Ontologies and Politics of Biogenomic 'Race'

Rasmus Grønfeldt Winther and Jonathan Michael Kaplan

Abstract: All eyes are turned towards genomic data and models as the source of knowledge about whether human races exist or not. Will genomic science make the final decision about whether racial realism (e.g. racial population naturalism) or anti-realism (e.g. racial scepticism) is correct? We think not. The results of even our best and most impressive genomic technologies underdetermine whether biogenomic races exist, or not. First, different sub-disciplines of biology interested in population structure employ distinct concepts, aims, measures and models, producing cross-cutting categorisations of population subdivisions rather than a single, universal biogenomic concept of 'race.' Second, within each sub-discipline (e.g. phylogenetics, conservation biology), genomic results are consistent with, and map multiply to, racial realism and anti-realism. Indeed, racial ontologies are constructed conventionally, rather than discovered. We thus defend a constructivist conventionalism about biogenomic racial ontology. Choices and conventions must always be made in identifying particular kinds of groups. Political agendas, social programmes, and moral questions premised on the existence of naturalistic race should accept that no scientifically grounded racial ontology is forthcoming, and adjust presumptions, practices and projects accordingly.

Keywords: biogenomic race, genomics, phylogenetics, racial ontology, racial realism

Introduction

A pervasive and not unreasonable belief about science is that, given enough time, it provides reliable knowledge about various aspects of the natural world. In general, this belief holds true: physics informs us about the nature of fundamental particles, space and time; chemistry provides us with knowledge about how simple and complex molecules work; and the biological sciences describe the variety of species, organs and biochemical molecules, and explain ecological, evolutionary, physiological and developmental mechanisms. But what about the existence of race in *Homo sapiens*? Will the biological sciences, in the fullness of time, explain the existence (or non-existence) of race

Theoria, Issue 136, Vol. 60, No. 3 (September 2013): 54-80
doi:10.3167/th.2013.6013605 • ISSN 0040-5817 (Print) • ISSN 1558-5816 (Online)

in our species? To which kinds of data and models will such sciences appeal in explaining that human races exist (or do not exist)? In our estimation, many biologists, doctors, politicians and political commentators – and the lay public alike – believe that the biological sciences will eventually grant an answer to whether one should accept subdivision in, or continuity of, human populations. The broad belief is that genomics will eventually establish the existence or not of biological races. In fact, many commentators have written as if the biological sciences had already done so and, ironically, in either direction (e.g. Livingstone and Dobzshansky 1962; Lewontin 1972; Edwards 2003; see also Winther 2011; Kaplan and Winther 2012). Today, all eyes are turned towards genomic data and models. If anywhere, *this* will be the science providing us with the answer.

We think not. On the contrary, we argue that even if it were possible to bracket the social and political implications of genomics and focus only on biological facts, human genomic data and models map in multiple ways to racial ontology. In other words, genomic results will never help us decide between accepting or rejecting the existence of race. This is so even when 'race' is taken as a strictly biogenomic concept, referring merely to any biologically legitimate subdivision below the species level. Nor will genomics allow us to adjudicate between different possible types of criteria used to identify races. That is, genomics cannot transparently speak to whether biogenomic races exist, or even what they *would be* should they exist. By investigating different sub-disciplines of the biological sciences interested in population structure (e.g. taxonomy, phylogenetics, conservation biology, and ecology), this strong claim is supported. We show how each employs distinct families of aims and norms of inquiry, using genomics for particular, local research questions. Robust associations of data, concepts and aims, and measures and models are found within each sub-discipline. Even then, one finds what we shall call *subdivision* (or racial realism such as racial populational naturalism, see Andreasen 2000, 2004; James 2011) positions, as well as *continuum* (or racial anti-realism such as racial scepticism – see Zack 2002; James 2011) perspectives, within each sub-discipline. For instance, taxonomists reasonably disagree about whether human populations below the species level are worth identifying, naming and writing papers about (i.e. taxonomic realists about biogenomic race), or whether differences between races of *Homo sapiens* are trivial, uninteresting, and irrelevant (i.e. taxonomic anti-realists about race), and would disagree even if the species in question was not of particular intellectual or moral interest to us. The robust differences between sub-disciplines of the biological sciences are investigated in this paper. Subdivision and continuum views are both reasonable interpretations and inferential consequences of each sub-discipline's use of genomics.

In short, this article examines how prominent sub-disciplines of the biological sciences employ genomic results to establish the existence (or not) of biogenomic human races. The results of even our most impressive genomic

technologies underdetermine the ontology of biogenomic race.[1] In fact, choices and conventions must be made to decide whether a realist or anti-realist gloss on the genomic results is desired. Hence, we call our position constructivist conventionalism about biological race. Note that this position is logically independent of any ontological position on the existence of social races, such as constructivism or error theories (see Mills 1998; Zack 2002; Mallon 2006). That is, both of these are consistent with realism or anti-realism about biogenomic race (e.g. logically, the same categories we construct socially may also turn out to have genomic objectivity). Our concern is with biogenomic rather than social races. Now, if we are right, science simply cannot respond unequivocally to the question of whether biogenomic races exist. Political agendas, social programmes and moral questions premised on the subdivision or continuity of our species should accept that no scientifically grounded racial ontology is forthcoming, and adjust presumptions, practices and projects accordingly.

A qualification imagining two families of parallel universes is worth making. Consider the following two sets of counterfactual situations, each of which would undermine our claim about the underdetermination of racial ontology by genomic science. First, if *Homo sapiens* were much less structured (e.g. smaller range, more admixture, younger species, etc.), the indeterminate field of genomics would strongly suggest a continuum perspective about races.[2] Second, if *Homo sapiens* were much more structured – perhaps in a world more like the Galápagos writ large, with hundreds of islands separated by large distances, with hundreds of thousands of years of evolution – there would very likely be clear sub-species-level structure, and genomic data and models would strongly converge on a subdivision, racial realist position. However, *Homo sapiens* is neither kind of species: it is neither totally substructured (i.e. living in extreme isolates) nor completely unstructured (i.e. thoroughly admixed and hybridised). As further discussed below, ours is a fairly admixed ('globalised') species, and a relatively young one, evolving in a continuous and clinal world, genetically and ecologically.

This article is organised as follows. In Section 2, we characterise four different biological sub-disciplines: taxonomy, phylogenetics, conservation biology, and ecology. The first two are theoretically and historically quite deep in the biological sciences. We describe them both in general and by examining two of their aspects critical for racial ontology: (1) measures and models vis-à-vis genetic variation, and (2) key concepts and aims for how to identify subspecies differences. Section 3 excavates each sub-discipline in detail and shows how genomic results underdetermine racial ontology in each of these fields. In so doing, it explores the role of lumping and lumpers, and splitting and splitters (Zerubavel 1996; Bowker and Star 1999). We argue that genomic results underdetermine racial ontology for two reasons. First, because different parts of the biological sciences focus on different meanings and measures of the genomic data; and, second, because within each sub-discipline, there can

be genuine, reasonable and irreducible disagreement about whether a subdivision or continuum view about race is the appropriate interpretation of human genomics.[3] This section is critical of simply reading realism or anti-realism off of the genomic results. Instead, our positive programme favours a constructivist conventionalism about biogenomic race. We conclude by discussing the implications of our analysis for political and social programmes.

