Howard University

From the Selected Works of Fatimah Linda Collier Jackson

February, 2008

Ethnogenetic layering (EL): An alternative to the traditional race model in human variation and health disparity studies

Fatimah Linda Collier Jackson, *Howard University*

SELECTEDWORKS™

Available at: <http://works.bepress.com/fatimah-linda-collier-jackson/6/>

ORIGINAL ARTICLE

Ethnogenetic layering (EL): an alternative to the traditional race model in human variation and health disparity studies

F. L. C. JACKSON

*Genomic Models Research Group, Biological Anthropology Research Laboratory,
Department of Anthropology, University of Maryland,
College Park, Maryland, USA*

(Received 1 May 2007; revised 6 December 2007; accepted 9 January 2008)

Abstract

Background: Traditionally, studies in human biodiversity, disease risk, and health disparities have defined populations in the context of typological racial models. However, such racial models are often imprecise generalizations that fail to capture important local patterns of human biodiversity.

Aim: More explicit, detailed, and integrated information on relevant geographic, environmental, cultural, genetic, historical, and demographic variables are needed to understand local group expressions of disease inequities. This paper details the methods used in ethnogenetic layering (EL), a non-typological alternative to the current reliance of the biological racial paradigm in public health, epidemiology, and biomedicine.

Subjects and methods: EL is focused on geographically identified microethnic groups or MEGs, a more nuanced and sensitive level of analysis than race. Using the MEG level of analysis, EL reveals clinical variations, details the causes of health disparities, and provides a foundation for bioculturally effective intervention strategies. EL relies on computational approaches by using GIS-facilitated maps to produce horizontally stratified geographical regional profiles which are then stacked and evaluated vertically. Each horizontal digital map details local geographic variation in the attributes of a particular database; usually this includes data on local historical demography, genetic diversity, cultural patterns, and specific chronic disease risks (e.g. dietary and toxicological exposures). Horizontal visual display of these layered maps permits vertical analysis at various geographic hot spots.

Results and conclusions: From these analyses, geographical areas and their associated MEGs with highly correlated chronic disease risk factors can be identified and targeted for further study.

Keywords: *Microethnic groups, biological lineage coalescence, GIS, computational anthropology, race, human biodiversity*

Introduction

As of March 2007, over 113 000 articles were published in the scientific literature and reported by PubMed to have used the classical Linnaean racial categories to define the populations studied. The use of the traditional race model is ubiquitous in biomedicine, epidemiology, and public health studies. Such archaic models do not well define existing human biodiversity however, and in fact, may, by their gross generality, obfuscate clinically important and evolutionarily more meaningful local patterns of human variation, particularly those relevant to specific disease risk and existing health disparities. Furthermore, the notion of there existing different human biological races today, from the perspective of scientific taxonomy, is unacceptable. Biological race is generally considered to be the equivalent of the taxonomic category subspecies. What variation exists among humans is not of the evolutionary magnitude to justify different racial or subspecific designations, but this variation may be of clinical relevance and public health significance. The evolutionary and the clinical are often two distinct levels of analysis.

The current concept and use of race grew out of an European worldview and intellectual paradigm used to classify human beings based on their observable physical appearance. This perspective became codified in European and European-influenced science. Later, other characteristics such as regional customs or behavioral differences, language, and intellect were linked to race. This linkage helped to reinforce the earliest notions of an existing hierarchy intimately infused with notions of European superiority. Many of the rules that govern the determination of race are still rooted in social and cultural traditions, rather than sound biological reasoning, hence the proliferation of 'races' in the literature and the ambiguity often observed between such groupings in the contemporary scientific literature. At best, what passes for 'race' is primarily a social construction with some biological aspects that depend on the specific context in which a particular phenotype is realized (see Jackson 1992). Such contexts can be genetic, ethnographic, demographic, historical, economic, or ecological.

Current studies of human variation often involve genetic analyses of the varying phenotype-relevant contributing components. These studies show that regardless of the way racial groups are classified, two unrelated people from the same 'racial group' are on average as different from each other as two random people from different 'racial groups'. As our understanding of human variation grows, it becomes increasingly more difficult to identify a clear link between existing biological diversity and the traditional applications of racial classification, especially as they relate to health (Marks 1994, 1995; Goodman 1997; Keita et al. 2004). An alternative is needed that systematically addresses locally generated patterns of human biodiversity, that can be aggregated to make more general regional statements, if needed, but that is not based on the biologically ambiguous macroethnic category of 'race'. This paper reports on an alternative to the race model, ethnogenetic layering (EL), its methods, applications, strengths, and limitations.

EL relies upon Geographic Information System (GIS) computational technology to digitize and horizontally layer various data (e.g. genetic, toxicological, historical, demographic, etc.) on a particular geographical region for subsequent multilayer vertical analysis. Geographical sites are identified by latitude and longitude. Geography is held constant while site- and group-specific data are evaluated statistically at these predetermined geographic coordinates. As a result, EL allows researchers to identify local patterns of contemporary patterns of human biological variation within socially constructed macroethnic aggregates and, in so doing, tease out previously hidden substructure and its likely causes.

Over the last 17 years, the Genomic Models Research Group in the Department of Anthropology at the University of Maryland has crafted a novel, sophisticated, and reproducible method for reconstructing and examining the human biocultural mosaic for the purpose of predicting disease susceptibility and risk in local microethnic groups or MEGs. In EL, the level of analysis is the MEG; most MEGs are made of related or coalesced biological lineages usually residing in close geographical proximity. *Thus, a MEG is a local constellation of genetically related biological lineages with shared cultural practices, environmental exposures, geographical residence, demographic status, and historical backgrounds.* Although MEGs are socially constructed, they often exhibit more within-group biological homogeneity and genetic correspondence than expected because most affiliates are from related biological lineages. This higher degree of biological continuity among members is often a result of recent localized multigenerational patterns of endogamy. The shared cultural practices within a MEG are often the result of reciprocal cultural exchange between group members. Augmenting these similarities are often shared health-related exposures to environmental toxins and communicable disease agents. In terms of health disparities, local aggregates of MEGs may have more in common with each other than with their general 'racial' cohorts 1000 miles away.

EL is particularly well suited to studies of human biodiversity in the US population because contemporary Americans represent a highly heterogeneous, exceptionally variable subset of humanity. Our recent ancestral origins from often identifiable and well-documented locations elsewhere in the world, our historical patterns of gene flow, and our defined opportunities for genetic drift have produced a mosaic that can be evaluated for biological lineage studies relevant to health and disease and investigations of recent group heritage. This information can also be used to reconstruct regional MEGs, the local geographical groupings of biological lineages that provide substructure to the larger, more geographically diffuse macroethnic groupings frequently mislabeled as 'race'.

Theoretical foundations

Recent common human origins

EL is based upon a set of theoretical assumptions underlying our methods. Our first presumption is that of a recent African origin for modern humanity (Horai et al. 1995; Armour et al. 1996). This scientific premise of a monogenic evolutionarily recent common African origin of approximately 150 000 years ago further suggests that all modern humans today are a single species and subspecies, *Homo sapiens sapiens*. This means, by definition, that the current diversity within our species is evolutionarily recent and not substantial enough to warrant separate subspecies categories. The taxonomic 'race' for all modern humans is the second *sapiens* designation.

