

Original Research Article

What Could You Do with 400 Years of Biological History on African Americans? Evaluating the Potential Scientific Benefit of Systematic Studies of Dental and Skeletal Materials on African Americans from the 17th Through 20th Centuries

FATIMAH JACKSON,^{1,2*} LATIFA JACKSON,^{1,3} CHRISTOPHER CROSS,^{1,4} AND CAMERON CLARKE^{1,2}

¹*Cobb Research Laboratory, College of Arts and Sciences, Howard University, Washington, DC*

²*Department of Biology, College of Arts and Sciences, Howard University, Washington, DC*

³*National Human Genome Center, College of Medicine, Howard University, Washington, DC*

⁴*Department of Anatomy, College of Medicine, Howard University, Washington, DC*

Objectives: How important is it to be able to reconstruct the lives of a highly diverse, historically recent macroethnic group over the course of 400 years? How many insights into human evolutionary biology and disease susceptibilities could be gained, even with this relatively recent window into the past? In this article, we explore the potential ramifications of a newly constructed dataset of Four Centuries of African American Biological Variation (4Cs).

Methods: This article provides initial lists of digitized variables formatted as SQL tables for the 17th and 18th century samples and for the 19th and 20th century samples.

Results: This database is dynamic and new information is added yearly. The database provides novel opportunities for significant insights into the past biological history of this group and three case study applications are detailed for comparative computational systems biology studies of (1) hypertension, (2) the oral microbiome, and (3) mental health disorders.

Conclusions: The 4Cs dataset is ideal for interdisciplinary “next generation” science research and these data represent a unique step toward the accumulation of historically contextualized Big Data on an underrepresented group known to have experienced differential survival over time. *Am. J. Hum. Biol.* 28:510–513, 2016. © 2016 The Authors American Journal of Human Biology Published by Wiley Periodicals, Inc.

The 21st century has been called the era of Big Data (global.sap.com). If it is properly contextualized, Big Data can provide a potential analytical advantage in research. It can identify interrelationships between variables, decrease analytical times, and provide predictive insights that would otherwise be unavailable. There is a paucity of data, big or otherwise, on African Americans (AAs). However, according to the U.S. Department of Health and Human Services (2014), AAs have a higher death rate compared to European Americans for cardiovascular disease, stroke, cancer, asthma, influenza and pneumonia, diabetes, and HIV/AIDS. Four hundred years of AA biological information can help unearth the science behind these health inequities and provide significant and substantial insights.

As of 2013, the District of Columbia had the largest percentage of resident AAs per total population in the United States at 51% (CDC, 2013). Our 19th and 20th century materials largely come from the DC area. Therefore, not only do we have 400 years but also a representative sample of AAs for the time period within which they were collected. This time period is important since our 400 years of samples were largely collected pre-passage of the 1952 Immigration and Nationality Act and before *Brown V. Board* (1954) and, therefore, provide the genetic integrity backing the terms colored, Negro, mulatto, black, and eventually AA. Over the last few decades, there has been a large influx of peoples from the African continent and elsewhere in the African diaspora into the United States. As a result, 4Cs may not be the best representation of current 21st century AAs. However, recent comments by Drew Gilpin Faust, president of Harvard University, suggest that “African and African American questions are at the core of world history and world identity as well as US history and US identity” (Faust, 2013 cited in Burke, 2014, p.16). Therefore, the need for a comprehensive historical biology

database centered on Africans and their North American descendants represents an important step in the right direction. To our knowledge, Howard University is the only institution of higher learning with access to this dimension of information on AAs historical biology.

BACKGROUND AND LITERATURE REVIEW

Our 17th and 18th century samples are derived from the New York African Burial Ground (NYABG) remains currently housed at Howard University and on loan from the National Park Service. The NYABG is the nation’s earliest and largest African burial ground (LaRoche and Blakey, 1997). These previously buried samples reflect African/AA biological diversity from the late 17th to late 18th centuries in New Amsterdam/New York. While there are over 400 burials, we have well-documented, archived biological materials from 250 individuals. Our 19th and 20th century samples come from the Cobb Collection (CC). The CC contains 699 individuals from the mid to late 19th and early to mid 20th centuries. It is the nation’s third largest collection of human skeletal remains and is the largest containing a majority of AA samples (83%). Together, these two collections of human skeletal remains

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Contract grant sponsor: National Park Service Cooperative Agreement; Contract grant number: H1171090106.

*Correspondence to: Fatimah Jackson, Ph.D., 415 College Street, NW, Department of Biology, E.E. Just Hall, Howard University, Washington, DC 20059. E-mail: fatimah.jackson@howard.edu

Received 22 July 2015; Revision received 12 November 2015; Accepted 28 November 2015

DOI: 10.1002/ajhb.22821

Published online 9 January 2016 in Wiley Online Library (wileyonlinelibrary.com).

represent 400 years of AA biological history. We have made it a priority to digitize the existing data on these collections in an interactive format that will allow for easy data curation and addition of supplemental data, facilitated searches and sharing, secure storage, and reliable data visualization without loss of privacy.