A Plurality of Concepts and Aims, and Measures and Models in the Biological Sciences

This section systematically compares ontologies and ontologising across a few particularly important biological sub-disciplines interested in population structure. While the four sub-disciplines identified cover many of the pertinent biological sub-disciplines that make heavy use of population structure, we are not exhaustive (e.g. we do not here cover certain parts of evolutionary genetics, but see Kaplan and Winther 2012), nor do we here provide further details of the status, history, epistemology and public recognition of the fields, as this would take us too far afield. These sub-disciplines matter because they are crucibles of practice shaping different understandings of racial realities. Indeed, each sub-discipline takes race to be real or *not*, according to its concepts and aims, and measures and models. Furthermore, racial realist or anti-realist interpretations emerge from, and are perfectly consistent with, the genomic results. Section (3) emphasises this point, for each sub-discipline. A complex story must be told about different ways that ontologies are (co-)created and (co-)constructed (Goodman 1978; Hacking 2002, 2007) by distinct, disunified (Galison and Stump 1996; Hacking 1996) sub-disciplines of the biological sciences (Winther 2012, 2013). Put differently, what follows excavates ontological assumptions (Kuhn [1962] 1970) situated in distinct (scientific) discourses (Foucault [1969] 1972).

We now turn to two critical aspects of each sub-discipline: concepts and aims, and measures and models. The first provides an overarching description, the second provides some technical details. A table helps organise the argument.

Very briefly, here is an overarching description of the concepts and aims (column 2) of each sub-discipline:

1. **Taxonomy** defines classes or types of individuals on the basis of diagnostic characters and names these classes, with the presumption of discovering the order of nature. It traces its Western intellectual origin at least to Aristotle.
2. **Phylogenetics** researches evolutionary relationships among classes or types of organisms (e.g. sub-species, species, genera, clades). Phylogenetics uses characters (predicates) acquired via molecular or morphological studies with the explicit goal of reconstructing the tree of life.

Biological Sub-discipline	**Concepts and Aims**	**Measures and Models**
Taxonomy	typology[4] and discover the order of nature	diagnostic characters
Phylogenetics	tree of life and reconstruct evolutionary history	differentiation measures (e.g., Nei's genetic distance)
Conservation Biology	biodiversity and preserve species and ecosystems	diversity measures (e.g., Jost's *D*)
Ecology	distribution and explain adaptation	fitness, function

3. **Conservation biology** examines biodiversity with the goal of preserving species and ecosystems, given limited public and private resources aimed at such purposes.
4. **Ecology** investigates the relations among species and between species and their environments. In particular, characterising the adaptation, number, and distribution of species are important goals of ecology.

Why is each field important to discussions about the status of groups and populations (and races) within biology? Ultimately, it is because each field provides distinct research frames and purposes for subdividing or lumping species. Respectively, each cares about: (1) the order of nature, (2) the origin and ongoing history of species and sub-species, (3) the conservation of endangered species and sub-species, and ecosystems, and (4) the adaptation and diversity of species and sub-species. But why might such scientific results matter when applied to our species, *Homo sapiens*? We return to this question in the conclusion. For now, simply note that the reality (or not) of human races is taken by many (e.g. hereditarians, social liberals, etc.) as directly pertinent to political and social projects of, for instance, biomedicine and education. We disagree doubly. Because of multiple mappings between racial ontology and genomic results, and because of underdetermination of political and social views by racial ontology, genomics neither justifies nor grounds political or social agendas.

Now that the plurality of concepts and aims in these sub-disciplines has been sketched, we turn to measures and models (column 3 of table above; for further detail see, e.g., Kaplan and Winther 2012). Because there are various types of measures (e.g. heterozygosity measures such as the Shannon Entropy measure and Wright's F_{ST} and diversity measures such as Jost's *D*), and modelling strategies of genomic variation (clustering analyses, heterozygosity partitionings and historical reconstructions), and because purposes and conventions of different researchers, research groups and research agendas that shape how measures and models articulate in particular applications vary radically, no ontology about biogenomic race follows automatically from genomic science.

Start by considering the distinction between *diversity* and *differentiation*, pertinent respectively to conservation biology and phylogenetics. Diversity is most naturally thought of as a measure of the heterogeneity of a system. In the case of genetic diversity, for example, it might be a measure of how many different alleles there are in a population, either at some particular locus or on average across all (or some set of) loci. Differentiation, though, is a measure of how different two units are. Two populations that share most of their common alleles are relatively undifferentiated, whereas two that do not share many alleles will be quite highly differentiated. Indeed, if differentiation is interpreted as a measure of genetic distance, a recursive nesting of populations – so useful to phylogenetics – can be performed.

Diversity and differentiation are independent of one another. In order to see this, consider two populations that are both very diverse at some locus – that is, each population has a large number of alleles at the locus or loci in question. Note that these two diverse populations can vary from (i) being entirely differentiated such that they share no alleles in common, to (ii) being undifferentiated such that each population has the same alleles at the same frequencies at each locus of interest. Moreover, two populations with very low diversity (say, because each is fixed for a single allele at the locus of interest) can vary from (i) being entirely differentiated such that each is fixed for different alleles, to (ii) being completely undifferentiated such that both are fixed for the same alleles. And, while it is obvious that two populations which differ with respect to their diversity must be somewhat differentiated (e.g. one population has 2, 10 or 100 alleles at a locus, the other 1 allele), their degree of differentiation can vary from low to complete. Diversity and differentiation are thus distinct measures of genomic variation, and lumping or splitting of populations based on one (e.g. diversity for conservation biology) will not identify the same populations as the lumping or splitting based on another (e.g. genetic distance, via differentiation, for phylogenetics) (Kaplan and Winther 2012).[5]

Moreover, groups identified in different ways also cross-cut. This is true, for example, of groups identified by diagnostic characters, on the one hand, and fitness to particular environments, on the other. In the on-the-ground practice of taxonomy it is common to use diagnostic characters to identify group members in the field and the laboratory. While taxonomy is often closely tied to phylogenetics, they are importantly distinct in both theory and practice. Taxonomy involves standard collection, naming, classification, and preservation practices of biological species. Natural history museums are repositories of 'remnant models,' which are products of taxonomic practices (e.g., Griesemer 1990). Such practices were not, and have not been, significantly affected by evolutionary theory. Indeed, Crombie (1994) distinguishes 'taxonomy' from 'historical derivation' styles of scientific thinking (see Hacking 2002, Winther 2012). Moreover, the existence of pattern cladistics (e.g., Nelson and Platnick 1981), as opposed to process cladistics, is further evidence that it would be a mistake to conflate taxonomy with phylogenetics (e.g., Hull 1988;

Winther 2009). Baum and Donoghue (1995, p. 560) criticize 'character-based' accounts of species in favor of 'history-based' phylogenetic species accounts. Yet, their description of the former is useful for our purposes: 'under 'charater-based' concepts [of species and sub-species], an organism is a member of a given species if and only if it possesses some character (i.e. an observable organismal attribute) or combination of characters. Generally, the origins of these characters are ignored ... as is the actual genealogy of the organisms in question.'[6] Taxonomy is interested in delineating diagnostic characters that are extremely particular and intra-specific. Especially given a combination of such characters, sub-species types are readily identified. Again, taxonomy *can be* phylogenetic (e.g., PhyloCode), but it need not be.

Now compare, in general, groups identified in taxonomy and in ecology. Unlike fitness and function measures used to identify *ecotypes* in the context of ecology (Pigliucci and Kaplan 2003), taxonomic types recognised by diagnostic characters need not have any sort of adaptive function. Indeed, as already pointed out by Darwin, the most reliable and informative characters for taxonomic typing are non-adaptive and neutral (e.g. Darwin [1859] 2001: 414–15, 433). This key contrast between neutral and purely phenomenological types (taxonomy) and fine-tuned, adaptive types (ecology) is critical to understanding how very different sorts of groups and sub-species ('races') would be identified by the genomic data. Different sorts of data (i.e. neutral vs. functional genome sequences) would have to be employed.