Epigenetic filtering of gene expression

A second theoretical assumption incorporated into EL is the notion that the expression of an individual's genetic background (the genotype) is modified as it passes through a series of various cultural and environmental filters to produce the phenotype or the expressed genotype. These filters provide the epigenetic context for gene expression. The more similar these filters over time, the more consistent their impact on the genetic background, all else

being equal. This assumption places local gene–environment outcomes at the forefront in defining a MEG. Figure 1 represents these filtering modifiers of the genetic background. Here, sociocultural, abiotic, and biotic environmental filters are crucial in evaluating most disease susceptibilities and human phenotypic variation because these filters regulate *what*, *when*, and *how* genes may be expressed. It is the expressed genotype or the phenotype that is the target of natural selection. For example, in this model, bioactive phytochemical exposure through the diet, such as cyanogenic glycosides through cassava (*Manihot esculenta*) ingestion, alters the physiological expression of a sickle cell genotype by irreversibly chemically altering (i.e. carbamylating) the hemoglobin (Jackson et al. 1988; Jackson 1990). When the genetic information is subjected to the influence of these dietary filters, the hemoglobin gene product is altered, exhibiting different physiological properties than could be predicted by solely studying the genotype.

Biological lineage coalescence

A third theoretical assumption of EL is that within designated geographical regions, specific human biological lineages can be identified and that under conditions of sustained geographical residence, the tendency is to observe coalescence or merging of these lineages over time. The coalescence of human biological lineages helps to explain the enhanced probability of finding local, MEG-specific disease markers but the difficulty in sustaining 'race-specific' disease markers across broader geographical regions. The historical and contemporary patterns of human biological lineage coalescence vary by geographical locale.

Epigenetic Filtering Influences on the Genotype as expressed in the Phenotype

Figure 1. The expression of the genetic background (= genome) is modified as it passes through the various filters resulting in an expressed genotype (= phenotype) that differs from that predicted solely by the genome. Redrawn from Jackson (2004).

Biological lineage coalescence is explained as follows: for each of us, the number of direct ancestors *increases* exponentially each generation (approximately every 20 years) as we go backward in time, while the number of humans alive on the planet *decreases* as we go backward in time. We each start off with two parents, then four grandparents, then eight great grandparents, then 16 great great grandparents, and so forth. In the year 1619 CE, only 1 year after a Dutch ship brought 19 captured Africans to a site near Jamestown, VA and the year before *The Mayflower* arrived at a site near Plymouth Rock, MA bearing English settlers, each of us alive today had, on average 2^{20} direct ancestors, or 1 048 576 individuals. The farther backward in time we travel, the greater the number of direct ancestors we have. However, the farther backward in time we go, the fewer the number of actual humans alive on the planet. Therefore, it becomes quite clear that the farther backward in time we go, the increased the probability that we share ancestors in common. If we contrast the graph of direct ancestors in Figure 2 with the estimated world population figures, it is clear that there must have been at least as much lineage coalescence or convergence among humans in the historical past as there is today. Sooner or later, each of our biological lineages merge

Figure 2. Depiction of increasing number of direct ancestors, as we go back in time, but decreasing world population number. This pattern results in the merging of biological lineages and the sharing of ancestors. Biological lineage convergence also highlights the evolutionary superficiality of 'racial' variation in humans. Redrawn from Jackson (2004).

Figure 3. Some US microethnic groups (MEGs) and their geographical locations (source: Census Bureau 2000).

(through gene flow) into the lineages of others *or* these lineages terminate and become extinct. The same is true, in a much more dramatic sense, with cultural lineages (which are more subject to rapid diffusion, change, and to extinction).

Nonrandom MEG dispersion over geographical space

A fourth theoretical assumption of EL is that MEGs are not uniformly distributed over broad geographical space. This nonrandom distribution of MEGs is the foundation for the nonrandom distributions of macroethnic aggregates often referred to as 'race'. Figure 3 highlights the geographical dispersion of various macroethnic (and some microethnic) groups based upon their majority presence in specific US counties (as reported by the United States Bureau of the Census). Here the divergent geographical locations of various European American, African American, and Native American MEGs are indicated. Figure 4 overlays the geographical foci of some specific chronic diseases on this demographic mosaic of MEGs. This overlay demonstrates that many of chronic disease disparities are primarily geographical, more than 'racial' *per se*. This suggests a major role for the filters of Figure 1 in modulating the gene-environment interactions that influence these disparities.

Local patterns of biological lineage coalescence are often the product of deep historical associations and affiliations among many MEGs. Such historical linkages are too frequently overlooked in the process of biomedical research, limiting the explanatory potential of such studies. EL requires a finer level of assessment than that presented by simply looking at macroethnicity or traditional race categories. The strategy of EL permits the identification of regional MEGs while concomitantly suggesting relevant environmental, genetic, historical, and cultural variables that may explain local chronic disease inequities. MEGs represent a more fine-tuned level of analysis in health disparities research and intervention. This is represented diagrammatically in Figure 5.

Figure 4. Regional variations in diseases with national health disparities.

EL emphasizes within-group differentiation far more than traditional macroethnic studies afford. Furthermore, the attention to regional nuance in EL offers a non-racial model for identifying regional genetic variation, gene-gene, and gene-environment interactions that may significantly predict the likelihood of chronic disease. EL is a conceptual innovation that recognizes that in any given area of the USA, many different constellations of ancestral groups have settled, often sequentially rather than simultaneously, and they have interacted over time. The intensity and duration of these interactions have varied in each area, as have the actual microethnic composition of the resident groups.

Methods

Calculating population density and substructure

Given the non-random geographical distribution of MEGs and the centrality of MEGs to EL analysis, what are the key steps in EL determination? The first questions addressed in EL are:

- (1) *Where* are the people located in a particular geographical area of interest?
- (2) *Who* are they?
- (3) *How* long have they been there?

Figure 5. Shared genes, environmental exposures, and cultural practices in a set of geographically proximate MEGs. Note that the shared elements are not necessarily synchronous across these microethnic groups.

- (4) *What else* is found where they are?
- (5) *How* do these various geographic variables interact and influence local patterns of human biodiversity?

In any one geographical region, as groups become more numerically dense, the potential for within-group substructure increases. An increase in local diversity leads to an increased potential for fissioning and an increase in substructure complexity. Two criteria are used to quantify the saturation and continuance densities of specific MEGs: MEG intensity and MEG duration. The relationship between these two criteria (intensity and duration) is as follows:

$$D_j = \sum \frac{A_{ji}}{\text{field}_{ji}} \quad (1)$$

where D_j equals the mean saturation density (intensity) of a particular MEG in a county, field_{ji} refers to the standardized land unit in county j at time i and A_{ji} equals the number of members of the specific MEG in county j at time i . Since there are different numbers of counties in the various states we have studied, which vary in size and date of incorporation, field_{ji} is a more accurate and standardized unit of MEG density. In most of our models we used the acre ($=0.405 \text{ ha}$, 4047 m^2) as the standardized land unit although many of the

other databases are designated by county or state. Our acre model allows us to consider US counties of both different and changing geographical size in evaluating the distributions of various American MEGs. Sufficiently saturated counties are those in which the intensity of a specific MEG is:

$$D_j \geq 3.0 \text{ individuals.} \quad (2)$$

This value allows us to include urban as well as appropriate rural communities for GIS depiction. To control for the overrepresentation of cities (where D_j is likely to greatly exceed three individuals) EL considers the continuance density (duration) as a measure of density stability over time. The eligibility of a county for inclusion in our studies is based on this model:

$$d_{ji} \geq 1.0. \quad (3)$$

That is, each temporal measurement of saturation density, d_{ji} , must equal or exceed one member of the MEG of interest per acre (or other standardized land unit). If a county did not maintain a d_{ji} of one since its incorporation, the stability of the continuance density is not considered to be adequate enough to warrant further EL determinations. In this way, we exclude seasonally transient groups from our analysis. In multiethnic settings such as the USA, an array of local MEGs is often represented at any one geographic site. The identification of these MEGs, their distribution patterns and history, background genetics, and health risk potential is the objective of EL.