METHODS

The 4Cs dataset was developed from existing and new data on the NYABG and the CC, both containing human skeletal and dental remains of predominantly of AA origin. Data were entered into an Excel file and then transferred to a SQL Server 2014 platform. Microsoft SQL 2014 was used for running the queries in the following fields of interest. Table 1 lists the digitized variables associated with samples from specific time periods, their digitization status, and additional comments.

RESULTS

As a consequence of the methods used, each individual in the 4Cs database has a growing wealth of demographic, biological, and historical data available for comparative analyses. A number of variables are currently under analysis at the CRL and nearby laboratories. Hence, a number of key variables are listed as incomplete and in process as new results become available monthly. A profile of the range of variability in the data is beginning to emerge and we are able to visualize the following sample case applications to demonstrate the kinds of interactive assessments now possible.

Sample case application: frequency variation in hypertension disease susceptibility over 400 years

An estimated 972 million people worldwide were hypertensive in 2000 and this number is predicted to increase to about 1.5 billion by 2025 (Hajar et al., 2006). In the United States, the prevalence of hypertension (HTN) among adults aged >20 years from 2009 to 2012 was 32.2% representing nearly 90 million people (CDC, 2014). HTN is dramatically high among some individuals and groups of recent African descent. The 4Cs database allows us to evaluate HTN across the years and assess its impact as a current major health disparity within an historical context.

The current high frequency of HTN and stroke among AAs suggests that there may be a contributing ancestral component. Assessing 400 years of genetic variation data should allow for the observation of any microethnic substructure (Microethnic substructure refers to within group population stratification that is the result of local patterns of social and cultural identity. Microethnic substructure occurs within macroethnic (i.e., socio-racial) groupings; Jackson, 2004, 2008) and the identification of relevant polymorphisms contributing to the etiology of HTN and stroke in this group. The historically recent 400 years of the transatlantic trade in enslaved Africans from West and West Central Africa and their transport to the Americas via the Middle Passage may have exposed these forced migrants to selection for electrolyte conservation (Jackson, 1991). Add to this, a diet in many parts of North America that is disproportionately high in sodium may have created the perfect setting for a gene X environment mismatch that would increase susceptibilities to hypertensive heart

disease. The people's ancestral background may now predispose them to an early-onset, salt-sensitive HTN and an increased vulnerability to stroke (Jackson, 1991). Ancestral background and previously adaptive susceptibility genes may account in large measure for the presence of the "Stroke Belt" centered broadly among AAs in the southeastern United States. The 4Cs database should allow researchers to examine hypotheses about HTN susceptibilities, identify the clusters of relevant common and rare genes over time, and track their frequency distributions. This can also allow researchers to monitor population changes in allele frequencies over time and across geographical space within the Washington, DC region.

Sample case applications: change and stability in the oral microbiomes over 400 years

Comparisons of the oral microbiome (Oral microbiome refers to the totality of the microfauna and microflora found within the oral cavity. Collection and analysis of the oral microbiome allows for the reconstruction of individual dietary and drug-use patterns as well as bacterial and viral exposures. Our oral microbiome analyses are currently in progress for all 4C's materials.) across 20 generations are possible with the 4Cs database and would be an important addition to the literature since changes in the oral microbiome over time may reflect the effects of exposure to environmental toxins, dietary patterns, and differential pharmaceutical use (see Dewhirst et al., 2010). Oral microbes are a fundamental part of human development and easily accessible part of the human microbiome (McLean, 2014). Studies of the oral microbiome could be central to the reconstructions of the effects of the built and social environments on the episome as the oral cavity and upper respiratory track are the main ports of entry of microorganisms into the human body (Macovei et al., 2015). Using the dental calculus (calcified tartar or plaque) (Weyrich et al., 2015), data in 4Cs will allow us to explore historical aspects of human diet and health within the New York City and Washington, DC regions and the presence of specific oral pathogens over time. Ancient human oral plaque preserves a wealth of biological data (Metcalf et al., 2014) and could provide great insights into diversity and change in this population. It may even be possible to reconstruct micro-ethnicity-specific bacterial signatures in the oral microbiomes from these dental and skeletal remains (see Mason et al., 2013).

Sample case applications: computational systems biology approaches in mental health disorders

Computational approaches were developed to address questions of complex disease phenotypes in the 4Cs populations, including a set of mental health disorders known to have a strong genetic component (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). These computational algorithms incorporate data from the 4Cs dataset, publicly available genetic polymorphism datasets, proteomic data, functional and pathway annotation datasets. The goal of these algorithms is to gain genomic insights into the functional contributors to complex diseases and to identify those regions of the genome that potentially may be implicated in disease phenotypes in AA populations. This approach is able to elucidate the intersections of multiple complex disorders with genetic components in human populations.