In summary, there are a variety of concepts and aims, and measures and models, for how to stabilise sub-species populations and groups, or races, within human and non-human species. The units this plurality produces often cross-cut. Thus, there is no single, stable way of identifying human races. Many kinds of racial ontologies are created. Let us now examine how each sub-discipline – with unique concepts and aims, measures and models – makes possible continuum (anti-realist) and subdivision (realist) ontological glosses.

Continuum vs. Subdivision Perspectives in the Four Sub-disciplines

We suggested above that genuine, reasonable, and irreducible disagreements in biology regarding the ontological status of particular sub-populations are common. Furthermore, these disagreements are not generally the result of disagreements over the basic biological facts, but rather over the correct interpretations of complex genomic data and models. Indeed, a prevalent form of disagreement is that between lumpers and lumping, on the one hand, and splitting and splitters, on the other hand. Zerubavel (1996: 421) explores the 'mental process' allowing 'grouping "similar" things together in a single mental cluster' (i.e. lumping) and, conversely, 'perceiving "different" clusters as separate from one another' (i.e. splitting). Cognitive and sociological processes of lumping suggest the continuum view; processes of splitting suggest subdi-

vision. In their investigation of the politics and ethics of 'classification and its consequences', Bowker and Star (1999: 45) note that lumpers are 'those who see fewer categories and more commonalities', while splitters are 'those who would name a new species with fewer kinds of difference cited'. We adopt Zerubavel's and Bowker and Star's characterisations of lumping and splitting, generalising splitters to those biologists who wish to name not just species but also sub-varieties or races with 'fewer kinds [and number] of difference'.

In this section, we flesh out how disagreements over the interpretation of biological data, including debate over whether lumping or splitting should be performed, play out in the human case. Confusion and convention in the case of non-human animals and plants carries over to the human case. We argue that the kinds of factors that result in real disagreements among biologists imply that the recognition of a particular subpopulation often involves decisions made on the basis of factors that go beyond narrow, empiricist conceptions of putatively transparent biological data.

Taxonomy: Naming Conventions and Population Structure

The question of the ontological status of biogenomic 'race' might seem to be a straightforward question of taxonomic practice. Are the populations that, in ordinary social discourse, we call 'races', the sorts of populations that a reasonable biologist would recognise as worthy of attention, and perhaps a name? Diametrically opposed answers to this question can be found in recent philosophical and biological literature.

Recently, Sesardic (2010) has suggested (without stating unequivocally) that the answer to this question is 'yes'. Sesardic argues that while there may not be much genetic variation between the populations on a 'locus-by-locus basis', that 'the aggregation effect of these inter-group differences' is such that a 'racial taxonomy' might well be supported (2010: 149). In this, he follows, in his own way, Edwards on the value of internal structure for identifying subpopulations (see Edwards 2003). Andreasen (2000, 2004) reaches a similar conclusion on the basis of genealogical considerations: human races are biologically real entities. She is a taxonomic splitter. 'Races' as used in some ordinary discourses are, according to Andreasen, the remnants of 'clades', (probably on 'their way out' due to increased migration and globalisation in the modern world; see Andreasen 2004: 431), that is, local breeding populations cut off from each other for significant periods of time. This kind of population structure gives rise to the genetic correlations that Edwards focuses on in his argument against Lewontin, and which Sesardic takes to be strong evidence against the biogenomic racial scepticism (aka continuum) position.

Templeton (1999) and Hochman (forthcoming) both argue that the degree of differentiation among human populations does not rise to the level required by biologists to recognise the populations as legitimate populations of interest. They are taxonomic lumpers, interpreting the human species as (roughly) a continuous genetic field. First, Templeton (1999: 634) shows that global

human F_{ST} is fairly low (= .156), somewhere between impalas and waterbucks, and nowhere close to the highly differentiated species of North American deer and gray wolf.[7] Moreover, he argues that 'human "races" do not satisfy the standard quantitative criterion for being traditional subspecies (Smith et al. 1997)' (ibid.: 635). Indeed, 'a standard criterion for a subspecies or race in the nonhuman literature under the traditional definition of a subspecies as a geographically circumscribed, sharply differentiated population is to have F_{ST} values of at least 0.25 to 0.30 (Smith et al. 1997)' (Templeton 1999: 633) Along similar lines, also interpreting the question to be about 'subspecies', Hochman argues that a subspecies level division generally requires far more differentiation than is found in human populations, including evidence of longer periods of (and more complete) genetic isolation. Hochman admits, however, that the 'debate surrounding racial naturalism does not take place against anything like a stable scientific backdrop' (Hochman forthcoming) because both the standards for recognising population structure below the species level and the legitimacy of the categories deployed in doing so remain contentious.

This is, unfortunately, precisely correct. If there were an orderly, agreed-upon system for recognising populations in biology below the species level, that system could simply be deployed in the case of the biogenomic race to yield an answer. That answer would be limited in that it would tell us only that biologists would or would not recognise a particular human population as a 'real' population worthy of attention. It would be an answer nonetheless. However, *pace* Templeton's seemingly confident pronouncements, there is no such agreement.

Turning to non-human organisms, population structure revealed by tiny differences is sometimes regarded as important (i.e. splitting). Consider for example the Cross River gorilla (*Gorilla gorilla diehli*). Bergl and Vigilant (2007) argue that, based on their analysis of the population structure within this particular subspecies of gorilla (using STRUCTURE, the same programme whose results are often cited as revealing human subpopulations; see Pritchard et al. 2000, Rosenberg et al. 2002, Feldman 2010), there is still some gene exchange between those subpopulations generally identified, and hence room for conservation efforts to consider larger potential effective population sizes than those implied by the small subpopulations considered independently.

Of interest here is not the argument regarding conservation efforts (though these arguments do reflect some of the difficulties we will suggest in the third section) but rather the fact that individual populations differing only in very small ways, and which furthermore continue to exchange genes with each other, are still considered in this case to be deserving of recognition as populations worthy of particular names and identifications. The 'reductio' that Hochman considers – that if we accept 'racial' designations on the basis of population structure, we would be forced to recognise even smaller subpopulations – is not here considered a reductio at all, but rather casually embraced

as part of a sensible analysis of population structure within a critically endangered subspecies. Again, genetic data underdetermines ontological interpretation, which is based more on conventional and pragmatic decisions about splitting and lumping than about the genomic facts.

In the end, whether human bio-genomic races are recognised as 'real' taxonomic units will depend less on the details of our population structure, and more on the conventions chosen regarding what kinds of populations we care to identify.

Phylogenetics: Getting Out What You Put In?

We now turn to attempts to reconstruct the history of human origins, diversity and migrations. Phylogenetics' charge is to reconstruct the history of life. Our basic argument in this section is that, again, the genomic data and models can be perfectly reasonably interpreted from continuum (lumping) and subdivision (splitting) perspectives. Genomic phenomena of *Homo sapiens* in the context of phylogenetics do not justify our choices regarding which populations to single out as worthy of particular attention. Indeed, as one of us has argued, perhaps more so in phylogenetics than in other sub-disciplines, we get out of the models what we put in (Winther 2009). Before we discuss the underdetermination of racial ontology by the results of phylogenetics, some preliminary discussion regarding phylogenetic data and models is in order.

Important work in the 1960s showed that molecular signals (genes or gene products – i.e. proteins) could be used to reconstruct phylogenetic history (Zuckerkandl and Pauling 1965), and results regarding the fairly early split (5 million, rather than 13 million) of humans and African great apes were put forward (Sarich and Wilson 1967).[8] Since then, this data has become further enriched as nucleotide sequences have become data-analysis targets: CNVs (Copy Number Variants), SNVs and SNPs (Single Nucleotide Variations, and Single Nucleotide Polymorphisms), and STR (Short Tandem Repeats). Moreover, entire genomes are being analysed (e.g. GWAS – Genome Wide Association Studies), and mitochondrial (matrilineal; e.g. van Oven and Kayser 2009) and Y-chromosome (patrilineal; e.g. Jobling and Tyler-Smith 2003) data have been added to the mix. Forms of genomic data useful for phylogenetics are quite complex indeed.

