique patterns of genomic variation that influence health and biology, and the localized scale of this population presents opportunities to elucidate complex interactions between genetic, environmental, and phenotypic variation. This three-part project used a combination of longitudinal growth, pedigree, genome-wide single nucleotide polymorphism (SNP), and cardiometabolic data to assess sources of variation in genomic patterns as well as anthropometric and cardiometabolic traits. Population genetic analyses clarify unique historical events over the past few centuries in this region of the Caribbean, and we have identified significant genetic components of phenotypic variation in Bwa Mawego, most specifically for body mass index (BMI).

Bayesian analyses of longitudinal growth data from 1997-2017 show that BMI varies widely between individual growth curves but is fairly consistent (repeatability=0.77) within individuals over time. Secular trends across this 20-year timespan do not show consistently increasing (or decreasing) BMI population-wide, and pedigree-derived heritability estimates indicate that up to 49% of the observed variation in BMI is explained by genetic variance in Bwa Mawego.

There is strong genetic structure in this admixed community, and four family-based clusters indicate that there are several well-defined family lineages in which to assess genetic associations. Admixture appears to have been localized between 5-6 generations (~130-150 years) ago in Bwa Mawego, and minimum estimates calculated in reference to the Human Genome Diversity Panel (HGDP) indicate that at least 40% of the

community's genetic ancestry is African, at least 20% is European, and more than 6% of the tagged SNP variation in Bwa Mawego correlates with Amerindian populations sampled in the HGDP.

Genomic patterns in Bwa Mawego correlate closely with samples from West African, French, and Karitiana/Surui populations. West African and French haplotypes reflect the island's 18th century French (and British) colonial occupation. Dominica has significantly more Amerindian ancestry than other Caribbean populations throughout the English-speaking Caribbean, in part reflecting its role as a destination for indigenous people fleeing colonial violence and seeking refuge in the island's steep, forested terrain. Karitiana and Surui groups are genetically distinct Amerindian peoples with globally-rare, deeply-diverged lineages shared with some indigenous groups in Australia and Oceania, and the affinity of haplotypes in Bwa Mawego to these samples indicates that the admixture composition in Dominica uniquely represents diverse global variation.

Ancestral genomic structure resulting from recent admixture is a source of phenotypic variation for height, weight, diastolic blood pressure, systolic blood pressure, and elevated glucose levels in Bwa Mawego. Genome-wide association (GWA) tests between 371,835 autosomal SNPs and cardiometabolic phenotypes revealed associations between 2 SNPs on chromosome 10 (rs10994198 and rs2393599) and BMI. These SNPs are within 1,000 base pairs of one another, downstream of the *ANK3* gene, and rs2393599 overlaps a regulatory enhancer. Any functional relationship between *ANK3* and body mass remains to be verified and characterized, and known epigenetic evidence shows that *ANK3* expression associates with stress and associated physiological stress responses

early in life, indicating that further exploration of *ANK3*'s role across genetic-environmental-phenotypic interactions is warranted.

Additional analyses of SNP data from this study could reveal unique signatures of selection across the admixed haplotypes sampled in Bwa Mawego to address the impact of evolutionary forces on patterns of genomic variation. In the future, assigning local ancestry across this SNP dataset will facilitate admixture mapping to further explore genotype-phenotype associations, which may uncover additional insights given the increased statistical power and potential to detect rare variants. More data is needed to capture environmental sources of variation on anthropometric and cardiometabolic outcomes in Bwa Mawego. In addition to characterizing more detailed cultural and biocultural sources of variation in health outcomes, this would facilitate epigenetic analyses of differential gene expression and inter-generational trends in this rural Caribbean community.

SUPPLEMENTAL MATERIALS

Figures

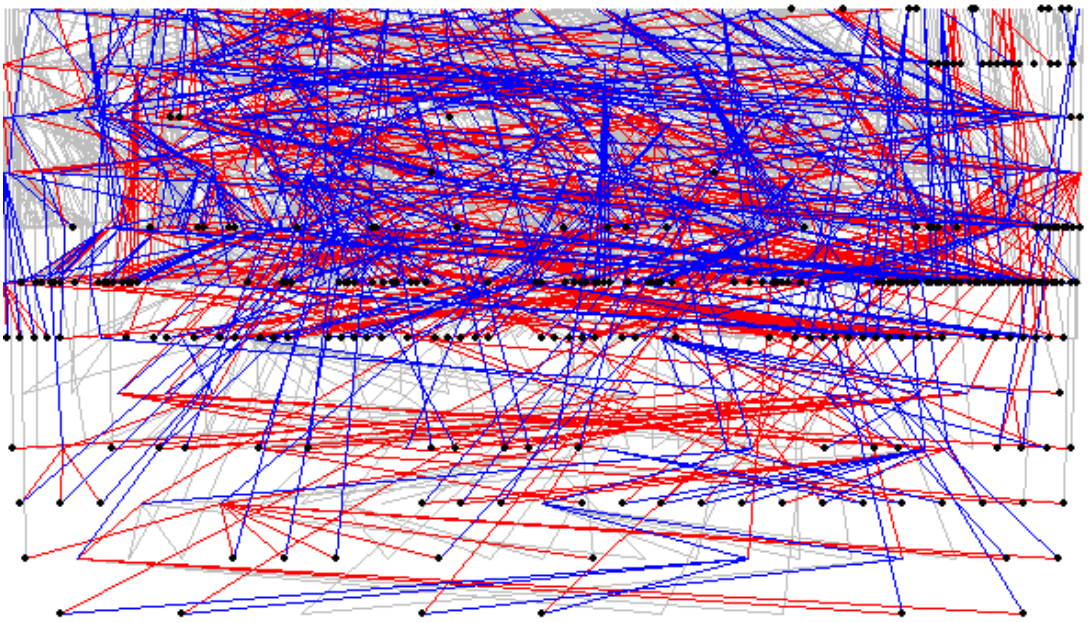

Figure S2.1. 11-generation pedigree of Bwa Mawego, Dominica. Lines indicate ancestor-descendant relationships. This complete population pedigree includes 1,455 individuals, and dots mark the 662 individuals who are related to and/or included in the 260 individuals for whom we have longitudinal growth data from 1997-2017.