TABLE 1. Variables associated with the 17th, 18th, 19th, and 20th century biological samples

	17th and 18th Century	19th and 20th Century	Completed	In process	In Prep
Identifier	Series identification Catalog Number	CRL Number Name Morgue Number Death Certificate Number	■ ■ ■ ■ ■		
Gender ^a	Assumed to conform to sex	Medical identity	■		
Sex	Determined molecularly and/or anatomically		■ Anatomy	■ Molecular	
Age Grouping	Age Grouping Age at Death	Age	■ ■		
Death Information	Year of Death	Date of Death Cause of Death Place of Death Autopsy Date Corpse Received Dissecting Comments	■ ■ ■ ■ ■ ■ ■		
Geographical Information	Grid Coordinates Elevation of highest skeletal element (feet AMSL)	Address at time of Death Birthplace Father's Birthplace Mother's Birthplace Duration of Residence in DC	■ ■ ■ ■ ■ ■	■ ■ ■	
Skeletal Information	Initial Status Initial Number of Pieces Initial Weight Skeletal Biology DNA Skeletal Element	Initial Status Initial Number of Pieces Initial Weight	■ ■ ■ ■	■	
DNA extraction	Cementum Extraction Amount of Bone needed for Isolation Amount of Bone returned for Curation Amount of DNA Extractable Amount of DNA needed per Analysis Amount of DNA used in Analyses Amount of DNA returned for Curation	Cementum Extraction DNA Skeletal Element Amount of Bone needed for Isolation Amount of Bone returned for Curation Amount of DNA Extractable Amount of DNA needed per Analysis Amount of DNA returned for Curation Amount of DNA used in Analyses		■ ■ ■ ■ ■ ■ ■ ■	
Oral Microbiome	Contaminating Bacterial Species Contaminating Viruses (dental calculus)	Contaminating Bacterial Species Contaminating Viruses Oral Microbiome Results (dental calculus) DNA Extraction and Isolation and Human DNA Quantification		■	■ ■ ■
Human DNA Isolation	DNA Extraction and Isolation Human DNA Quantification			■ ■	
Sequencing Results Environmental Analyses	Sequencing Optimization for NGS and Sanger platforms Grave Soil Analysis	Marital Status Occupation prior to Death Additional clinical and/or Historical Comments Geospatial Analyses	■ ■ ■	■ ■	
Ethnic Identity	African Regional Origins Individual Historical Profile Historical Ethnographic Information		■		■ ■

^aGender, in contrast to sex, refers to the social and cultural identity of the individual while sex refers to the biological designation of the individual based on genetics, anatomy, and physiology.

Currently, this unique computational systems biology approach is being used to identify genomic regions involved in schizophrenia, bipolar disorder and depression disorders in AAs. The 4Cs dataset has a number of individuals whose samples were derived from psychiatric hospitals, and were noted to have had mental health disorders during their lifetime. By performing these compu-

tational analyses with publicly available AA genetic polymorphisms, it is possible to identify those regions of particular import in AA manifestations of mental health disorders (see Jackson, 2014). This information will then be used to survey these genomic regions in the 4Cs population, allowing us to evaluate changes in the genomics of psychiatric disorders over 400 years.

DISCUSSION

Relevance of database for advanced biohistorical studies

The 17th and 18th centuries represent the very beginnings of the emergence of African descended peoples in the Americas. Most Africans arrived in the Americas as a result of the transatlantic trade in enslaved individuals, a trade which became economically and demographically prominent by the year 1700 AD (Manning, 2004). African labor was in high demand in the New World and provided a broad range of agricultural, domestic, and artisanal tasks. The impacts of these social and economic conditions should be reflected in the skeletal, dental, and molecular biology of these individuals. Four hundred years of biological data allow for the development of complex research designs; ones that address both the psychological concerns of the descendant AA communities and are also responsive to pressing scientific issues, both applied and theoretical. Four hundred years represents approximately 20 generations, counting the average time between generations as 20 years. This extended time depth permits a panoramic vision of health disparities, disease susceptibility, and microevolutionary processes. With data from 400 years, researchers can investigate a wealth of hypotheses, particularly concerning the biological and cultural transition from African to AA identities (LaRoche and Blakey, 1997; see Gomez, 1998) in the North American context.

The 4Cs data can also allow researchers to examine patterns of gene flow between major population clusters in the United States. For example, tests of various admixture models for AA populations can be facilitated by application of the 4Cs database.