Regarding phylogenetic modelling, two influential lineages or families of modelling in twentieth-century intellectual history can be found: cladistic and probabilistic. Cladism holds that parsimony is the best method for inferring trees. In order for classifications to be natural and objective, they must refer to systematisations of the order of nature captured in our cladograms, which show a nested clade structure (see Winther 2012 section 6.2.2). As Darwin argued, 'all true classification is genealogical' ([1859] 2001: 420). In contrast, probabilistic phylogenetics employs for instance genetic distance measures (sometimes derived from genetic differentiation measures, as explained earlier) to construct trees. According to Felsenstein, 'One of the foundations of

numerical work on phylogenies was the remarkably creative work of Anthony Edwards and Luca Cavalli-Sforza … Both had been students of the famous statistician and population geneticist R. A. Fisher. They were trying to make trees of human populations from gene frequencies of blood group alleles' (2004: 125). To provide a flavour of the impact, Felsenstein in his 2004 chapter 'A digression on history and philosophy' continues: 'Edwards and Cavalli-Sforza's paper of 1964 is remarkable in that it introduces the parsimony method, the likelihood method, and the statistical inference approach to inferring phylogenies, all in one paper' (2004: 128). In short, cladistic and probabilistic methods are standard modelling strategies in phylogenetics.

Now that we have provided some background regarding the data and modelling of phylogenetics in general, let us return to the identification of populations and groups in our species. Here are a few reasons for why we should not expect phylogenetics to provide unequivocal answers for racial ontology: (i) comparison of extreme lumpers and splitters; (ii) data choice; (iii) choice of models and measures; (iv) choice of tree vs. trellis topology; and (v) getting out what you put in (i.e. using anthropological and linguistic information).

First, consider the extreme cases for lumping vs. splitting. On the one hand, there is work on the evolution of *Homo sapiens,* as a single, entire, typological species, compared with, for instance, chimpanzees (Yamamichi et al. 2012).[9] This kind of lumping has been fairly standard in work on human evolution. On the other hand, significant strands of contemporary work on human phylogenetics takes subdivision and splitting very seriously, appealing to various anthropological and linguistic sources of information (see below) to identify 26 (Nei and Roychoudhury 1993) or 42 (Cavalli-Sforza et al. 1994) populations. Again, decisions and conventions related to the purposes of the phylogenetic analysis have to be made in deciding whether to engage in extreme lumping or splitting. The facts do not on their own provide answers for these questions.

Second, data choice can provide distinct answers for how and what to split (or lump). For instance, different lineages will be traced depending on whether mitochondria or Y-chromosome data is being tracked. Indeed, it turns out that we can track the differential migration paths and mechanisms for women and men (e.g. Wells 2003). Perhaps a 'total evidence' approach (e.g., Eernisse and Kluge 1993) could adjudicate some of these data conflicts. However, the variety of data seems to point to a diversity of incongruous lineage signatures, with no single, neat tree possible. Moreover, given kin and marriage relations, war, invasions and forced migrations, various methodological decisions have to be made regarding what could possibly count as reasonable populations, useful for data collection and phylogenetic reconstruction.

Third, different modelling methods and genetic distance measures are used by the interlocutors in this debate: Nei and collaborators use Neighbour-Joining modelling, whereas Cavalli-Sforza and collaborators use UPGMA, i.e. Unweighted Pair-Group Method with arithmetic Averages (reviewed in

Felsenstein 2004); and the Nei distance measure contrasts with Cavalli-Sforza's F_{ST}-based genetic differentiation measure. This is significant because when mid-level clusters (or clades) within *Homo sapiens* are compared across these methods, they fail to correspond perfectly. Nei and collaborators are rather interested in determining the major races, and are more 'lumpers' at this level of granularity than Cavalli-Sforza and colleagues, who often tend simply to wish to make inferences about the high-level details of the human phylogenetic tree, and are more 'splitters'. When one compares their results, one observes that both have 'Africans' as outgroups (see, e.g., figures 2.3.2.A in Cavalli-Sforza et al. 1994: 78; figure 2 in Nei and Roychoudhury 1993: 932). However, the Cavalli-Sforza et al. phylogeny separates out Southeast Asians and Pacific Islanders from everybody else remaining, whereas the Nei et al. phylogeny separates out Caucasians from everybody else remaining. This is a remarkable difference at such a deep level in the tree. The gross topology (and cladistic groupings) simply does not match between these two influential phylogenies.[10]

Fourth, a tree topology may not be the right topology. While this is not the place to enter the territory of graph or network theory (but see Huson et al. 2010), there is an important discussion about whether, given processes of admixture and migration, the evolution of *Homo sapiens* diversity can ever be represented by a well-behaved tree, as per the Out-of-Africa hypothesis, or whether we need a 'trellis' model indicating repeated episodes of population isolation and population hybridisation in *Homo sapiens* and *Homo erectus*, as an instance of a multi-regional hypothesis.[11] If hybridisation (mechanism) and branch reticulation (topology) have been as common as Templeton and others suggest, then indeed it becomes even less clear how to subdivide *Homo sapiens*. Again, such an anti-realist argument need not carry the day. Legitimate modelling strategies using particular measures can still solidify a phylogenetically based racial ontology out of the indeterminate genomic field (e.g. figure 5 of Mountain and Cavalli-Sforza 1997: 712; Agrawal and Khan 2005). Again, the particular research paradigms and purposes embraced explain more about the particular decisions made regarding model choice than do the genomic facts.

Fifth, in modelling one often has to be careful about simply getting out what one has put in to the model. How are the groups used in human phylogenetics determined in the first place? Which criteria and definitions, measures and models are used?[12] Cavalli-Sforza and collaborators are fairly extreme splitters and, in their phylogenetic investigations of human evolution, employ other sources of data besides genomics: 'Candidates [of "human demes"] could be ethnographic units (e.g. tribes) or geographically defined clusters of people (villages, towns, cities). They are all usually endogamous to some degree and may come closer to the definition of a deme, but there are always many possible, embarrassing choices' (Cavalli-Sforza et al. 1994: 21). Indeed:

> Our main criterion in pooling ['aboriginal', Cavalli-Sforza et al. 1988] populations for generating higher categories was geographic, but it was clear that, especially for populations from the developing world, the geographic criterion had to be supplemented with general anthropological information of some kind ... We decided to resort to linguistics when other criteria failed since there is a certain amount of parallelism between the linguistic and genetic evolution of populations. (Cavalli-Sforza et al. 1994: 22)

The phylogenetics of *Homo sapiens* already 'puts in' certain groups identified by criteria alien to genomics. Modelling circularity of you-get-out-what-you-put-in exists in this case. Clade-cutting, as it were, is strongly conventionalist.[13] How many populations does a clade need in order to count as a population, group, or race? Again, it is not clear how genomics is supposed to (in any way) give us natural races of *Homo sapiens* in phylogenetics.

In this section a number of conventional choices that need to be made in human phylogenetics have been reviewed: data choice; choice of models and measures; choice of tree vs. trellis topology; and getting out what you put in (i.e. using anthropological and linguistic information). The decisions made about each of these provide better explanations of why particular interlocutors choose to adopt racial realism or anti-realism than do the genomic facts to which they appeal. In phylogenetics, genomics also underdetermine racial ontology. Constructivist conventionalism is the appropriate ontological gloss vis-à-vis biogenomic race in phylogenetics.