Databases used in EL analysis

EL depends upon a sampling and evaluative protocol that incorporates data from the five key sources outlined in Table I. Relevant databases include validated historic, geographic, cultural, genetic, and health risk source references. Table I indicates some sample sources and sample source references for each of these kinds of datasets. Depending upon the research question, as new databases are available, these are integrated into EL analysis. Data are then digitized and incorporated into GIS maps as discussed below.

GIS data integration and analysis

Ethnogenetic layers are developed from the datasets listed in Table I to provide useful spatial depictions of the interface of ethnic identity, ancestral background, and environmental conditions while holding geography constant. GIS allows us to manage and analyze data for spatial patterns based on physical relationships (proximity, intensity, etc.) as well as non-spatial data related to the space. The quality of the data collected is the only limitation in regards to the resolution of the space being investigated. In the current EL studies, many of the external databases are reported at the state and county levels. However, it is possible to look at finer resolutions with more data collection, as we have done with our acre-level determinations used in our historical database.

The science of GIS allows for its multidisciplinary applications such as in its use with EL. Analysis of these data permit the sampling and modeling of data from diverse sources. Data are reclassified and/or interpolated to generate surface layers that represent

Table I. Databases consulted for ethnogenetic layering of three regions of the USA.

Database	Sample sources	Sample source references
Historic	Archived US Customs records, Census Bureau data, US and European Military records, plantation diaries, narratives of enslaved and previously enslaved persons, early 20th Century and post-emancipation letters, Native American tribal registries, BIA historical documents, oral histories	University of Michigan-based initiative the Inter-University Consortium for Social and Political Research (IUCSPR)'s Historical Census Data Browser, hosted by the University of Virginia's Fisher Library, RG 45 Naval Records Collection of the Naval Records and Library, Archives I, Washington, DC, RG94 Records of the Adjutant General's Office, 1780-1917, American Freedmen's Inquiry Commission
Geographic	Historical demographic records, US Census Bureau data, historical regional maps	Harvard University's CD ROM on the trans-Atlantic Slave Trade, TIGER/Line® Files American Fact Finder Census, Ethnologue
Cultural	Regional ethnographies, oral histories, folklore, local ethnographic surveys	Slavery Illustrated UNC-CH digitization project, WPA Slave narratives, Federal Writers' Project
Genetic	Public databases, regional and country origin studies, specific evaluations of particular local groups of interest	HyrBase++, mitomap, Sorenson Molecular Genealogy Foundation, US Dept of Justice mtDNA Population Database
Health risk	WHO database on environmental toxicity, American Cancer Society reports, CDC reports, EPA pesticide use reports	ECOTOX Database System, INTER, Pesticide Action Network, MDL Toxicity Database, NCI PDQ Database

one particular facet of the EL process. By creating multiple layers in the GIS model, we can determine patterns of groups of effect from particular factors. This can be done with processes such as union or more complex network analysis. All of these capabilities are built into the GIS software. This in turn provides a broad context from which potential MEGs can be subsequently studied.

In our research, all data are digitized and entered into a spreadsheet using Microsoft Excel software and converted into DB4 format, in order to conform to ArcView file formats. EL maps are developed from these spreadsheets to provide useful spatial depictions of the interface of the variables under consideration (while holding geography constant). GIS is used to transform our various databases into raster and vector maps using either AutoCAD (AutoDesk), ArcView (ESRI) or Arc/Info (ESRI). ArcView's Spatial Analyst facilitates the sampling, modeling, and gridding of raster-based data (reclassification, interpolation, creation of surfaces), as well provides an array of statistical assessments. In our studies of three US geographical regions, the maps produced were created with ESRI ArcView 3.2 software, with Network, Spatial, and 3D analyst extensions installed. The programs were

run on three separate workstations. The primary two workstations used were AMD Athalon processor-based Windows NT workstations. Each had 320 MB of system memory and ran at a system bus of 800 mHz. The third workstation was an Intel Pentium-based Windows NT workstation with 128 MB of system memory and running at a system bus speed of 500 mHz.

All base state shapefiles were attained from United States Census TIGER Census Data. Each shapefile was of each separate county contained within each state. These countywide data files were then merged together using ArchView's GeoProcessing Extension.

All the data from Table I are divided into spatial data and non-spatial data. Non-spatial data are made into Excel spreadsheets and/or database tables. Spatial data are collected and inputted into GIS software. The spatial data comes in several forms. Data can be raster data, vector data, and tabular data with spatial information. Raster data are a grid of squares (pixel) where each pixel has a value. The value of each pixel is used to represent a color in the GIS software. The visualization of these colors either form a picture like a historical map or it can form a condition display for the value of the pixels represented. The GIS software can then analyze the conditions for spatial patterns or patterns based on the relationships defined by non-spatial tables. Vector data is viewed as points, lines, and/or areas (polygons). The difference between the vector versus raster maps is how data is handled spatially. Raster data treats all features the same within a given pixel. The resolution (the geographic space covered by a pixel) treats all features as one value. Pixel by pixel comparison for patterns and changes works well if your data is fairly homogeneous within your specified resolution. If the data are not homogeneous within a given resolution it may be better to use vector data. Vector data treats features as a whole regardless of how much space they occupy. These individual features can be attributed with non-spatial attributes and then be analyzed for patterns and interactions with other features using the GIS software. The third type of spatial data are tables with spatial attributes. An example might be a list of a condition (e.g. hypertension, lead poisoning) that has occurred in many geographical locations. The table then has a latitude and longitude (or similar spatial location) for each occurrence. The GIS software can then take that table and geocode (plot the spatial information and create a point, line, or polygon) those locations. All of these data types are used along with the non-spatial data in the GIS software to create ethnogenetic layers.

Results

Example MEGs

Using EL, geographically defined relevant MEGs replace the broad stroke of 'race', providing more detail and specificity. In Louisiana, for example, MEGs include (but are not limited to) the Cajun, Cane River Creole, Clifton Chocktaw, Appalachee, Coushatta, Houma, Red Bones, and Sabines. Each of these MEGs has its own constellation of associated environmental risk factors affecting the expressed genotypes of the people. For example, using the traditional racial categories, Cajun peoples are simply classified as 'Whites' but using EL, we are able to include information on their 200 year group history since migrating to the southeast USA from Canada, local dietary patterns, toxicological profiles, extent of endogamy, and current genetic susceptibilities (see Pollitzer et al. 1977; McDowell et al. 1992; Zlotogora 1993; Oriol 1995; Garg et al. 1998; Scacheri et al. 1999; Waldmeier 2002; ACOG Committee on Genetics 2005; Oetting and Garrity 2006).

In the South Carolina Lowland areas, relevant MEGs include (but again are not limited to) the Brass Ankles, Buckheads, Clay Eaters, Lumbees, Lowland Eastern Band

Cherokee, Sandhillers, Turks, Marlboro Blues, Mestees, Red Legs, Free Moors, Cros, and Gullah/Geechee peoples (see Jackson 2003, 2004). Using the traditional racial categories, Gullah/Geechee peoples are simply classified as 'Blacks' but using EL, we are able to include information on their predominant African cultural origins from the Upper Guinea coast region (Senegambia, Windward Coast, and Sierra Leone), their West Central African (Kongo and Angola) baseline biological origins, their contemporary dietary patterns, current chronic disease susceptibilities, mtDNA haplotypes, and other regional genetic polymorphisms (Parra et al. 2001; McLean et al. 2003, 2005; Ely et al. 2006).

Table II details these and additional US MEGs, including their geographical locations.