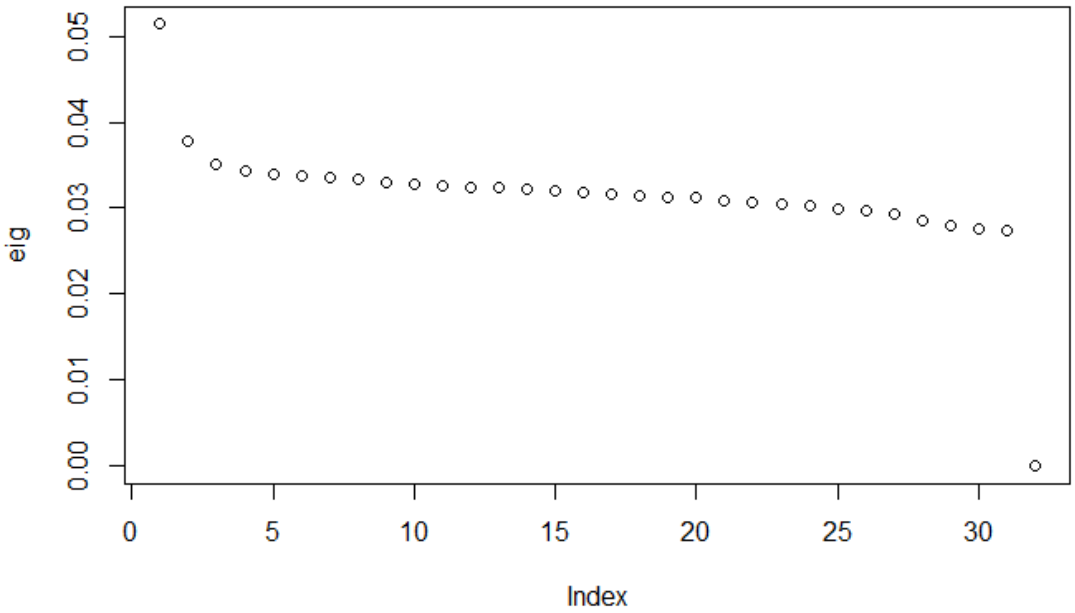

Figure 4.1. Scree plot of eigenvalues for 32 Principal Components derived from 31,968 independent SNPs.