Conservation Biology: Lewontin and the Alien

> To effectively conserve regional biodiversity, conservationists need to know how diversity is distributed geographically within the region. Does the region consist of many distinct communities, or is it homogeneous? How much does each community contribute to the regional diversity? How different are the communities? The answers to these questions determine how conservation resources should be allocated among sites in the region. (Jost et al. 2010: 65)

In 1972, Lewontin published the remarkable 'The Apportionment of Human Diversity', famously concluding that (i) the vast majority (more than 85%) of the total genetic variation in the human species existed within any given population, (ii) relatively little (only about 6%) could be attributed to variation between the 'major' continental groups ('races'), and (iii) the remainder (approx. 8%) was found among different populations within each continental group. One might quibble with Lewontin's choice of markers (proteins rather than gene sequences), and with the measures chosen (a Shannon entropy measure rather than a true diversity measure; see Kaplan and Winther 2012), but the basic result that the vast majority of genetic variation exists within any particular population has held up startlingly well to reanalysis with different data sets and measures (see Nei and Roychoudhury 1972; Barbujani et al.

1997; Jorde et al. 2000; Brown and Armelagos 2001). What has not always been so clear, however, is what to make of this claim. What, precisely, is our interest in assessing between- versus within-population diversity?

One hint at the interpretative difficulties facing claims of this sort is provided by Lewontin himself. In a later work (co-authored with Rose and Kamin), Lewontin remarks that the results of analysing between- and within-population genetic variation imply that 'if everybody on earth became extinct except for the Kikiyu of East Africa, about eighty-five percent of all human variation would still be present in the reconstituted species' (Rose et al. 1984: 126). Again, while one might quibble with some of the details underlying this claim (for instance, the Kikiyu of East Africa might very well represent significantly more than the 'average' 85% of the total variation), it is the thrust of this comment that we find the most telling. Comparisons of the genetic variations within and between populations are most at home in conservation biology. Indeed, it is in conservation biology that arguments surrounding what constitute 'true' measures of diversity are the most prevalent (and where much of the most interesting work on these questions has been done; see Lowe et al. 2004; Jost 2008). Lewontin's comment also carries more than a hint of the conservation biologist's focus.

When encountering a wide-ranging species facing the destruction of much of its habitat, it is the charge of the conservation biologist to ask, 'what would happen if we could only save *this part*? OK, how about *this part*?' Lewontin's comments can be read as suggesting that a conservation biologist (perhaps an alien, even) forced to save only a small subpopulation of *Homo species* could reasonably pick any small subpopulation, nearly at random, and be confident that she would not be losing much of the extant genetic variation in so doing. It is this thought experiment that we wish to take seriously in this section. At least some of the arguments surrounding the ontological status of biogenomic races depend on our answering the question 'what would biologists say about saving this population, if it was anything except humans?' Our answer? 'It depends.'

The alien conservation biologist thought experiment makes perspicuous the implications of the claim that any particular population could practically stand in for the whole of humanity without any great loss. An alien lumper might well point towards the genetic homogeneity of our species, and the large portion of within-population variation, and lump all of humanity together as a single population with no meaningful subdivisions. However, one can easily imagine conservation biologists more committed to saving particular aspects of diversity. For example, we can imagine conservation biologists less willing to give up on unique alleles – conservation biologists for whom saving only 85% of the extant genetic variation at the allele level would represent an unacceptable loss. Or, we can imagine a conservation biologist who is unwilling to give up unique populations that, while genetically similar at the level of individual alleles, have particular collections of allele frequencies that make them special; that is, a conservation biologist who is committed to preserving population

structure (a quintessentially splitter position). It is imaginable that another flavor of conservation biologist might look not towards measured diversity, but towards history, and want instead to find and preserve 'phylogentic subspecies' (for example), however similar they turned out to be, genetically or phenotypically. An ontology of subdivision or continuity is not automatically given by the biological or genomic facts of conservation biology. Choices have to be made, guided by certain norms and conventions.

In the first case, a conservation biologist who wanted to preserve more than the average 85 per cent of the variation Lewontin alluded to might note the following. Firstly, since the vast majority of genetic diversity exists within Africa,[14] any populations chosen should probably be African; and secondly, a collection of populations from Africa should be preferred to a collection of populations drawn from the various continents. Certainly, if our goal were actually to reconstitute the species, this strategy would have much to recommend it. While we might lose many particular *combinations* of genes, the vast majority of the genes themselves would be present, and combinations could of course be recreated. In the same way, a pigeon fancier faced with only being able to save a very few pigeons would be wise to pick a small but genetically diverse wild population, from which the full range of exciting pigeon varieties could almost always be recreated, given much time and effort (see Stringham et al. 2012; Shapiro et al. 2013).

But even saving (nearly) every individual allelic variant would not preserve the full range of diversity. Particular populations, with different frequencies of allele combinations, can be unique and interesting, even if none of the individual genes are themselves unique. This is roughly the situation given human population structure. There is nothing in the practice or the theory of conservation biology that would speak against choosing populations from each continent, but certainly nothing in practice or theory would require one to do that.[15]

Ecology: Splitters and Ecotypes

Turn to ecology. Here, again, we wish to raise the question of whether a 'splitter' ecologist, treating the human population like some other population of organisms, would find interesting divisions within our species worthy of attention, and whether those divisions might fall out along the lines of 'race' as ordinarily understood. Our answer is 'it depends'. There is no way to answer this question without knowing the particular interests of the ecologist in question, and those interests are determined by factors that go beyond an attention to the biological details as ordinarily understood. Ecologists will in some contexts write about groups that comprise multiple different species, and consider the implications that changes to groups might have; Jost et al. (2010) for example, in defending a particular interpretation of diversity, apply their measure to a particular 'Neotropical fruit-feeding butterfly guild' in Ecuador (2010: 68).

Put crudely, one might, as an ecologist, be curious about the role that a particular kind of grass, say, *Paspalum notatum*, plays in pasture environments

(see e.g. Daurelio et al. 2004), or one might be interested in how heavy-metal tolerance evolves in grasses of this type (see e.g. Teng et al. 2008). From the perspective of understanding the ecology of large-scale environments, the latter is unlikely to ever be more than a trivial wrinkle. But from the perspective of someone interested in the evolution of specific ecotypes within populations, of course, those trivial wrinkles are where the action is. Again, how ecologists subdivide species into functional ecotypes, group several species under a single biome or under a single functional guild will vary depending on research group and research question. Moreover, one of us has previously defended treating some human populations as 'ecotypes' (Pigluicci and Kaplan 2003). Again, the decision to do so, and thereby be a splitter, depends critically on one's conventional and purpose-oriented interest in relatively small populations adapted to local conditions.

Conclusions: Political and Social Implications

This paper has argued that no neutral appeal to biological practice can settle the question of the legitimacy of treating the human species as being subdivided into various biological sub-divisions. Even within those biological fields in which these questions emerge in the case of non-human biological entities, debates between biologists in favour of 'lumping' biological entities (stressing the continuous nature of a population, species, etc.) and 'splitting' (stressing the existence of structured groups based on genomic variation, however slight) can be found. While there are exemplars of species that cannot reasonably be divided, and exemplars of species that nearly everyone would agree contain important divisions, the human case, along with many other species of interest, lies firmly in the 'gray zone', where divisions are possible but not required.