EL is able to detect historically important local gene flow between formally distinct groups, frequently mediated by socio-economic circumstances. For example, in areas such as southeastern Virginia, the Low Country of the Carolinas, and around Galphintown near Savannah, Georgia, ancestors of indigenous Native American communities (especially from the five 'civilized' tribes: Choctaw, Chickasaw, Cherokee, Creek, and Seminole) and the ancestors of local African Americans historically shared a common slave experience for over a hundred years (Minges 2001). In major historical trade zones, such as Galphintown, communities emerged of combined West/West Central African-local Native American (primarily Cherokee) heritage that reflected long intertwined Native American and African biological and cultural lineages. In terms of human variation, these unique circumstances created interesting local patterns of human biodiversity. EL is able to distinguish these local influences from those, for example, affecting other geographically distinct MEGs elsewhere. The ability of EL to distinguish local influences is based upon a central theoretical premise that local interactions among resident groups have created unique microevolutionary niches within which local groups have fused, to varying degrees, due to shared environmental exposures, shared cultural patterns, and some degree of gene flow. We have termed this phenomena 'bleeding' as its phenotypic consequences are reminiscent of the transfer of color from one garment to another during washing. Within these local settings, emerging MEGs have frequently been exposed to shared environmental (e.g. pollution) and sociocultural (e.g. occupational) constraints and have developed some level of shared biological and cultural adaptive responses. Many of the MEGs currently encountered in lowland South Carolina (such as though around Galphintown) are a direct product of such historical interactions.

Cultural fluidity in the formation of MEGs

Cultural distinctiveness is a highly transient state. MEGs that come in contact with each other invariably influence each other's behaviors. However, when we limit our analyses to gross 'racial' designations, we often miss key shared cultural behaviors that can dramatically explain regional patterns of chronic diseases. For example, prior to the devastating and disruptive effects of Hurricanes Katrina and Rita in 2005, in the Mississippi Delta region, there were numerous distinct indigenous Native American, African American, European American, and Asian American microethnic groups. If we take, for example, just three distinct local MEGs, the Cajun, Cane River Creole, and Clifton Choctaw peoples, we notice that they all share an important cancer-associated behavior (Kapadia et al. 1978), the regionally ubiquitous dietary consumption of sassafras (*Sassafras albidum*). Sassafras was originally a Choctaw domesticated used for medicinal purposes. Its use spread to the Acadia French Canadian (Cajun) refugees from Nova Scotia (who settled in the bayous and natural levees of southern Louisiana in the 1700s) and among local African American Creoles

Table II. Some examples of US microethnic groups.

Name of microethnic group	Current geographical locations (states marked in bold)	Additional comments
Adamstown Indians	King William, Virginia	Also known as the Upper Mattaponi People
Brandywines	Southeast Prince George's County and northeast Charles County near the Patuxent River, Maryland	Piscataway linguistic retentions in English (see Louden 2000)
Brass Ankles	Coastal Dorchester, Colleton, Berkeley, Orangeburg, and Charleston also Etiwan Island, South Carolina	Found in swampy coastal areas, sandy foothills of the Piedmont, and on the coastal plain. Possible combination of Cusabo, Yemassee Native American Indians and local Europeans and African Americans (see Berry 1945). Closely related to the Brass Ankles (see Estabrook and McDougle 1926)
Buckheads	Bamberg, South Carolina	(see Berry 1960; Kessler and Ball 2001)
Cajans	Mobile and Washington Counties in Alabama	Also referred to as: 'Bald Knobbers' and 'Ozarkers' in Missouri, or 'Sang Diggers' in Virginia (see Buck 1925; Twyman 1971; Corwin 1999)
Carmel Indians	Highland County, Ohio and Mcgoffin County, Kentucky	Combination of Choctaw, Chatot Native American Indians, Creoles, and Africans
Clay Eaters	Central Piedmont sections of Georgia and South Carolina	High incidence of diabetes (LA); unique hemoglobin variant (AL) (see Bowman et al. 1967; Stuart et al. 1994; Gordon et al. 1999)
Clifton Choctaw	Rapides Parish, Louisiana	
Coushatta	Alabama and Louisiana	
Cros	Marlboro, Dillon, Marion, and Horry counties, South Carolina	
Free Moors	Various locations in South Carolina	(see Dunlap and Weslager 1947)
Guineas	Barbour and Taylor counties in West Virginia	Sometimes called Maileys (see Dunlap and Weslager 1947)
Gullah and Geechee	Coastal and Lowland areas of South Carolina and Georgia , including the Sea Islands	Historical range included Coastal and Lowland North Carolina to north coastal Florida (see Argyropoulos et al. 1998; Parra et al. 2001; McLean et al. 2003, 2005; Spruill 2004; Hamilton 2007)
Houma	Terrebonne and Lafourche Parishes southwest of New Orleans, Louisiana (near Montagne, Golden Meadow, and Dulac-Grand Caillou)	Largest Native American Indian group in Louisiana; severely impacted recently by Hurricanes Katrina and Rita (see Coughlin 1998)

(continued)

Table II. Continued.

Name of microethnic group	Current geographical locations (states marked in bold)	Additional comments
Issues	Amherst and Rockbridge counties, Virginia	Originally from hill country of Virginia
Jackson Whites	Ramapo Mts, Orange and Rockland counties in New York , Bergen, Morris, and Passaic counties, New Jersey	Combination of Lenape, Mensee, Tuscarora, Hessian, English, Dutch, and Hackensack River Valley African Americans; sometimes called the 'blue-eyed Blacks' (see Cohen 1974)
Lumbee (Croats, Croatans)	Robeson County, North Carolina (near Lumberton) also Bladen, Columbus, Cumberland, Macon, Hoke, and Sampson counties, Halifax county, Virginia , Marlboro, Dillon, Marion, Horry counties, South Carolina	Largest group of Native American Indians east of Mississippi River; originally lived in swamplands. Mixture of not only Croatan and Cheraw but also Pamlico, Hatteras, and Catawba Indians (see Swanton 1934; Dial 1975)
Marlboro Blues	Chesterfield, South Carolina	Closely related to the Brass Anles (see Dunlap and Weslager 1947)
Melungeons	Hancock county, Tennessee (Newman's Ridge, Clinch Mountain, Copper Ridge, Cumberland Range) also Cocke, Nashville, Franklin, Grundy, Chattanooga, Hawkins, Knoxville, Marion, Meigs, Morgan, Overton, and other counties in Tennessee . In Virginia known as Ramps; found in Giles, Lee, Russell, Scott, Washington, and Wise counties, Virginia . Also found in Kentucky	Linguistic ties with Turkish. Admixture with Saponi and other local Native American Indians, local European Americans, and local African Americans (see Price 1951; Ivey 1977; Kennedy and Kennedy 1994; Winkler 2004)
Mestees	South Carolina	(see Nassau et al. 1994)
Nanticoke-Moors	Sussex county, Delaware (around Millsboro) and Kent county, Delaware (Chesterwold) and Cumberland county, New Jersey (Bridgeton)	(see Price 1953; Blakey 2001)
Oklahoma Choctaw	Throughout Oklahoma	Increased susceptibility to systemic sclerosis, scleroderma, recent genetic bottleneck (Henderson et al. 2002; Herrick and Worthington 2002; Zhou et al. 2002, 2003; Mayes 2003; Wang et al. 2004)
Red Bones	Richland, South Carolina also in Calcasieu, Rapides, Beauregard, Vernon and Allen parishes, Louisiana	May be originally from South Carolina (based on surname analysis)
Red Legs	Orangeburg, South Carolina	

(continued)

Table II. Continued.