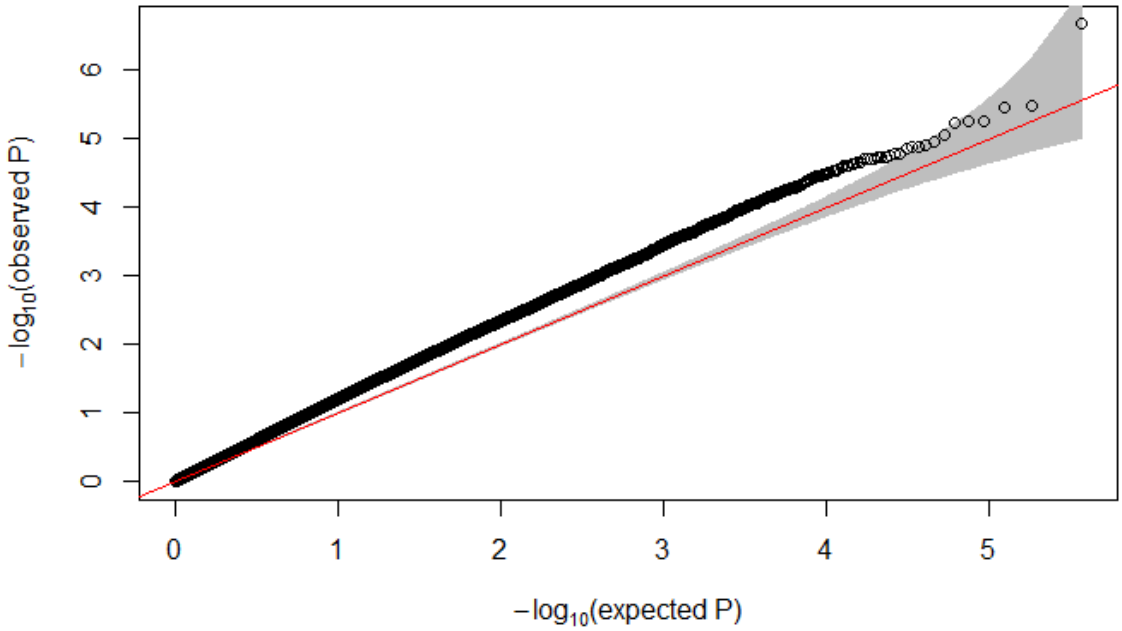

Figure S4.2. Weight Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

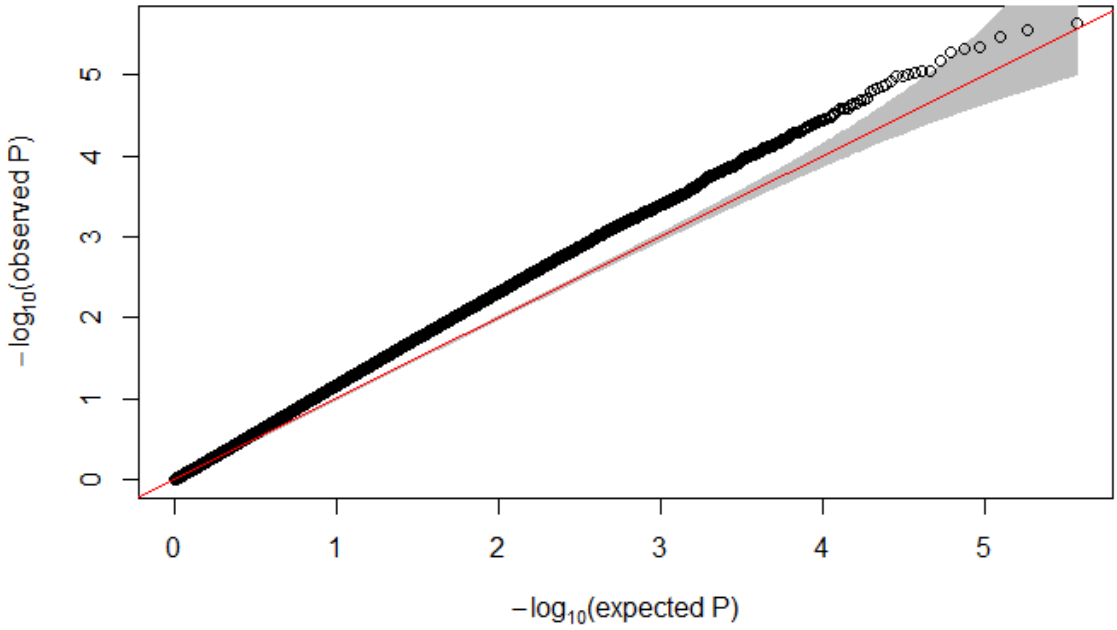

Figure S4.3. Height Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

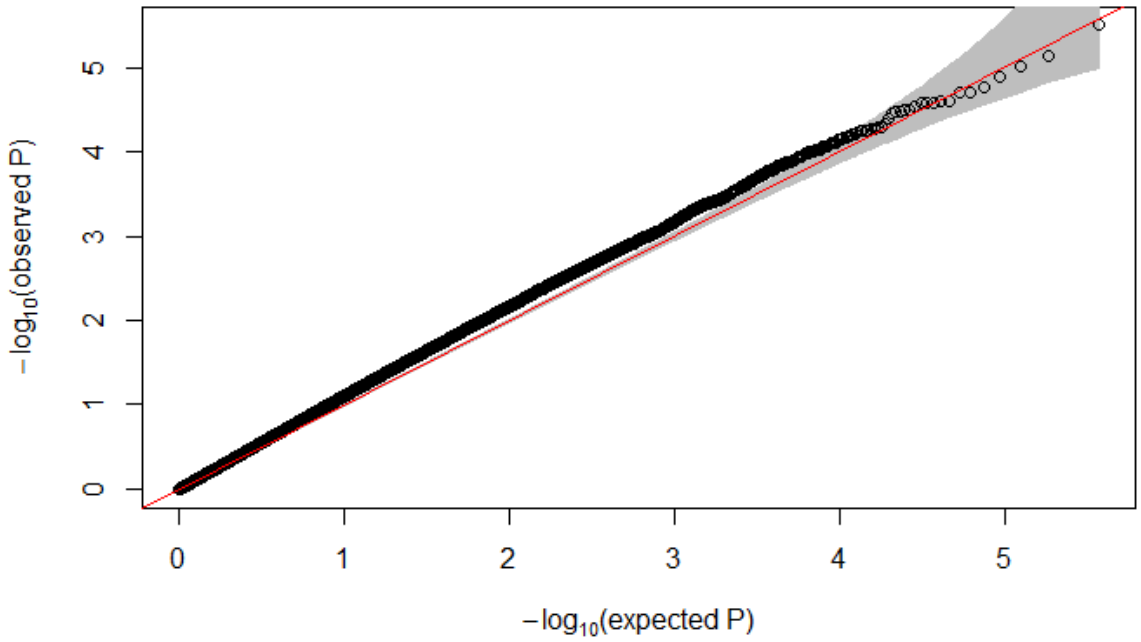

Figure S4.4. MUAC Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

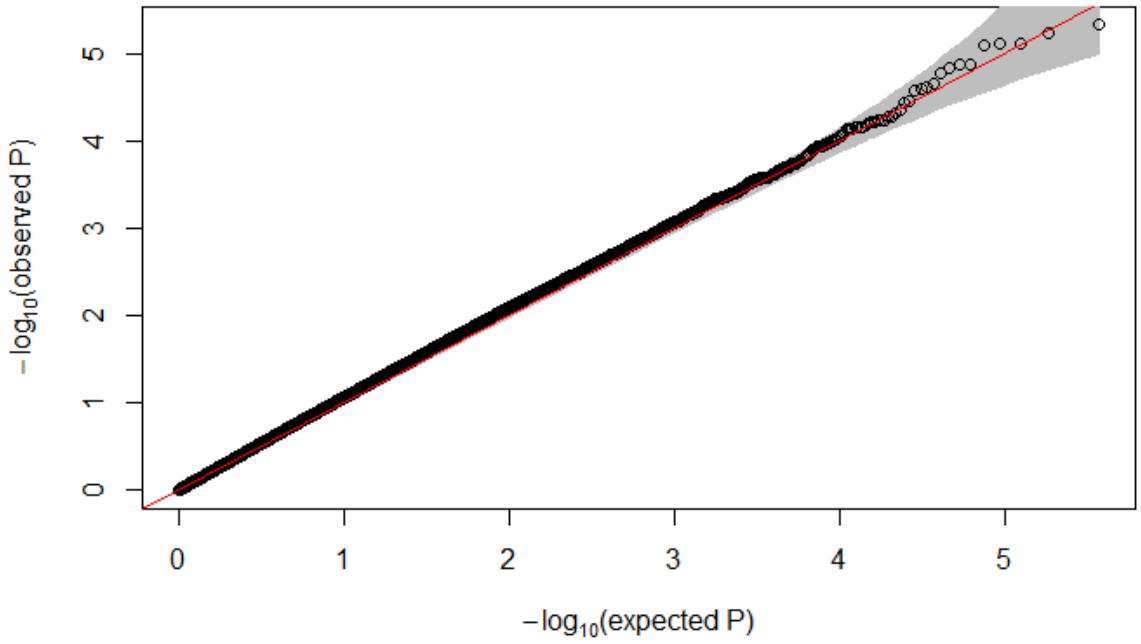

Figure S4.5. Triceps Skinfold Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

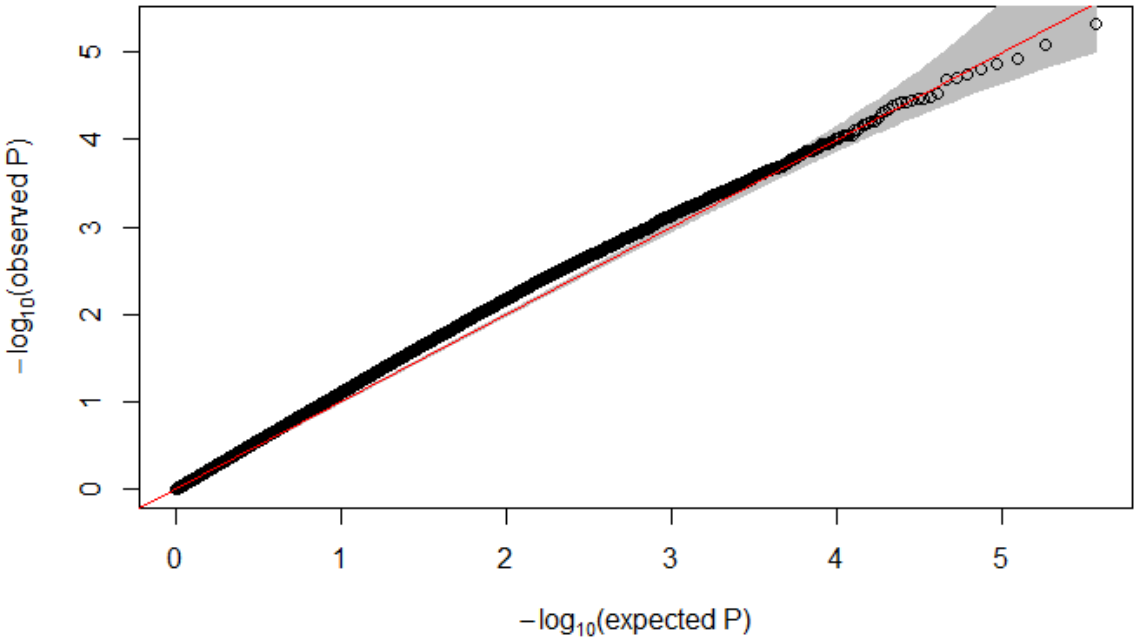

Figure S4.6. Systolic Blood Pressure Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

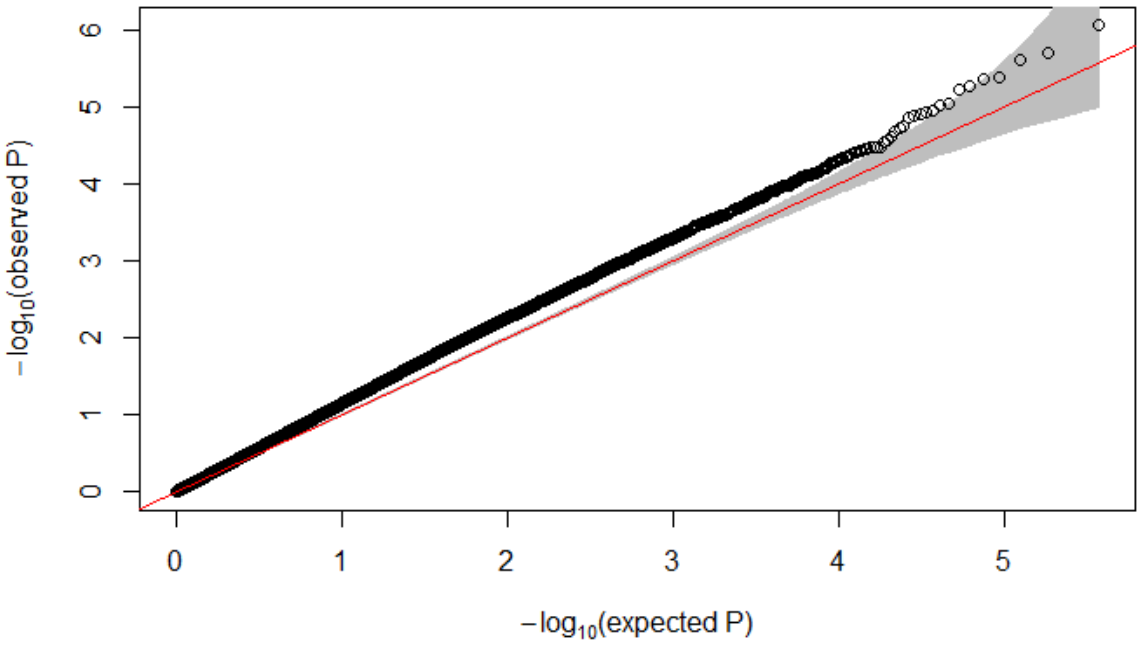

Figure S4.7. Diastolic Blood Pressure Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