At this point, it is worth pausing to reflect on why the ontological status of biogenomic human races matters. The use of population structure to reconstruct the history of the human species, including our history of migrations and the history of gene exchange within and between geographical areas, is obviously of some real intellectual interest. Similarly, elucidating the existence of internal population structure more generally may be of some intrinsic concern. And if the debates in the literature were really about whether there were particular human populations, identifiable by the subtle genetic differences that resulted from the existence of population structure, the answer might be provided by the analysis of straightforward biological and genomic data. To return to an example developed above, biologists do sometimes need to decide if some particular tiny subpopulation, restricted to some small valley, of an already rare subspecies of gorilla, is in fact sufficiently different from the population on the other side of a ridgeline to be worth treating as a separate population. And, utilising a variety of techniques for detecting population structure, they can come to decisions vis-à-vis lumping or splitting, and give

reasons for the decisions that they come to, which other biologists can challenge on intellectual and technical grounds (though they rarely do so on moral grounds). But if our ability to make this kind of subdivision were really the only question in the human case, the vigour with which it is pursued would be entirely mysterious.

So what is the issue in the human case, if it is not that of characterising the mere existence of population structure? The vast majority of subpopulations that can be 'discovered' by genomic clustering analysis are of no real interest; they are trivial effects of real but unimportant population structure. In contrast, the ordinary racial categories deployed in social discourse and social power structures are important. In the U.S., for example, the difference in average life-prospects between native-born Black Americans and native-born White Americans is sufficient to shock the conscience. Answers to questions about the reality of biogenomic race are *supposed to* speak to such disparities between socially identified (and socially meaningful) 'races' – in both the normative and descriptive senses of the 'supposed to'.

It is for this reason that discussions of the biogenomic reality of race degenerate quickly into suspicion, and often into outright accusations, of ulterior motivations. Thus, arguments surrounding the viability of splitting populations immediately run into questions regarding the reasons for doing so. The question, in the human case, was never 'can this population be split into two populations distinguishable on the basis of the subtle results of internal population structure?' For example, Sesardic's argument in defence of the existence of biogenomic races immediately segues into a defence of the so-called 'hereditarian' position on IQ differences (Sesardic 2010). That position – the claim that the 'average' difference in scores on standardised tests between populations reflect 'native' endowments – is rightly viewed with deep suspicion by many people and groups interested in social justice. It is a view that was deployed by, for example, Herrnstein and Murray to defend a society with a strongly hierarchical hereditary class structure. But that argument, and many others, demands far more than the existence (or not) of detectable population structure. To pretend that such arguments are about population structure when they really concern particular views about our moral responsibility to those currently severely disadvantaged in our society seems misguided.

Similarly, while the discussions inspire perhaps less vitriol, the controversy over the implications of biogenomic racial realism are no less real in biomedicine. Again, the question is not simply whether some differences among arbitrary populations smaller than the entire human species exist, nor even whether subpopulation-level differences might have medical consequences. This is for a couple of reasons. First, because no one disputes that such populations sometimes exist, and have biomedical consequences (Tay-Sachs disease, Thalassemia, lactase persistence, HIV resistance); and second, because questions regarding the force of these populations can (sometimes) be answered (fairly) straightforwardly. But, again, the disparities in health-outcomes between Black

and White Americans are dramatic, and not easily attributed to simple monofactorial causes. This has led some researchers to posit that there is something in the 'average' genome of biogenomic subpopulation that make up 'Blacks' in the U.S. that accounts for their poor health (Risch et al. 2002; Collins et al. 2003; regarding analogous arguments for the 'Mexican' genome, see Silva-Zolezzi et al. 2009; critique in López-Beltrán 2011). Others have argued vociferously that the differences in health-outcomes are best accounted for not on the basis of shared genetic risk factors, but rather on the basis of shared social circumstances (namely, racism and the legacies of racism) (see Gravlee 2009; Kaplan 2010; for broader discussion of the way racism becomes internalised as a consequence of various social factors, including colonialism, see Fanon [1952] 2008). This matters because it would appear that if the root of the health-disparities is a difference in inborn metabolic pathways (say), a biomedical response is most obviously appropriate. However, if the origin of the health disparities is racism, a political and social response would seem more appropriate. In both cases, people of genuine good will wish to address these health disparities. Nevertheless, the moral significance of those health disparities is starker if we accept that they are the result of an actively racist society.

Return to multiple underdetermination. If our main line of argument is correct, then no social and political agenda, with associated policy recommendations, can ever hope to turn to genomics as the source and justification of racial ontology. Either lumping or splitting positions can be defended in each of the four sub-disciplines excavated in section 3. Even once a whole host of conventional decisions are made regarding biogenomic continuum or subdivision perspectives, the relationship between those ways of categorising the world and the categories we call 'races' in social contexts remain contentious. Moreover, to get from claims about the ontological status of the social categories we call 'races' to any political positions or policies that we care about requires many further assumptions and decisions. It is also for this reason that both liberal and conservative agendas can make use of either the realist or anti-realist positions. For instance, left-wing 'race consciousness' movements and right-wing fascistic agendas can both capitalise on the (putative) existence of biogenomic race. Conversely, the (supposed) non-existence of biogenomic race can be leveraged to support either liberal affirmative action policies responding to the unequal outcomes that it is argued are the result of racism and the legacies of racism, or to support conservative anti-affirmative action policies that take the choices, decisions and behaviours of individuals, against a claimed backdrop of broad equality of opportunity, to be the source of current inequalities. Multiple mappings indeed.

Now, because positions in this debate are exceedingly complex and can lead to potential misunderstandings through bad faith, false consciousness or misreading, intentional or not, it is important to flag our own political position explicitly, qua authors. We find the so-called 'hereditarian' position with respect to the differences in average IQ scores between the 'races' abhorrent –

intellectually, empirically and morally. We firmly reject the 'conservative' position that the severe differences in average life prospects between native-born Black and White Americans could possibly be justified by any biological facts.[16] But our dismissal of the hereditarian position is based not on our denying the existence of population structure in humans, nor on any simple biological principle or fact to which one could appeal. Rather, our informed rejection of hereditarianism emerges from multiple lines of evidence in the social sciences, philosophy and biology.

Is biology then completely irrelevant to political and social agendas? This paper has defended a constructivist conventionalism only about biogenomic race. Racial ontology is (forever) underdetermined by our finest genomic results. Therefore, political programmes cannot turn to genomics for a guiding, univocal racial ontology. The biogenomic race concept is a *fata morgana* (see note 1). Biology is broader than genomics, though. Where policy recommendations pertinent to 'social race' really do rest on biological facts, such facts will not be claims about the overall structure of human genomic variation, but very specific claims concerning individuals subject to systematic social-environmental influences. Such facts can and need to take account of far more than genomic data, including development, physiology, epigenetics, local ecology and environment. In a broad view, biological processes can be seen as part of the complex developmental systems of the human condition (Levins and Lewontin 1985; Oyama 2000; Oyama et al. 2001). They should be included within a full and accurate account of such systems, where they are placed in their appropriate social and political context and not illegitimately averaged.

Rasmus Grønfeldt Winther, Associate Professor of Philosophy in the University of California, Santa Cruz, investigates the promises and dangers of scientific theory. He has lectured and published on a wide variety of topics in philosophy of science as well as on science more generally. He is the History and Philosophy of Biology book series editor at Pickering and Chatto (London), and PI of the "Genomics and Philosophy of Race" Research Cluster, a collaborative research project involving UC Santa Cruz, UC Davis, and Stanford University.

Jonathan Michael Kaplan, Associate Professor of Philosophy at Oregon State University, is the author of *Making Sense of Evolution* (with Massimo Pigliucci, University of Chicago Press, 2006) and *The Limits and Lies of Genetic Research* (Routledge Press, 2000. He is one of the founding editors of the on-line, open-access journal "Philosophy and Theory in Biology.