Name of microethnic group	Current geographical locations (states marked in bold)	Additional comments
Sabines	Terrebonne and Lafourche Parish, Louisiana	Live near Sabine River; combination of Houma, French, and African
Sandhillers	Richland County, South Carolina	Maybe connected to the Croatans
Skeetertown Indians	Virginia	(see Berry 1945; Tindall 2003)
Turks	Sumter County, South Carolina	Slavic group from Lusatia region of Eastern Germany (see Fliess 1988, 1991)
Wends	Throughout Central Texas	
Wesorts	Charles and Prince Georges counties, Maryland	Descended from local free-African Americans with Chesapeake Bay Native American Indians and European Americans (see Gilbert 1946; Beale 1972; Pollitzer 1972)

descended from gene flow between African and European Francophone settlers before 1803. In these MEGs, dried ground sassafras leaves are used as gumbo filé for flavoring and thickening sauces and soups as well as spring tonics and medicinal teas. Gumbo filé and other sassafras products contribute important cross-cultural ingredients in Cajun, Creole, Choctaw, and other Mississippi Delta cuisines. The plant contains high levels of safrole (Kamdem and Gage 1995; Carlson and Thompson 1997). Safrole, a hepatotoxin, is a methylenedioxy phenyl compound, inhibits human cytochrome P450 enzymes (Ueng et al. 2005) and is considered carcinogenic (Hilker 1981). Recently, the USDA officially classified sassafras as carcinogenic in certain individuals. Its use is associated with increased susceptibilities to liver (Miller and Miller 1983), esophageal, and pancreatic cancers. In fact, the incidence of pancreatic cancer is unusually high in the Mississippi Delta region (Blot et al. 1978; Falk et al. 1988), and this shared regional behavior pattern may play a role. If researchers relied on traditional racial groupings rather than the EL-generated MEGs, the potential contribution of an important shared cultural behavior, the regional consumption of sassafras products, would have been missed in efforts to understand a regionally important chronic disease disparity. With EL, sassafras usage 'links' MEGs from three different 'racial' groups, making sassafras consumption more predictive of pancreatic and liver cancer prevalence than 'race' *per se*.

Discussion

The old approach: traditional racial models

The traditional race-based models used in epidemiology, public health, and biomedicine rarely define these groups. In such classic race-based studies, this part of the research design (i.e. the precise definition of the population studied) is glossed over to get to the 'evidence-based' data. And yet, as scientists, *who* we study is as important as *what* we study and *why*. Disease risk generalizations made about any one of these 'race-identified groups'

Table III. The traditional approach to depicting health disparities in chronic disease morbidity and mortality between ill-defined macroethnic (racial) groups. The data on Native American Indians have been recently contested as they are based on a set of non-representative groups.

Racial group	Cancer incidence and death rates (number of new cases each year) per 100 000 individuals (1992–1999)		
	Both sexes	Males	Females
African-American	526.6 (267.3)	703.6 (369.0)	404.8 (204.5)
European-American	480.4 (205.1)	568.2 (258.1)	424.4 (171.2)
Asian/Pacific Islander	348.6 (128.6)	408.9 (160.6)	306.5 (104.4)
Hispanic/Latino	329.6 (129.2)	393.1 (163.7)	290.5 (105.7)
American Indians/Alaska Native	244.6 (128.6)	277.7 (154.5)	224.2 (104.4)

Modified from National Cancer Institute data presented at: http://newscenter.cancer.gov/BenchMarks/archives/2002_05/related_article.html

may not be relevant for members of the same so-called 'racial cohort' located some geographical distance away. This mindset is tolerated because new alternative models have not been developed to provide a sustainable alternative to the reigning race paradigm. This traditional paradigm would have us believe that there is no significant lineage coalescence between macroethnic (racial) groups, and that variation *within* a 'racial group' is less than variation *between* such groups. This is part of the fallacy of racial thinking that continues to pervade the human sciences (Keita and Kittles 1997) and is reflected in the traditional macroethnic or racial clustering of chronic disease morbidity and mortality. Table III presents the typical presentation of cancer mortality and morbidity at the traditional 'race' level of analysis by gender.

Table III is misleading because it aggregates highly diverse macroethnic groups under simple racial rubrics. For example, the category 'Hispanic' is problematic as a separate 'race', since 'Hispanic' denotes a linguistic status rather than a specific pattern of biological ancestry, genetic background, or even exposure to a specific set of environmental filters. EL suggests that classic 'racial' terms (e.g. those used in Table III) are of limited value in addressing the health disparities in cancer (and in other chronic diseases) because conceptually, the traditional approach is insensitive to critical variables associated with within 'race' variability. Many of these critical variables are included in Table I. The traditional race approach ignores the various MEGs that make up and at times overlap various macroethnic 'races' (as presented in Table II) and the dynamic reality of biological lineage coalescence between macroethnic and microethnic groups in the formation of many MEGs.

EL allows researchers to 'unpack' within-group 'racial' variation in order to recognize local microethnic patterns of contemporary population substructure. As we deconstruct these traditional racial aggregates, we are better able to predict regional chronic disease risks for disorders with longstanding health disparities such as hypertension in the Carolina Coast region, diabetes in the southwestern US, pancreatic cancer in the Mississippi Delta region, and prostate and breast cancers in the Chesapeake Bay region (see Jackson 2003, 2006, 2007).

Figures 3 and 4 demonstrate that layering of demographic and clinical data can more accurately reflect some of the complex interactions between local MEGs and chronic disease and may begin to clarify some of the regional nuances in disease etiology and epidemiology. For example, the focus of non-Hodgkin's lymphoma among New Hampshire

and Vermont males suggests that EL of this region could possibly identify the specific local MEGs among the English- and French-descended residents where this form of lymphatic cancer is concentrated. These analyses could then more rapidly identify the specific relevant biological and cultural lineages involved as well as relevant environmental exposures and could lead to a better understanding (and control) of the underlying contributors to the disease etiology.

Rationale for the new EL approach

As evidenced by the internal variability uncovered by the application of EL in just a few US geographical regions (Table II), some regional MEG clusters have deep historical roots while others are of more recent construction. EL allows researchers to access this substructure and integrate it into evaluations of regional human biodiversity and health disparities. American residential and mating patterns have not produced a genetically homogeneous nation, the metaphoric melting pot. Rather, geographical, cultural, political, social, and economic forces have contributed to a populace that remains regionally heterogeneous, rather like a poorly tossed salad. This heterogeneous distribution is characterized by different combinations of MEGs across geographical space and different degrees of MEG persistence through time. The descendants of the various founding groups that have contributed to the contemporary American gene pool are not uniformly distributed across space nor equally represented through time. By identifying the founding groups and evaluating their local descendants within a broad genetic and non-genetic context, EL provides a backdrop for evaluating the contemporary population substructure at both biological and cultural levels, holding geography constant.

It has been projected that by the year 2015 a major demographic shift will have occurred in the US. By that time, only 66% of US citizens will be classified (using the traditional 'racial' model) as European American or 'White', 13% as African American or 'Black', 15% as Hispanic or 'Latino', 5% as Asian and Pacific Islander, and 1% as American Indian, Inuit, or Aleut or 'Native American'. In terms of resolving chronic disease risk and health disparities, these population changes will continue to be best evaluated at the level of MEGs. This is because the peopling of America continues to be a process of multiple layering, group on top of group, in a colonization process that started with the earliest Native American residents and continues to be subsequently augmented with specific Europeans, Africans, Asians, and others from particular parts of the world. Today, across the American landscape, there are important regional variations in local ancestral origins that, in turn, set the stage for regional differences in potential genetic susceptibilities and non-genetic vulnerabilities.

The impact of this inherent regional biodiversity on disease expression has not been sufficiently quantified, perhaps because the necessary conceptual and technological tools have not been available until recently. EL represents a major step in computational anthropology. Computer technology permits the quantification of an expanded integrative perspective of human biodiversity. The ease of identification and characterization of MEGs can be incorporated into the research design of studies in epidemiology, public health, and medicine as well as the behavioral and social sciences.