CONCLUSIONS

The 4Cs dataset allows for the novel study of AA biological variation over 400 years of North American existence. As the data expand in this database, we expect that they will become increasingly heterogeneous and complex. As computational biologist Lawrence Hunter has said, "Getting the most from the data requires interpreting them in light of all the prior knowledge." (cited in Marx, 2013). Research collaborations that provide for interdisciplinary investigations are encouraged as such efforts will increase our longitudinal insights into the contemporary health disparities, explore evidence for selection and genetic drift in the North American environment, genomically reconstruct the African and non-African regions of origin of these New World Africans, and evaluate the epigenomic impact of life in early and contemporary America on AA genome expression patterns. To our knowledge, the Cobb Research Laboratory (The website for the W. Montague Cobb Research Laboratory is www.cobbresearchlab.com.) at Howard University is the only academic institution with such a breadth of biological samples and with such great potential for illuminating the AA past. It is our hope that the 4Cs dataset will be accessed by a diverse and international array of researchers such that the database will expand over time

and will benefit from the intellectual efforts of many scientists and citizen scholars.

ACKNOWLEDGMENTS

We thank the National Park Service for access to the NYABG remains. Competing Interests: The authors declare no competing interests. Author Contributions: The 4Cs database and this paper were conceived by Fatimah Jackson, Professor and Director of the W. Montague Cobb Research Laboratory. Dr. Latifa Jackson contributed the computational systems biology insights. Mr. Christopher Cross contributed anatomical data which were incorporated into the dataset encoded by Mr. Cameron Clarke.

LITERATURE CITED

- Burke KM. 2014. The Hutchins Center Honors Inauguration. New York: Hutchins Center for African and African American Research at Harvard University; Thornwillow Press, Ltd.
- CDC. 2013. CDC Health Disparities and Inequalities Report — United States. Available at: <http://www.cdc.gov/mmwr/pdf/other/su6203.pdf>, accessed on June 18, 2015.
- CDC. 2014. Health, United States, 2014: At A Glance. Available at: <http://www.cdc.gov/nchs/data/abus/abus14.pdf#059>, accessed on June 17, 2015.
- Cross-Disorder Group of the Psychiatric Genomics Consortium. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45(9):984–994.
- Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade WG. 2010. The human oral microbiome. *J Bacteriol* 192(19):5002–5017.
- Gomez M. 1998. Exchanging our country marks: the transformation of African identities in the Colonial and Antebellum South. Chapel Hill: University of North Carolina Press.
- Hajar I, Kotchen J, Kotchen T. 2006. Hypertension: trends in prevalence, incidence, and control. *Annu Rev Public Health* 27:465–490.
- Jackson F. 2004. Human genetic variation and health: ethnogenetic layering as a way of detecting relevant population substructuring. *Br Med Bull* 69:215–235.
- Jackson FLC. 1991. An evolutionary perspective on salt, hypertension, and human genetic variability. *Hypertension* 17(1):1129–1132.
- Jackson FLC. 2008. Ethnogenetic layering (EL): an alternative to the traditional race model in human variation and health disparity studies. *Ann Hum Biol* 35(2):121–144.
- Jackson LF. 2014. Addiction, mental health, and infectious disease: a complex web of genetic interactions. Ph.D. Dissertation. Ann Arbor: Drexel University, ProQuest.
- LaRoche CJ, Blakey ML. 1997. Seizing intellectual power: the dialogue at the New York African Burial Ground. In: McDavid C, Babson DW, editors. In the realm of politics, prospects for public participation in African American and plantation archaeology, Vol. 31. *Historical Archaeology*. p 84–106.
- Macovei L, Chen T, Teles F, Hasturk H, Paster BJ, Campos-Neto A. 2015. The hidden 'microbiome of the human healthy oral cavity and upper respiratory tract. *J Oral Microbiol* 7:26094.
- Manning P. 2004. Slavery and African Life: Occidental, Oriental, and African Slave Trades. African Studies Series. Cambridge: Cambridge University Press.
- Marx V. 2013. Biology: the big challenges of big data. *Nature* 498:255–260.
- Mason MR, Nagaraja HJ, Camerlengo T, Joshi V, Kumar PS. 2013. Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. *PLoS One* 8(10):e77287.
- McLean JS. 2014. Advancements toward a systems level understanding of the human oral microbiome. *Front Cell Infect Microbiol* 4:98.
- Metcalf JL, Ursell LK, Knight R. 2014. Ancient human oral plaque preserves a wealth of biological data. *Nat Genet* 46(4):321–323.
- U.S. Department of Health and Human Service African American Profile in the U.S. 2014. Available at: <http://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=61>, accessed on June 16, 2015.
- Weyrich LS, Dohney K, Cooper A. 2015. Ancient DNA analysis of dental calculus. *J Hum Evol* 79:119–124.