Acknowledgments

Winther is grateful to Ian Hacking, Amir Najmi, Michael J. Wade and Bøllemis for ongoing discussions on these matters; he is partially supported by a faculty research grant from the Academic Senate Committee on Research at the University of California, Santa Cruz. Kaplan thanks the Oregon State University Department of Philosophy and Center for the Humanities, whose support helped make this work possible. Both appreciate the critique of two anonymous reviewers, Marcus Feldman, Lucas McGranahan, Matthew Kopec, Massimo Pigliucci, Elliott Sober and undergraduate students in Winther's 'Philosophy of Race' and 'Classifying Persons: Philosophy, Politics, and Possibilities/Problems' (UC Santa Cruz, Winter 2013) courses. This paper is written fully jointly.

Notes

1. For the purposes of this paper, we take 'bio-genomic race' to refer to any legitimate subdivision of populations below the species level made on the basis of genomic differentiation. This generic concept has four meanings or inflections, characteristic of the four biological sub-disciplines here excavated – e.g., taxonomic race, ecotypical race, etc. We do not wish to make, multiply or regiment the intension and use of concepts. The bio-genomic race concept is here used because many interlocutors believe in it, and in its power to finally resolve the question of whether race exists. Bio-genomic race will never, and can never, be the judge (or jury) of the existence (or not) of 'race', in any sense of that term. Indeed, the very concept and its sub-concepts, lose force and coherence through our analysis. Actual biological practice shows them to be unstable *fata morganas*. There is nothing there.
2. Even then, however, as long as there was some population structure, potentially very small, genomic methods would find differences quite effectively and so splitters might still exist.
3. While not denying that within these disciplines, additional information from non-genomic sources can sometimes tip the balance with respect to what is considered the proper ontological stance towards particular sub-populations, or sub-populations in general, our focus in this paper is primarily on judgments made about clusters, populations, clades, groups, etc., emerging from genomic data and methodology, as in recent years, these have been regarded as the most critical with respect to human bio-genomic races.
4. Here the concept of 'typology' is used in the morally neutral conceptual and methodological sense of standard taxonomy. Think of the transcendental anatomy of early 19th century German Naturphilosophie, e.g., by Goethe and Lorenz Oken, or of the British morphologist Richard Owen. One goal in this research program was to identify the same specific organ—i.e., a particular homologue—in different species. For instance, inner ear bones of mammals were found to be homologues ('the same') to various jaw bones of reptile species. Darwin gave homology a dynamic, evolutionary interpretation but the sameness of typology is strictly speaking logically independent of the mechanisms or causes of that sameness. (For recent uses and characterizations of typology see Rieppel 1994, Brigandt 2009, Love 2009, Winther 2009.) We use 'typology' as a basic methodological and ontological concept of

taxonomy, and not as synonymous with the morally laden term 'essentialism.' Thanks are due to Matthew Kopec for bringing this potential confusion to our attention.

5. As we note in (Kaplan and Winther 2012) there is substantial confusion in the literature regarding what, exactly, particular measures (e.g. F_{ST}) in fact measure. The relationship between what gets called within- and between-population diversity, given a particular measure, varies substantially given different measures.
6. If the community of scholars cannot solve the species problem (because of a plurality of definitions, concepts, and identification criteria), how could we hope to solve the subspecies problem? The latter conceptually inherits many of the same problems as the former, and adds its own, including the meanings and measures of intra-specific formal analyses. Put differently, technical arguments about subspecies subdivision and splitting parallel arguments about species subdivision and splitting. (We thank Michael J. Wade and Elliott Sober for this observation.) More broadly, just like social and political arguments are underdetermined by genomics of race, so they are underdetermined by the genomics of species and our responsibilities and duties, qua Homo sapiens, to sister species (e.g., chimpanzees, bonobos) and beyond.
7. Keinan and Reich (2010) provide a slightly lower F_{ST} of approximately .12. What matters is that we fall somewhere in the middle of the range of most species. For instance, the fixation index of many marine species is significantly below .5, even frequently approaching 0 (e.g., Weersing and Toonen 2009, Fig. 1, p. 6).
8. Indeed, Lewontin and Hubby (1966) was an influential early use of molecular technologies to assess population genetic variation, in fruit flies. This is a research paradigm within which one can also place Lewontin (1972).
9. Given rich data and complex statistical models based on coalescent theory using maximum likelihood estimation, Yamamichi et al. (2012) estimated effectively instantaneous speciation between humans and chimpanzees 6.1 million years ago. Differing somewhat, Patterson et al. (2006) remain typological, but model human-chimp speciation as a complex process with multiple hybridization.
10. As a further comparison of measures and models cross cutting (section 2 above), in the context of Cavalli-Sforza's collaborative work, consider the relation between taxonomy and phylogenetics. In discussing their choice of how they clustered 42 populations into 9 clusters for 'reducing the complexity of the data' and 'simpl[ifying] discussion' (Cavalli-Sforza 1994, p. 80), they note: 'The nine clusters chosen differ in their genetic homogeneity, but we are interested in establishing history not in generating a classification scheme.' (p. 79) Indeed, they insist on the 'difference between taxonomy and phylogenetic analysis' and the appropriateness of abstracting away from differences in intra-cluster genetic homogeneity for purposes of the latter, though not the former.
11. See Templeton 1997, 1999. Adams 2008 usefully discusses how both Out-of-Africa and multiregional hypotheses can be represented with trees, but only the multiregional hypothesis can, broadly speaking, also be represented with a trellis. Relethford 1998 and Wells 2003 review these two hypotheses further.
12. On philosophical concerns regarding the empirical base of evolutionary inferences, see Sober 2008.
13. One might argue that straightforward genomic clustering á là Rosenberg et al. (2002) could permit the formation of clusters upon which we could then do phylogenetic inference. This is unlikely. The Bayesian clustering techniques Rosenberg and collaborators use are premised on particular methodological and ontological assumptions and interests, that need to be (conventionally) accepted in performing STRUCTURE clustering. One consequence of these is that STRUCTURE does not work well for more than roughly 10 clusters (i.e., $K = 10$). Moreover, further analysis of the promises and limits of using anthropological and linguistic categories in human phylogenetics is required (see, e.g., Winther 2011 and references therein).

14. Contemporary humans vary, on average, by about 1:1000 nucleotides; this is around an order of magnitude less variation than occurs in many other species (see, e.g., Li and Sadler 1991; Cognato 2007). All the measures of genetic variation discussed below rely on this small amount of variation when applied to humans. This variation is not distributed equally in humans; people whose ancestors are of recent African origin, for example, differ on average by about 1:900 nucleotides; people whose ancestors were of recent European origin differ by only about 1:1600 nucleotides. Edwards is correct when he notes that 'It is not true, as Nature claimed, that "two random individuals from any one group are almost as different as any two random individuals from the entire world"' (2003, p. 801) but not perhaps in the way he intended; interestingly, the 'average' person of recent African descent is more likely to share an arbitrary allele with an 'average' person of recent European descent (will differ less) than he or she is with another arbitrarily chosen person of African descent (African v. European difference is about 1:1050, compared with 1:900 difference within Africa) (see Yu et al 2002).
15. It is of course logically possible to admit that a putative population is a legitimate biological population, and simultaneously deny that it ought to be protected or conserved. In practice, however, arguments regarding the value of protecting populations tend to be closely tied to arguments regarding the biological legitimacy of the population in question. For example, Scharpf (2000), analyzing some of the arguments over the existence and biological uniqueness of the 'Alabama Sturgeon,' suggests that it is often the desire to not protect a population that leads to claims regarding the biological illegitimacy of the claimed population (see also Campton et al 2000). Moreover, the wealth of species concepts subsumed by the 'surrogate species' concept (e.g., 'flagship species,' 'umbrella species,' and 'population indicator species'; see Caro and O'Doherty 1999), implies that in conservation ecology there is conventionalism not only about the reality or not of populations at different granular levels (including the species level), but also about which kind of species (and populations) we should be monitoring and conserving.
16. A certain paradox may seem to suggest itself. How can we argue against dichotomous, essentialist thinking vis-a-vis bio-genomic race for most of this paper, and now talk about 'black/white' health and life prospect disparities? Admittedly, there is intragroup variation in social and economic factors. Moreover, more groups in contemporary North American society (e.g., Native Americans / First-Nations Peoples, Latin Americans, at various levels of granularity, etc.) must be taken into account in this discussion. However, unlike the case of genomics, for political, social, and economic phenomena, there are clear mechanisms (e.g., racism and allosteric load, see Fanon [1952] 2008, Gravlee 2009, Kaplan 2010) and measures and metrics (e.g., political representation, wealth) indicating significant intergroup variation and, especially in the action of those mechanisms, much less intragroup variation than in the genomics case. While dichotomous and averaging thinking should be resisted, a dualistic, bimodal descriptive strategy is still all-too-painfully appropriate for the morally reprehensible and prudentially unwise racial clustering of political, social, and economic phenomena and mechanisms.