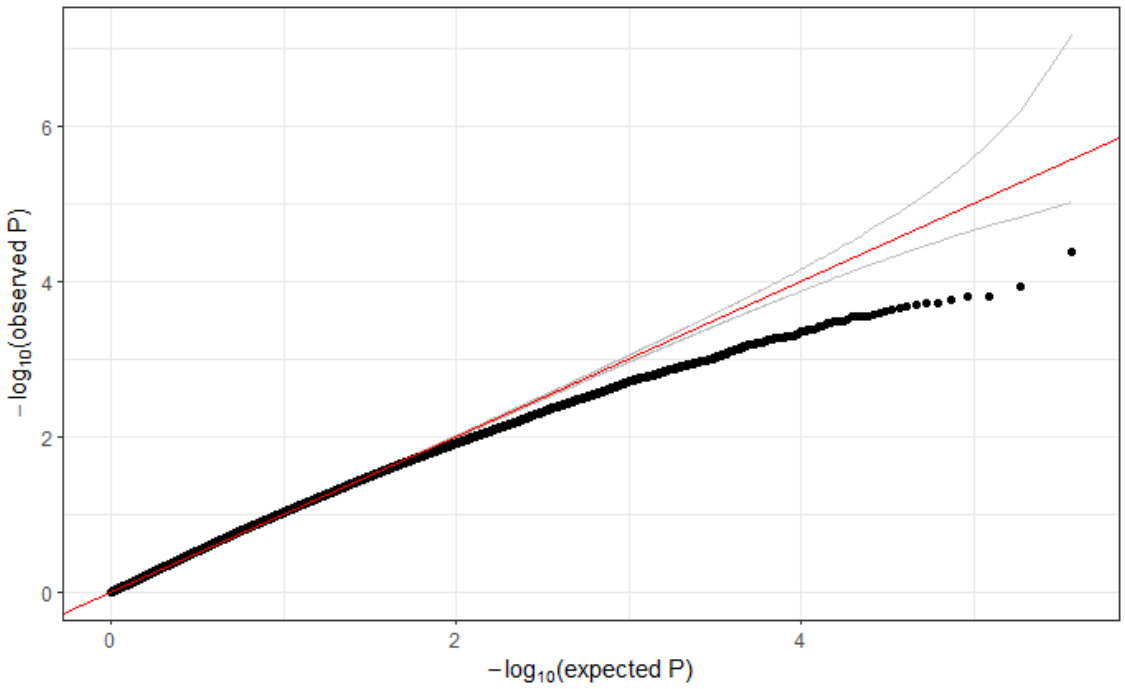

Figure S4.8. Diabetic Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

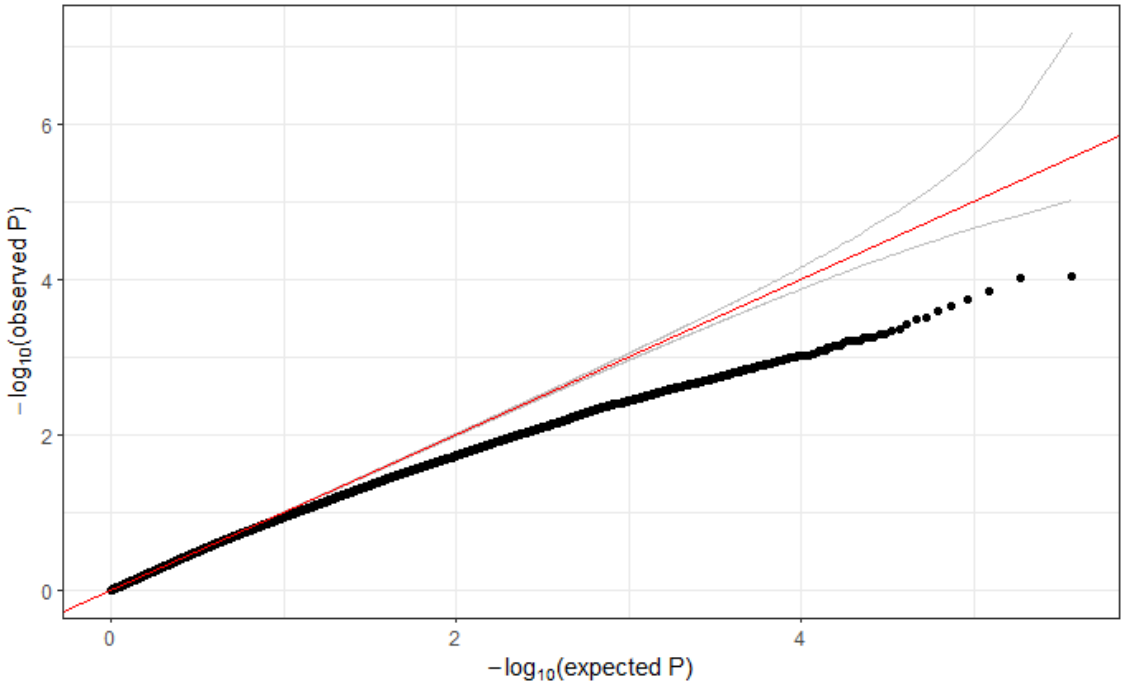

Figure S4.9. Glucose Test Q-Q Plot of p-values for 371,835 SNPs. Expected p-values are plotted in red with the 95% confidence interval in gray.