These discoveries make racial designations based on a limited number of biological and cultural traits scientifically problematic. The use of 'race' as the primary level of analysis in epidemiology, public health, and biomedicine is actually a handicap; local distinctions may be glossed over and key genetic and environmental markers obscured. Truly in a biological and cultural sense, we remain 'racially' labile, usually having many identities, and

being most consistently grounded by geography. The more we understand about the adaptive fine-tuning of human biology to the abiotic, biotic, and sociocultural environmental filters that are able to transform the genetic message into the expressed genotype or the phenotype, the more difficult it is to match biological diversity with the traditional 19th and 20th century sociological categories of human variation that have guided so many modern epidemiological, public health, and biomedical investigations. We can do better in the 21st century; EL is an effort to provide that research option.

Strengths of EL as a research model

Ethnogenetic layering can benefit disease research in many ways. These include:

- (1) *Finding the 'holes' in macroethnic or race-based disease research:* the lack of adequate disease-related research on many east coast Native American Indian groups and local southeastern African American and European American groups is often rooted in the presumption of 'racial' homogeneity within groups and significant between-race heterogeneity. This assumption often leads to the practice of substituting the results of one local 'racial' group for another similarly colored in a different locale. However, this is often problematic. In fact, even comparing two North Carolina Native American groups, Messler and colleagues (1999) have found different responses between Eastern Band Cherokee (in the western part of the state) versus Lumbee (in the central and eastern part of the state) to a cervical cancer intervention program (see Messer et al. 1999). The key lesson of EL is that each group has to be evaluated within its own unique local historical, socio-economic, ethnographic, and genetic context. This is also true with such highly diverse groups as Hispanic/Latino Americans and Asian Americans.
- (2) *Distinguishing 'expected' from 'unexpected' disparities, signaling potential local influences.* Prostate cancer incidence and mortality rates are known to vary worldwide. In the USA, prostate cancer is the most common malignancy affecting men and is the second-leading cause of cancer death (Crawford 2003). While the risk of developing prostate cancer is usually associated with advancing age, African American ethnicity, and a positive family history, these generalizations do not hold everywhere. EL can uncover and further define local conditions that might influence these specific health disparities. For example in some areas of the Mississippi Delta region, measures of prostate cancer among African American men are equal to that among European men (Witte et al. 1999; Martin et al. 2003). This is not the usual pattern reported for elsewhere in the USA.
- (3) *Providing an integrated, microscopic analysis of regional chronic disease patterns based on the 'as lived' experiences of local groups:* for example, in a recent study of lifestyle and prostate cancer in older South Carolina men (Sanderson et al. 2004), researchers found that occupation (production, transportation, and material moving versus military) and diet (lycopene consumption) had more of an influence on prostate cancer risk than race per se. Once a cancer diagnosis has been made, it is clear that understanding the 'as lived' experience of patients is essential in refining disease prognosis. Hamilton and Sandelowski (2004) reported that among southeastern African Americans with cancer, many of the most important sources of social support were types not even emphasized in the current literature. They emphasize the need to 'reconceptualize social support for African Americans with cancer and to refine instruments to include these different manifestations of social support.'

- (4) *Improving our insights on the influence of local cultural patterns on chronic disease epidemiology:* Quantifying the biological impact of specific cultural behaviors is a promising although currently underdeveloped area within biological anthropology. EL can accelerate the discovery of cultural patterns of potential biological significance, the extent of their distributions, and the likely impact on specific disease processes. For example, a cultural behavior observed in North Carolina is the use of smokeless tobacco by eastern band Cherokee women (Spangler et al. 2001). These researchers suggest that this practice may play a role in the high incidence of breast cancer in this region. From an EL perspective, since there is a high incidence of breast cancer among Native Americans in general (Paltoo and Chu 2004), a key question might be whether the use of smokeless tobacco in these eastern band women is a limited microethnic cultural distinction or is found among all regional Cherokee-influenced biological lineages, among all local Algonquin speakers, among all Carolina Native Americans, etc.? In other words, are there similar comparative MEGs without this specific cultural behavior that could be identified using EL to determine the impact of this cultural behavior on breast cancer etiology?
- (5) *Identifying the likelihood for shared marker genes among diverse local microethnic groups:* EL is a first step in the search, in affiliated microethnic groups, for shared chronic disease susceptibility and resistance genes since the process of EL relies, in part, on the historical reconstruction of microethnic identities. The longer and more profound the history of genetic exchange between groups, the more likely marker genes will be shared between linked microethnic groups. The same is true for cultural behaviors and environmental exposures with potential importance for chronic disease health disparities. Figure 5 depicts this 'bleeding' of traits from one MEG to another in the process of MEG formation.
- (6) *Identifying the magnitude of real 'difference' between adjacent MEGs.* On the other hand, MEGs with little gene flow and/or cultural exchange between them might be expected to share few specifically geographically restricted alleles. An example of this is provided by Quinones and colleagues (1999) in their comparison of differences in susceptibility genes to lung cancer (specifically, cytochromes P-450 CYP2E1, CYP1A1 and glutathione S-transferase Mu (GSTM1)) between European French and Spanish European-Native American Chilean populations.

Limitations of EL

There are a number of important limitations to EL. The major caveats are listed below:

- (1) *MEGs are dynamic units; therefore, material on MEGs needs to be periodically reevaluated and updated.* A MEG may experience an influx of new genes, for example, through migration and this would need to be documented and factored into the genetic and other databases.
- (2) *Cultural behaviors are very fluid. New behaviors can rapidly spread through MEGs and can cross MEGs within a generation.* These can lead to generational discrepancies in health risk.
- (3) *Diversity exists within MEGs but is only indirectly measured by EL.* MEGs represent the local group level of analysis; the data generated on an MEG should not be uncritically applied to every member of an MEG. MEG phenotypic measures represent a series of means, but variation (e.g. standard deviations of these means)

could diverge importantly. The individual level of analysis is superior in assessing individual disease risk, for example.

- (4) *MEG cultural and biological histories may not be entirely synchronous.* EL should reconcile discontinuities between databases but since EL is based on evidence-based data, a priority system is initiated with respect to verification of databases. Facts must override folklore. For example, many MEGs who have resided at a specific location for an extended number of generations may develop a folk history that misrepresents their original migrations from elsewhere. However, their genetic background data may suggest prior ancestral residence on a different continent or in a different ecosystem.
- (5) *The temporal limitations of EL must be kept in mind.* Although geography is held constant (more or less), other aspects of the EL maps can change. For example: in Ward 9 in New Orleans, the geography was shifted by Hurricanes Katrina and Rita and their aftermath (e.g. the compromising of the levees). When geography is not constant, this has to be taken into consideration, as all other variables may also shift radically as well (within a very short period of time).
- (6) *The impact of an MEG in a region is not always proportional to their size in the population.* EL could inadvertently overlook culturally important but numerically minor groups in regional searches such as the 19th Century Chinese immigrants to the Mississippi Delta. For example, consider the huge culinary and architectural influence of the Italians although their actual numbers in the US are proportionately small.

Applications of EL

We have applied EL to identify founding populations in the Chesapeake Bay region who may be responsible, in part, for the high frequency of aggressive breast cancer observed among local African American women and women from specific groups from the Bight of Bonny, West Africa (Jackson 2007). We have applied EL to identify founding populations of the Carolina Coast region who may have disproportionately contributed to the high rates of hypertension and stroke observed in this area (Jackson and Inasi n.d.). This application of EL identifies certain ancestral groups in West Central Africa who may have disproportionately contributed salt-sensitive genes to certain Lowland Carolina MEGs. We have applied EL to identify the founding populations in the Mississippi Delta region and begun to identify cultural behaviors that contribute to the high rates of pancreatic and liver cancer in this region (Jackson et al. n.d.).