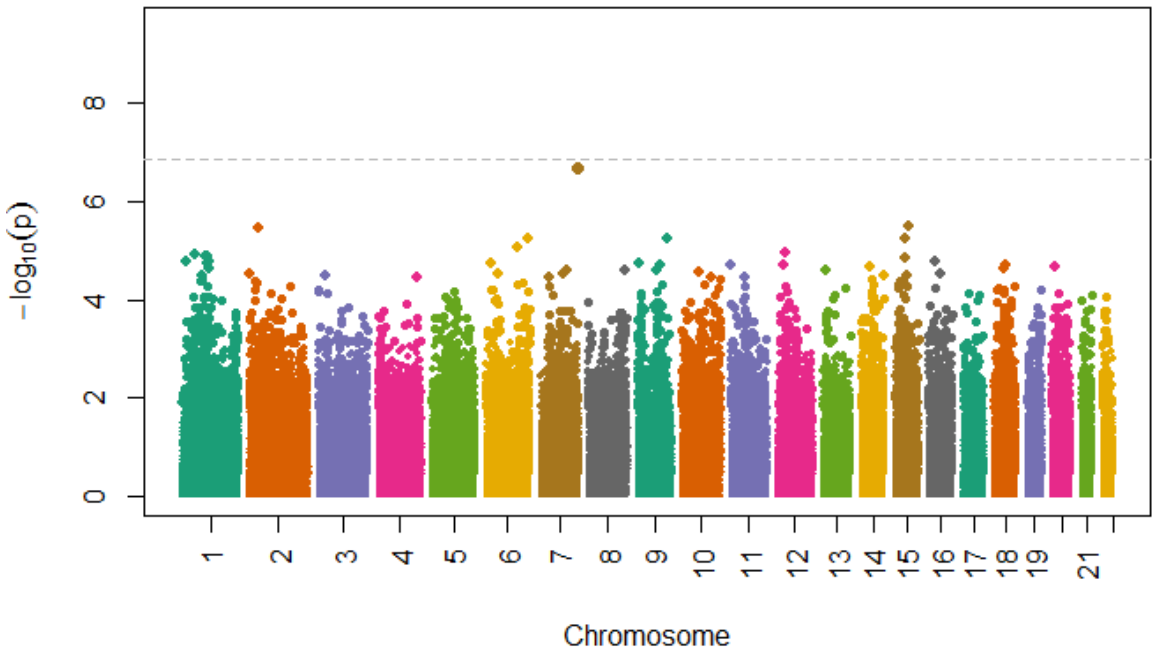

Figure S4.10. Weight Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

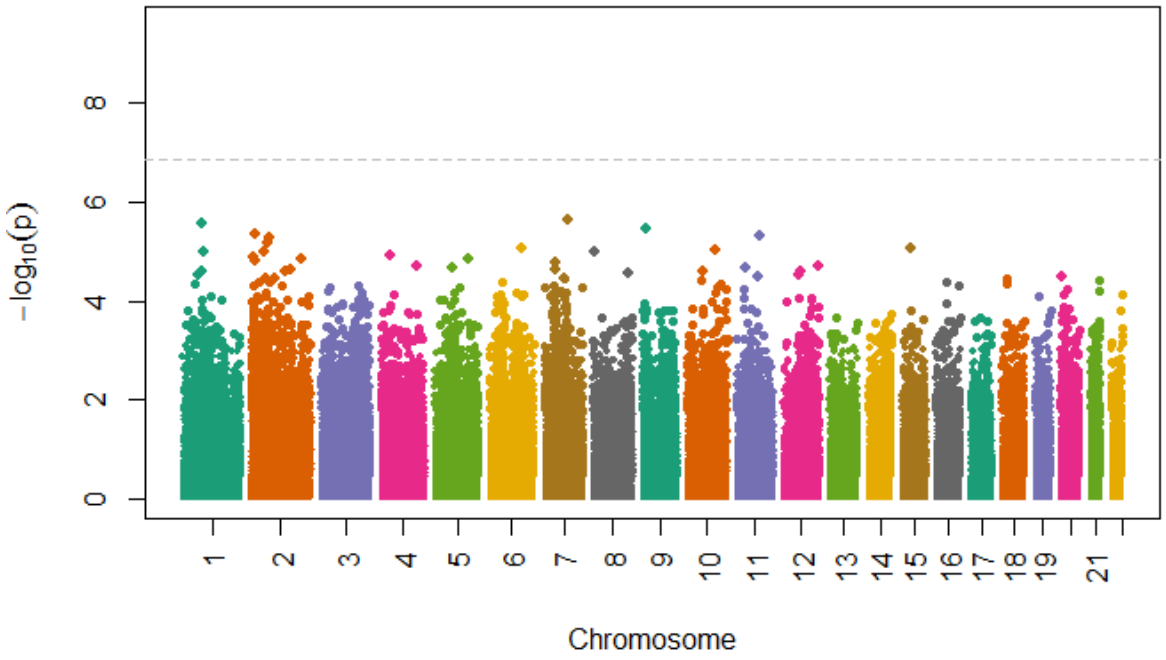

Figure S4.11. Height Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

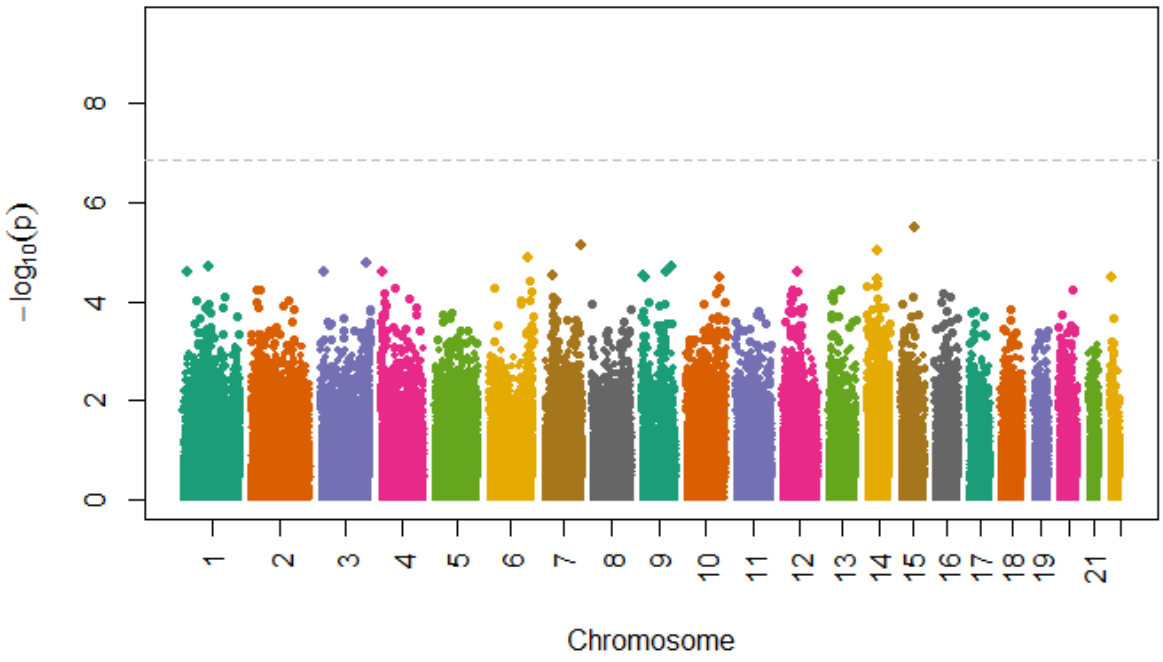

Figure S4.12. MUAC Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

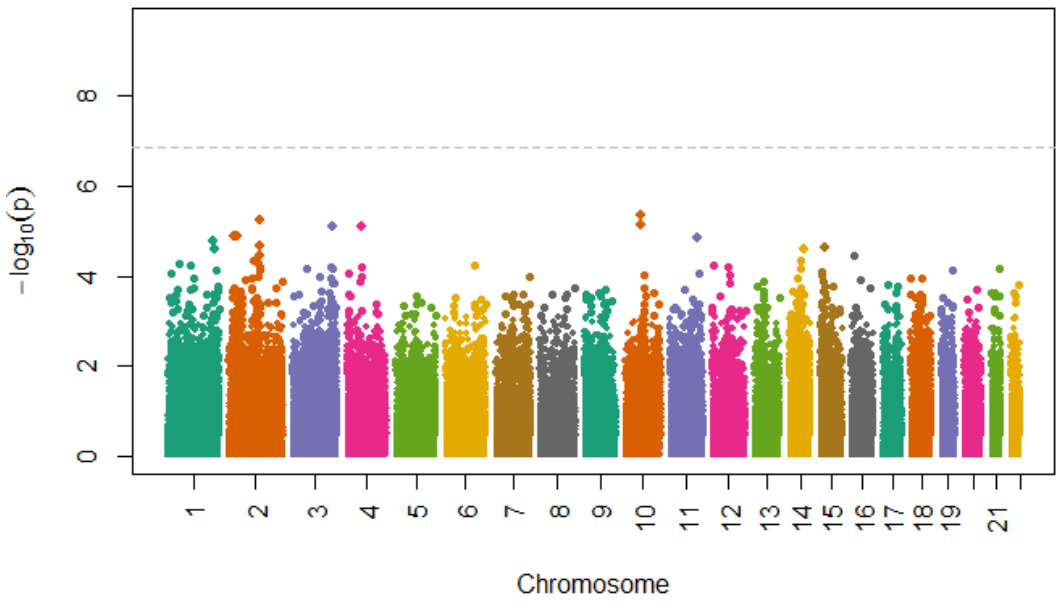