Summary

EL is a new non-typological tool used to identify geographical regional substructure in macroethnic groups. Population substructure is recognized to help identify and assess the risks underlying the health disparities observed in a specific geographical area. By focusing on groups that live in and/or have ancestral origins in specific regions, we have developed a strategy to collect and analyze geographical patterns of biological lineage data and microethnic affinity within an ethnohistorical framework. Regional frequencies of significant biocultural factors correlated with health outcomes are identified to develop a predictive model for assessing group susceptibilities in health disparities research.

Acknowledgements

Support for this research has come from a University of Maryland General Research Board Summer Award to the author and from Oasis Scientific Research Grant no. 401A-05 to the Genomic Models Research Group, Department of Anthropology, University of Maryland. Mr Timothy Goddard is appreciated for his important contribution to the GIS-related sections of the manuscript.

References

- ACOG Committee on Genetics. 2005. ACOG committee opinion. Number 318, October 2005. Screening for Tay-Sachs disease. *Obstet Gynecol* 106:L893-L894.
- Argyropoulos G, Brown AM, Willi SM, Zhu J, He Y, Reitman M, Gevaio SM, Spruill I, Garvey WT. 1998. Effects of mutations in the human uncoupling protein 3 gene on the respiratory quotient and fat oxidation in severe obesity and type 2 diabetes. *J Clin Invest* 102:1345-1351.
- Armour JAL, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, Bertranpetit J, Paabo S, Jeffreys AJ. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet* 13:154-160.
- Beale C. 1972. An overview of the phenomenon of mixed race isolates in the United States. *American Anthropologist* 74:704-710.
- Berry B. 1945. The Mestizos of South Carolina. *Am J Soc* 51:34-41.
- Blakey ML. 2001. Bioarchaeology of the African diaspora in the Americas: Its origins and scope. *Annual Review of Anthropology* 30: 387-422.
- Blot WJ, Fraumeni JF Jr, Stone BJ. 1978. Geographic correlates of pancreas cancer in the United States. *Cancer* 42:373-380.
- Bowman BH, Barnett DR, Hite R. 1967. Hemoglobin G-Coushatta: A beta variant with a delta-like substitution. *Biochem Biophys Res Commun* 26:466-470.
- Buck PH. 1925. The Poor Whites of the Ante-Bellum South. *Amer Hist Rev* 31:41-54.
- Carlson M, Thompson RD. 1997. Liquid chromatographic determination of saffrole in sassafras-derived herbal products. *J AOAC Int* 80:1023-1028.
- Census Bureau 2000. www.census.gov/main/www/cen2000.html (Accessed October 15, 2007).
- Cohen DS. 1974. The Romapo maintain people. Rutgers University Press, USA. p. 197.
- Corwin T. 1999. The Clay Eaters. *Augusta Chronicle*, 18 March 1999.
- Crawford ED. 2003. Epidemiology of prostate cancer. *Urology* 62:(6 Suppl 1)3-12.
- Coughlin SS. 1998. Implementing breast and cervical cancer prevention programs among the Houma Indians of southern Louisiana: Cultural and ethical considerations. *Health Care Poor Underserved* 9:(February)30-41.
- Dial AL, Eliades DK. 1979. The only land I know: A history of the Lumbee Indians. Indian Historian Press, San Francisco. USA.
- Dunlap AR, Weslager CA. 1947. Trends in the naming of tri-racial mixed-blood groups in the eastern United States. *Am Speech* (April) 22:81-87.
- Edward T, Price ET. 1953. A geographic analysis of white-Indian-Negro racial mixtures in the Eastern United States. *Ann Am Assoc of Geographers*, Vol. 43, No. 2, p. 136.
- Estabrook AH, McDougale IE. 1926. *Mongrel Virginians*. Baltimore: Williams & Wilkins.
- Ely B, Wilson JL, Jackson F, Jackson BA. 2006. African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biol* (12 October) 4:34.
- Falk RT, Pickle LW, Fonthan ET, Correa P, Fraumeni JF Jr. 1988. Life-style risk factors for pancreatic cancer in Louisiana: a case-control study. *Am J Epidemiol* 128:324-336.
- Fliess KH. 1988. Fertility, nuptiality, and family limitation among the Wends of Serbin, Texas, 1845-1920. *Fam Hist* 13:251-263.
- Fliess KH. 1991. Mortality transition among the Wends of Serbin, Texas, 1854-1884: changes in pattern of death from parochial records. *Soc Biol* 38:266-276.
- Garg RK, Filozof E, Etheredge G, Maney E. 1998. Health behavior in a Cajun population. *J Cult Divers* 5:89-93.
- Gilbert WH. 1946. Memorandum concerning the characteristics of the larger mixed-blood racial islands of the eastern United States. *Social Forces* 21(4):438-447.

- Goodman AH. 1997. Bred in the bone. *The Sciences* (March/April):20-25.
- Gordon DL, Cook CB, Scheer WD, Oalmann M, Boudreau DA, Borne D, Gatchair-Rose A, Weems AH, Turner MN, Robinson EM, et al. 1999. Diabetes and obesity in the Louisiana Coushatta Indians. *Ethn Dis* 9:48-58.
- Hamilton K. 2007. The taste of the sun: Okra saip in the Greechec tradition. *Callaloo* 30:75-86.
- Hamilton JB, Sandelowski M. 2004. Types of social support in African Americans with cancer. *Oncol Nurs Forum* 31:792-800.
- Henderson JN, Crook R, Crook J, Hardy J, Onstead L, Carson-Henderson L, Mayer P, Parker B, Petersen R, Williams B. 2002. Apolipoprotein E4 and tau allele frequencies among Choctaw Indians. *Neurosci Lett* 324:77-79.
- Herrick AL, Worthington J. 2002. Genetic epidemiology: systemic sclerosis. *Arthritis Res* 4:165-168.
- Hilker DM. 1981. Carcinogens occurring naturally in food. *Nutr Cancer* 2:217-223.
- Horai S, Hayasaki K, Kondo R, Tsugane K, Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* 92:532-536.
- Ivey S. 1977. A scribed ethnicity and ethnic display event: The Melungeons of Hancock county. *Tennessee Western Folklore*. 36:88-108.
- Jackson LC, Oseguera M, Medrano S, Kim YL. 1988. Carbamylation of hemoglobin *in vivo* with chronic sublethal dietary cyanide: Implications for hemoglobin S. *Biochem Med Metab Biol* 39:64-68.
- Jackson FLC. 1990. Two evolutionary models for the interactions of dietary cyanogens, hemoglobin S, and falciparum malaria. *Am J Hum Biol* 2:521-532.
- Jackson FL. 1992. Race and ethnicity as biological constructs. *Ethn Dis* 2:120-125.
- Jackson FL. 2003. Ethnogenetic layering: A novel approach to determining environmental health risk potentials among children from three US Regions. *J Child Health* 1:369-386.
- Jackson FL. 2004. Human genetic variation and health: ethnogenetic layering as a way of detecting relevant population substructuring. *Br Med Bull* 69:215-235.
- Jackson F. 2006. Illuminating cancer health disparities using ethnogenetic layering (EL) and phenotype segregation network analysis (PSNA). *J Cancer Educ* 21:69-79.
- Jackson FLC. 2008. Ancestral links of Chesapeake Bay region African Americans to specific Bight of Bonny (West Africa) microethnic groups and increased frequency of aggressive breast cancer in both regions. *Am J Hum Biol*, 20:165-173.
- Jackson FLC, Inasi J. Unpublished paper 2007. Explaining the Stoke belt: investigating the founding ancestral links to West Central Africa and the increased frequency of hypertension and stroke in the Carolina Coast region.
- Jackson FLC et al. Unpublished paper 2007. Opportunities for gene flow and biological lineage coalescence among the founding populations of the Mississippi Delta region as revealed through ethnogenetic layering.
- Kamdem DP, Gage DA. 1995. Chemical composition of essential oil from the root bark of *Sassafras albidum*. *Planta Med* 61:574-575.
- Kapadia GJ, Chung EB, Ghosh B, Shukla YN, Basak SP, Morton JF, Pradhan SN. 1978. Carcinogenicity of some folk medicinal herbs in rats. *J Natl Cancer Inst* 60:683-686.
- Keita SOY, Kittles RA. 1997. The persistence of racial thinking and the myth of racial divergence. *Am Anthropol* 99:523-544.
- Keita SOY, Kittles RA, Royal CD, Bonney GE, Furbert-Harris P, Dunston GM, Rotimi CN. 2004. Conceptualizing human variation. *Nat Genet* 36:(11 Suppl)S17-S20.
- Kennedy B, Kennedy RV. 1994. *The Melungeons: The resurrection of a proud people*. Macon, GA: Mercer University Press.
- Kessler JS, Ball DB. 2007. *North from the maintains: A folk history of the Carmel Melungeon Settlement*. Highland Country, Ohio. Mercer University Press, USA.
- Louden ML. 2000. African-Americans and minority language maintenance in the United States. *The J Negro History* 85:223-240.
- Marks J. 1994. Racial categories and cultural constructs masquerading as biology. *Natural History* (December) 32-35.
- Marks J. 1995. *Human biodiversity: Genes, race, and history*. New York: Aldine Press.
- Martin B, Cheli C, Pollard S, Sullivan J, Goodman J, Kokatnur M, Ward M, Martin L, Mercante D, Rayford W. 2003. Similar age-specific PSA, complexed PSA, and percent cPSA levels among African American and white men of southern Louisiana. Prostate specific antigen. *Urology* 61:375-379.
- Mayes MD. 2003. Scleroderma epidemiology. *Rheum Dis Clin North Am* 29:239-254.
- Messler L, Steckler A, Dignan M. 1999. Early detection of cervical cancer among Native American women: A qualitative supplement to a quantitative study. *Health Educ Behav* 26:547-562.