Figure S4.13. Tricep Skinfold Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

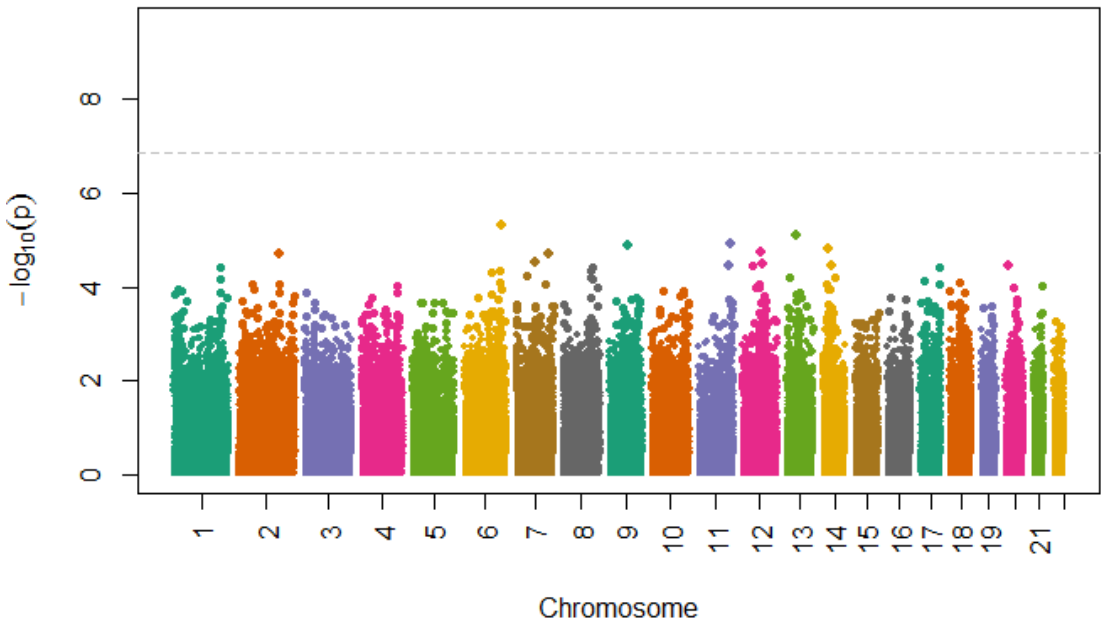

Figure S4.14. Systolic Blood Pressure Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

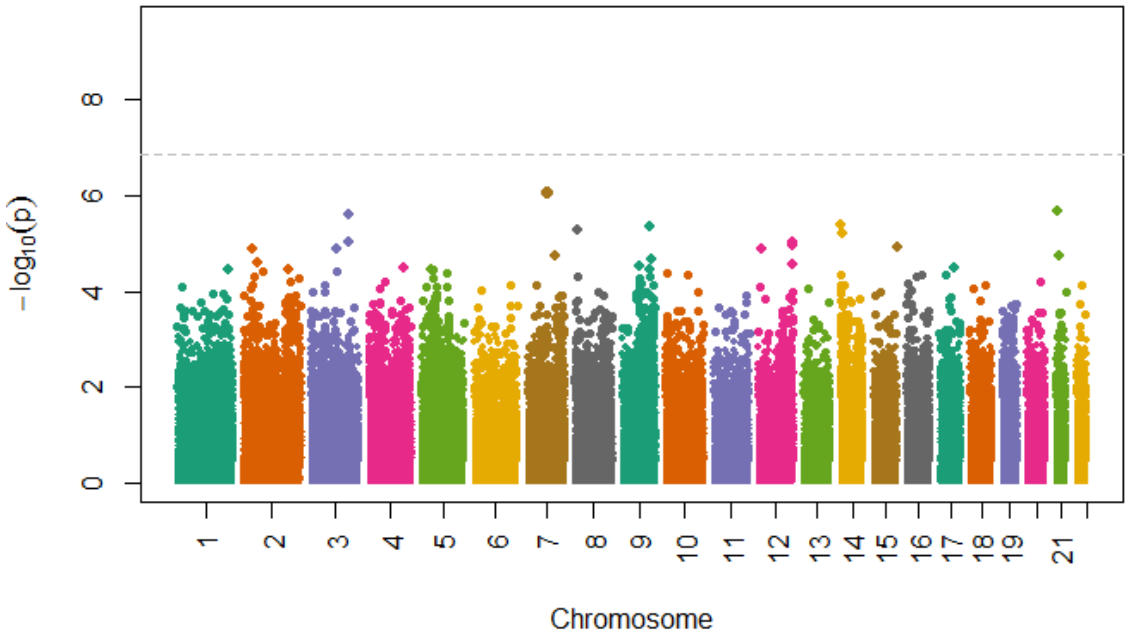

Figure S4.15. Diastolic Blood Pressure Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