- McDowell GA, Mules EH, Fabacher P, Shapiro E, Blitzer MG. 1992. The presence of two different infantile Tay-Sachs disease mutations in a Cajun population. *Am J Hum Genet* 51:1071-1077.
- McLean DC Jr, Spruill I, Gevaso S, Morrison EY, Bernard OS, Argyropoulos G, Garvey WT. 2003. Three novel mtDNA restriction site polymorphisms allow exploration of population affinities of African Americans. *Hum Biol* 75:147-161.
- McLean DC Jr, Spruill I, Argyropoulos G, Page GP, Shriver MD, Garvey WT. 2005. Mitochondrial DNA (mtDNA) haplotypes reveal maternal population genetic affinities of Sea Island Gullah-speaking African Americans. *Am J Phys Anthropol* 127:427-438.
- Miller JA, Miller EC. 1983. The metabolic activation and nucleic acid adducts of naturally-occurring carcinogens: Recent results with ethyl carbamate and the spice flavors safrole and estragole. *Br J Cancer* 48:1-15.
- Minges P. 2001. Beneath the underdog: Race, religion, and the trail of tears. *Am Indian Q* 25:453-479.
- Nassau ME, Price ET, Gilbert WH. 1994. Melungeons and other Mestee groups. <http://www.multiracial.com/readers/nassau.html> (Accessed October 26, 2007).
- Oetting JB, Garrity AW. 2006. Variation within dialects: A case of Cajun/Creole influence within child SAAE and SWE. *J Speech Lang Hear Res* 49:16-26.
- Oriol MD. 1995. Cajun traditions and their impact on health care. *J Cult Divers* 2:27-30.
- Paltoo DN, Chu KC. 2004. Patterns in cancer incidence among American Indians/Alaska Natives, United States, 1992-1999. *Public Health Rep* 119:443-451.
- Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, et al. 2001. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 114:18-29.
- Pollitzer WS. 1972. The physical anthropology and genetics of marginal people of the southeastern United States. *American Anthropologist* 74:719-734.
- Pollitzer WS, Namboodiri KK, Coleman WH, Finley WH, Leyshon WC, Jennings GC, Brown WH. 1977. The Cajuns of Southern Alabama: Morphology and serology. *Am J Phys Anthropol* 47:1-6.
- Price ET. 1951. The Melungeons: A mixed-blood strain of the southern Appalachians. *Geogr Rev* XLI:256-271.
- Quinones L, Berthou F, Varela N, Simon B, Gil L, Lucas D. 1999. Ethnic susceptibility to lung cancer: Differences in CYP2E1, CYP1A1 and GSTM1 genetic polymorphisms between French Caucasian and Chilean populations. *Cancer Lett* 141:167-171.
- Sanderson M, Coker AL, Logan P, Zheng W, Fadden MK. 2004. Lifestyle and prostate cancer among older African-American and Caucasian men in South Carolina. *Cancer Causes Control* 15:647-655.
- Scacheri PC, Garcia C, Hebert R, Hoffman EP. 1999. Unique PABP2 mutations in 'Cajuns' suggest multiple founders of oculopharyngeal muscular dystrophy in populations with French ancestry. *Am J Med Genet* 86:477-481.
- Spangler JG, Michielutte R, Bell RA, Dignan MB. 2001. Association between smokeless tobacco use and breast cancer among Native-American women in North Carolina. *Ethn Dis* 11:36-43.
- Spruill I. 2004. Project Sugar: A recruitment model for successful African-American participation in health research. *J Natl Black Nurses Assoc* 15:48-53.
- Stuart CA, Smith MM, Gilkison CR, Shaheb S, Stahn RM. 1994. *Acanthos Nigricans* among Native Americans: an indicator of high diabetes risk. *Am J Public Health* 84:1839-1842.
- Swanton JR. 1934. Probable Identity of the Croatan Indians, IN Senate Reports, 73rd Congress, 2nd Session, Calendar No. 229, Report No. 204, Washington, p. 5.
- Tindall GB. 2003. *South Carolina Negroes: 1877-1900* University of South Carolina Press.
- Twyman RW. 1971. The clay eater: a new look at an old Southern enigma. *J Southern Hist* 37:439-448.
- Ueng YF, Hsieh CH, Don MJ. 2005. Inhibition of human cytochrome P450 enzymes by the natural hepatotoxin safrole. *Food Chem Toxicol* 43:707-712.
- Waldmeier VP. 2002. Laissez les bon temps rouler (let the good times roll): Can embedded Cajun cultural health beliefs be changed through traditional health education? *Cult Divers* 9:113-117.
- Wang N, Zhou X, Tan FK, Foster MW, Arnett FC, Chakraborty R. 2004. Genetic signatures of pre-expansion bottleneck in the Choctaw population of Oklahoma. *Am J Phys Anthropol* 124:373-379.
- Witte MN, Kattan MW, Albani J, Sharp DS, Estham JA, Morton Jr. 1999. Race is not an independent predictor of positive surgical margins after radical prostatectomy. *Urology* 54:869-874.
- Winkler W. 2004. *Walking toward the sunset: The Melungeons of Appalachia*. Macon, GA: Mercer University Press.

- Zhou X, Tan FK, Wang N, Xiong M, Maghidman S, Reveille JD, Milewicz DM, Chakraborty R, Arnett FC. 2003. Genome-wide association study for regions of systemic sclerosis susceptibility in a Choctaw Indian population with high disease prevalence. *Arthritis Rheum* 48:2585-2592.
- Zhou X, Tan FK, Reveille JD, Wallis D, Milewicz DM, Ahn C, Wang A, Arnett FC. 2002. Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. *Arthritis Rheum* 46:2990-2999.
- Zlotogora J. 1993. Is the presence of two different Tay-Sachs disease mutations in a Cajun population an unexpected observation? *Am J Hum Genet* 51:1071-1077.