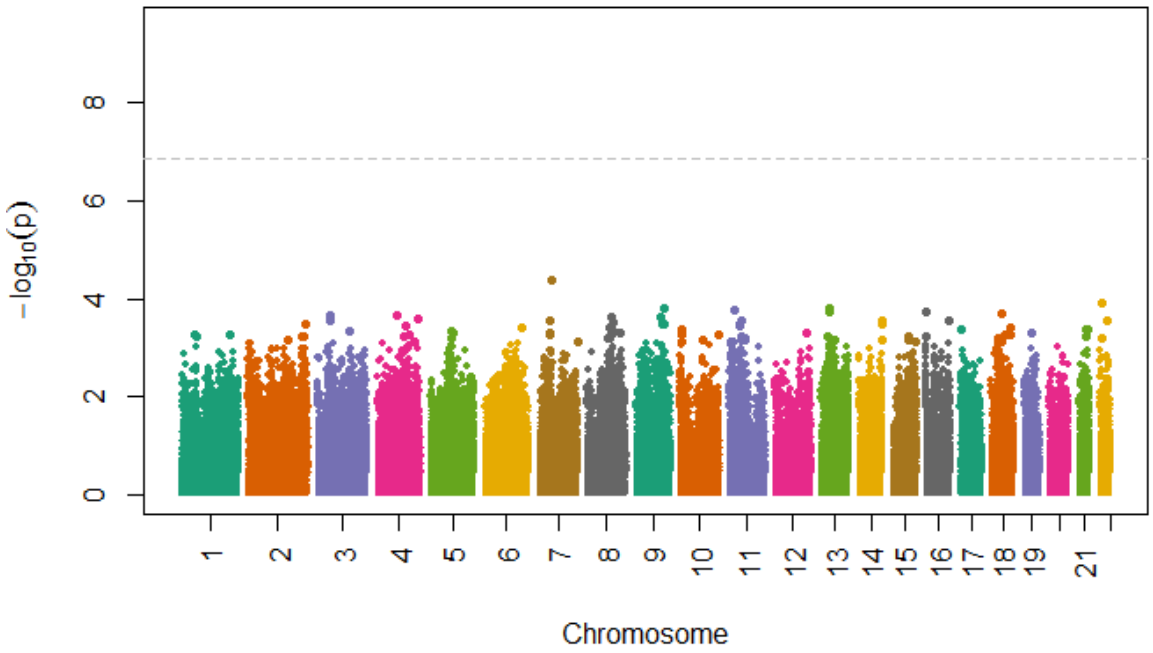

Figure S4.16. Diabetic Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

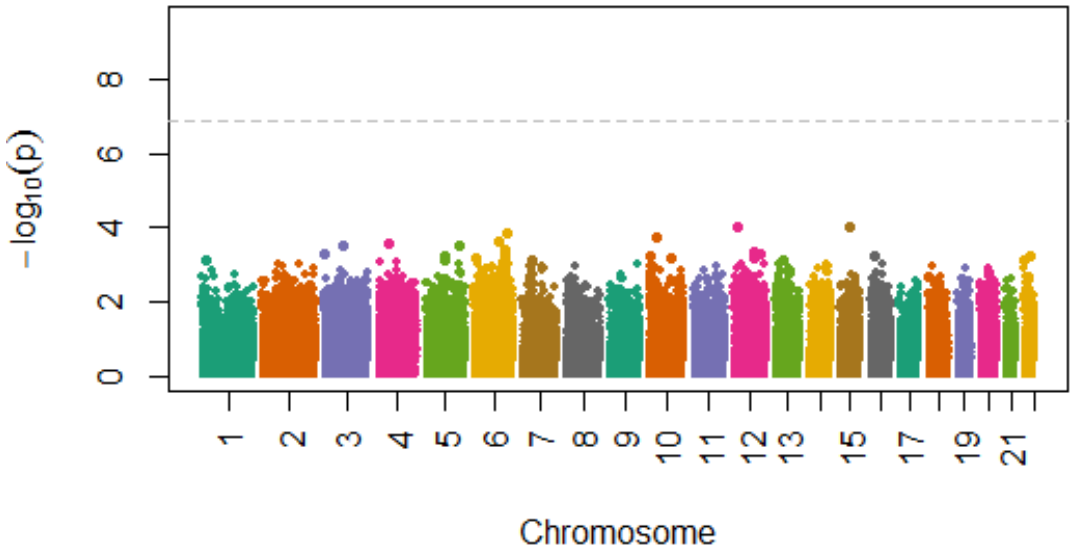

Figure S4.17. Glucose Test Manhattan plot for 371,835 autosomal SNPs. Dashed line marks the p-value threshold of 1.34×10^{-7} (Bonferroni correction at 0.05 significance).

Tables

Table S2.1. Population-wide pedigree statistics.

Individuals	1455	Maternal grandmothers	879
Maternities	1097	Maternal grandfathers	842
Paternities	1062	Paternal grandmothers	696
Full sibs	1487	Paternal grandfathers	727
Maternal sibs	1992	Maximum pedigree depth	11
Maternal half sibs	505	Founders	328
Paternal sibs	1733	Non-zero F	114
Paternal half sibs	246	F > 0.125	1
Mean pairwise relatedness	0.0085		

Parental, grand-parental, and sibling relationship counts include all relationships among individuals, therefore one female may have multiple maternities, sibships, etc. Values of F denote inbreeding coefficients.

Table S4.1. Fixed effect results from linear and generalized linear mixed models fit with `fitNullModel` in `GENESIS`. The matrix of pairwise kinship coefficients (GRM) was included as a random effect in all models to control for family relatedness.

FIXED EFFECT COEFFICIENTS

	Beta	SE	χ^2	p-value	Beta	SE	χ^2	p-value
	BMI				HEIGHT			
Intercept	22.601	1.283	310.251	1.93E-69	142.859	3.096	2128.792	0.00E+00
Sex (male)	-3.493	1.037	11.338	7.60E-04	10.171	2.511	16.412	5.10E-05
Age	0.083	0.022	14.307	1.55E-04	0.289	0.053	29.879	4.60E-08
PC1	-0.640	5.051	0.016	8.99E-01	-35.729	12.207	8.567	3.42E-03
PC2	-1.748	4.889	0.128	7.21E-01	19.643	11.820	2.762	9.66E-02
	WEIGHT				TRICEP SKINFOLD			
Intercept	49.515	3.974	155.227	1.25E-35	14.909	1.202	153.920	2.41E-35
Sex (male)	-1.221	3.198	0.146	7.03E-01	-8.276	1.003	68.104	1.55E-16
Age	0.349	0.067	26.932	2.11E-07	-0.002	0.021	0.013	9.09E-01
PC1	35.543	15.606	5.187	2.28E-02	-2.342	4.831	0.235	6.28E-01
PC2	15.506	15.094	1.055	3.04E-01	6.557	4.684	1.960	1.62E-01
	MUAC				SYSTOLIC BP			
Intercept	9.168	0.388	557.930	2.37E-123	99.554	4.858	419.982	2.46E-93
Sex (male)	-0.243	0.316	0.593	4.41E-01	5.581	3.549	2.473	1.16E-01
Age	0.040	0.007	35.449	2.62E-09	0.698	0.081	73.375	1.07E-17
PC1	-2.474	1.534	2.601	1.07E-01	-26.243	18.997	1.908	1.67E-01
PC2	1.584	1.486	1.135	2.87E-01	36.142	17.900	4.077	4.35E-02
	DIASTOLIC BP				DIABETIC			
Intercept	65.739	3.006	478.110	5.51E-106	-6.002	1.634	13.493	2.00E-04
Sex (male)	3.751	2.266	2.740	9.79E-02	-1.307	0.901	2.101	1.47E-01
Age	0.314	0.052	36.915	1.23E-09	0.060	0.023	6.713	9.60E-03
PC1	32.668	11.966	7.454	6.33E-03	2.079	5.673	0.134	7.14E-01
PC2	14.987	11.230	1.781	1.82E-01	6.680	4.999	1.786	1.81E-01
	GLUCOSE TEST							
Intercept	-6.024	2.192	7.551	6.00E-03				
Sex (male)	-2.217	1.270	3.047	8.09E-02				
Age	0.063	0.033	3.725	5.36E-02				
PC1	-6.078	7.098	0.733	3.92E-01				
PC2	16.059	8.187	3.848	4.98E-02				

Table S4.2. Association score test statistics for SNPs with the 5 smallest p-values. freq=minor allele frequency, MAC=minor allele count, Score.SE=Score standard error, Score.Stat=Score z-test statistic, Est=effect size estimate per copy of minor allele, Est.SE=effect size standard error, PVE=proportion of phenotype variance explained

variant.id	chr	pos	n	freq	MAC	Score	Score.SE	Score.Stat	Score.pval	Est	Est.SE	PVE
						WEIGHT						
rs11761744	7	153374029	159	0.10	32	1.36	0.26	5.19	2.12E-07	19.74	3.80	0.18
rs1579730	15	70953639	159	0.23	72	1.77	0.38	4.65	3.26E-06	12.27	2.64	0.14
rs7607123	2	33416865	159	0.14	43	1.43	0.31	4.63	3.51E-06	15.07	3.25	0.14
rs10123214	9	122815896	159	0.11	34	1.26	0.28	4.54	5.54E-06	16.43	3.62	0.13
rs12898513	15	60984546	149	0.14	43	1.28	0.28	4.54	5.64E-06	16.16	3.56	0.13
						HEIGHT						
rs10269661	7	89486165	159	0.13	41	-1.56	0.33	-4.72	2.34E-06	-14.31	3.03	0.15
rs875727	1	70908055	159	0.24	76	-2.30	0.49	-4.69	2.72E-06	-9.57	2.04	0.14
rs2997550	9	7446748	159	0.41	130	2.59	0.56	4.65	3.40E-06	8.32	1.79	0.14
rs13424719	2	10963283	159	0.44	141	-2.60	0.57	-4.59	4.46E-06	-8.09	1.76	0.14
rs2509814	11	95096350	159	0.12	63	-1.86	0.41	-4.58	4.70E-06	-11.27	2.46	0.14
						MUAC						
rs1579730	15	70953639	159	0.23	72	17.93	3.85	4.66	3.15E-06	1.21	0.26	0.14
rs11761744	7	153374029	159	0.10	32	11.94	2.66	4.49	7.19E-06	1.69	0.38	0.13
rs11621381	14	61241678	159	0.20	64	16.11	3.64	4.43	9.61E-06	1.21	0.28	0.13
rs9371382	6	155653262	155	0.48	149	-18.48	4.24	-4.36	1.31E-05	-1.03	0.24	0.12
rs7645376	3	188207114	159	0.12	37	11.48	2.67	4.30	1.70E-05	1.61	0.38	0.12
						TRICEP SKINFOLD						
rs10994198	10	60096277	158	0.29	90	6.11	1.33	4.58	4.65E-06	3.43	0.75	0.14
rs11685957	2	134015266	158	0.15	47	4.63	1.02	4.54	5.63E-06	4.45	0.98	0.14
rs2393599	10	60097103	158	0.11	35	4.12	0.92	4.48	7.42E-06	4.87	1.09	0.14
rs2902280	3	176855946	156	0.22	70	4.78	1.07	4.47	7.71E-06	4.19	0.94	0.14
rs1398548	4	66492500	158	0.11	35	3.79	0.85	4.46	8.10E-06	5.25	1.18	0.14
						SYSTOLIC BLOOD PRESSURE						
rs11968187	6	155857851	134	0.46	124	1.64	0.36	4.57	4.82E-06	12.75	2.79	0.16
rs9595638	13	47280373	138	0.17	47	1.26	0.28	4.46	8.30E-06	15.73	3.53	0.15
rs2846690	11	128939032	138	0.30	84	1.55	0.35	4.38	1.19E-05	12.37	2.82	0.14
rs7861242	9	84300627	138	0.20	55	-1.28	0.29	-4.35	1.36E-05	-14.82	3.41	0.14
rs13379337	14	34755224	138	0.32	88	-1.56	0.36	-4.32	1.54E-05	-12.00	2.78	0.14
						DIASTOLIC BLOOD PRESSURE						
rs17153801	7	77466146	137	0.23	63	-2.57	0.52	-4.92	8.88E-07	-9.42	1.92	0.18
rs232405	21	21437359	138	0.21	58	2.23	0.47	4.75	2.05E-06	10.10	2.13	0.17
rs1842840	3	157010268	138	0.42	116	2.73	0.58	4.72	2.41E-06	8.13	1.73	0.17
rs3811259	14	22312080	137	0.22	59	-2.28	0.49	-4.61	4.13E-06	-9.32	2.02	0.16
rs4836864	9	116818456	138	0.26	72	2.68	0.58	4.59	4.34E-06	7.89	1.72	0.16
						DIABETIC						
rs17170837	7	37017687	159	0.13	42	4.89	1.06	4.59	4.24E-05	4.31	0.94	0.46
rs4822763	22	26664677	159	0.15	47	4.84	1.17	4.14	1.20E-04	3.53	0.85	0.37
rs2636870	9	114339766	159	0.27	86	6.36	1.59	4.00	1.59E-04	2.52	0.63	0.35
rs12429891	13	43062428	159	0.12	38	4.64	1.18	3.95	1.59E-04	3.36	0.85	0.34
rs7117475	11	11485544	156	0.12	36	4.41	1.05	4.21	1.71E-04	4.02	0.95	0.39
						GLUCOSE TEST						
rs8023607	15	61829409	103	0.07	14	3.06	0.70	4.35	9.19E-05	6.21	1.43	0.62
rs10845474	12	12085383	103	0.18	36	5.68	1.39	4.09	9.60E-05	2.95	0.72	0.55
rs3823311	6	151444161	103	0.13	26	3.61	0.98	3.67	1.41E-04	3.73	1.02	0.44
rs12257289	10	16228813	102	0.13	27	4.50	1.13	3.99	1.81E-04	3.53	0.89	0.52
rs12526978	6	109987833	103	0.09	19	3.48	0.98	3.54	2.26E-04	3.59	1.02	0.41

VITA

Monica Keith (née Ahsan) is originally from Plano, TX. She earned a Bachelor of Science in Anthropology with a minor in Biology from Texas State University in 2012. As an undergraduate, Monica interned at the Texas Biomedical Research Institute in San Antonio where she first began to run laboratory and statistical analyses of health and genetic data. Upon receiving a Life Sciences Fellowship, Monica began her graduate studies at the University of Missouri in Fall 2012.

Monica completed a Master's thesis in Anthropology at MU titled "Genetic and Maternal Effects on Neonatal Survival in the Western Lowland Gorilla", developing programming and modeling skills with pedigree data. She then began her dissertation project to explore genetic variation in the Caribbean and research biocultural influences on cardiometabolic health. In addition to researching local health concerns in rural Dominica, Monica collaborates with other anthropologists to analyze biocultural sources of variation on health outcomes in unique ecologies, focusing on small-scale populations.

Monica taught lecture and laboratory courses in Anthropology and Biological Sciences at MU from 2016-2020. Her current publications include articles in the American Journal of Physical Anthropology, the American Journal of Human Biology, and Philosophical Transactions of the Royal Society, Biological Sciences.