Bibliography

Adams, J. U. 2008. 'Human evolutionary tree', *Nature Education* 1(1). <http://www.nature.com/scitable/topicpage/human-evolutionary-tree-human-evolutionary-tree-Human-Evolutionary-Tree-417>.

Agrawal, S. and F. Khan. 2005. 'Reconstructing recent human phylogenies with forensic STR loci: A statistical approach', *BMC Genetics* 6:47. doi: 10.1186/1471-2156-6-47

Andreasen, R. O. 2000. 'Race: Biological reality or social construct?' *Philosophy of Science* 67:S653-S666

Andreasen, R. O. 2004. 'The cladistic race concept: A defense', *Biology and Philosophy* 19: 425–42.

Barbujani, G., A. Magagni, E. Minch, L. L. Cavalli-Sforza. 1997. 'An apportionment of human DNA diversity', *Proceedings of the National Academy of Sciences* 94: 4516–9.

Baum, D.A. and M. J. Donoghue. 1995. 'Choosing among alternative "phylogenetic" species concepts', *Systematic Botany* 20: 560–73.

Bergl, R.A. and L. Vigilant. 2007. 'Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*)', *Molecular Ecology* 16: 501–16.

Bowker, G. C. and S. L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Brigandt, I. 2009. 'Natural Kinds in Evolution and Systematics: Metaphysical and Epistemological Considerations', *Acta Biotheoretica* 57: 77-97.

Brown, R. A. and G. J. Armelagos. 2001. 'Apportionment of racial diversity: A review', *Evolutionary Anthropology* 10: 34–40.

Campton, D. R., Bass, A. L., Chapman, F. A., Bowen, B. W. 2000. 'Genetic distinction of pallid, shovelnose, and Alabama sturgeon: emerging species and the US Endangered Species Act', *Conservation Genetics* 1: 17-32.

Caro, T. M., O'Doherty, G. 1999. 'On the Use of Surrogate Species in Conservation Biology', *Conservation Biology* 13: 805-814.

Cavalli-Sforza, L., A. Piazza, P. Menozzi and J. Mountain. 1988. 'Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data', *Proceedings of the National Academy of Sciences* 85: 6002–6.

Cavalli-Sforza, L., P. Menozzi and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Cognato, A. I. 2007. 'A standard DNA taxonomy for insects?' *USDA Forest Service Proceedings RMRS-P-45*: 11–12.

Collins, F. S., E. D. Green, A. E. Guttmacher and M. S. Guyer for the Institute USNHGR. 2003. 'A vision for the future of genomics research', *Nature* 422: 835–47.

Crombie, AC. 1994. *Styles of Scientific Thinking in the European Tradition* (Vols. I–III). London: Duckworth.

Darwin, C. R. [1859] 2001. *On the Origin of Species by Means of Natural Selection or the Natural Selection of Favoured Races in the Struggle for Life.* Cambridge, MA: Harvard University Press.

Daurelio, L.D., F. Espinoza, C. L. Quarin and S. C. Pessino. 2004. 'Genetic diversity in sexual diploid and apomictic tetraploid populations of Paspalum notatum situated in sympatry or allopatry', *Plant Syst. Evol.* 244: 189–99.
Edwards, A. W. F. 2003. 'Human genetic diversity: Lewontin's fallacy', *BioEssays* 25: 798–801.
Edwards, A. W. F. and L. L. Cavalli-Sforza. 1964. 'Reconstruction of evolutionary trees,' in V. H. Heywood and M. McNeill (eds), *Phenetic and Phylogenetic Classification*, London: McNeill, 67–76.
Eernisse DJ, Kluge AG. 1993. 'Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology', *Molecular Biology and Evolution* 10: 1170-1195.
Fanon, F. [1952] 2008. *Black Skin, White Masks*. New York: Grove Press.
Feldman, M. W. 2010. 'The biology of ancestry: DNA, genomic variation, and race', in Markus, H. R., and Moya, P. M. L. (eds), *Doing Race: 21 Essays for the 21st Century*. New York: W.W. Norton.
Felsenstein, J. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer Associates Inc.
Foucault, M. 1969. *L'archéologie du savoir*. Paris: Gallimard. (*The Archaeology of Knowledge*, 1972, translated by A. Sheridan Smith, New York: Harper and Row.)
Galison, P. and D. Stump. 1996. *The Disunity of Science: Boundaries, Contexts and Power*. Stanford: Stanford University Press.
Goodman, N. 1978. *Ways of Worldmaking*. Indianapolis, IN: Hackett.
Gravlee, C. C. 2009. 'How race becomes biology: Embodiment of social inequality', *American Journal of Physical Anthropology* 139: 47–57.
Griesemer, J. R. 1990. 'Modeling in the Museum: On the Role of Remnant Models in the Work of Joseph Grinnell', *Biology and Philosophy* 5: 3-36.
Hacking, I. 1996. 'The Disunities of Science,' in P. Galison and D. Stump (eds), *The Disunity of Science: Boundaries, Contexts and Power.* Stanford: Stanford University Press, 37–74.
Hacking, I. 2002. *Historical Ontology*. Cambridge, MA: Harvard University Press.
Hacking, I. 2007. 'Kinds of People: Moving Targets', *Proceedings of the British Academy* 151: 285–318.
Hochman, A. (forthcoming). 'Against the new racial naturalism', *The Journal of Philosophy*.
Hull, D. 1988. *Science as a Process.* Chicago: University of Chicago Press.
Huson, D. H., R. Rupp and C. Scornavacca. 2010. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge: Cambridge University Press.
James, M. 2011. 'Race', *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), E. N. Zalta (ed.), <http://plato.stanford.edu/archives/win2012/entries/race/>.
Jobling, M. A. and C. Tyler-Smith. 2003. 'The human Y chromosome: an evolutionary marker comes of age', *Nature Reviews Genetics* 4: 598–612.
Jorde, L. B., W. S. Watkins, J. Bamshad, M. E. Dixon, C. E. Ricker, M. T. Seielstad and M. A. Batzer. 2000. 'The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data', *American Journal of Human Genetics* 66: 979–988.
Jost, L. 2008. 'GST and its relatives do not measure differentiation', *Molecular Ecology* 17: 4015–26.

Jost, L., P. DeVries, T. Walla, H. Greeney, A. Chao and C. Ricotta. 2010. 'Partitioning diversity for conservation analyses', *Diversity and Distributions* 16(1): 65–76.

Kaplan, J. M. 2010. 'When socially determined categories make biological realities: understanding Black/White health disparities in the U.S.', *Monist* 93: 281–97.

Kaplan, J. M. and R. G. Winther. 2012. 'Prisoners of abstraction? The theory and measure of genetic variation, and the very concept of "race"', *Biological Theory* 7(1). <http://philpapers.org/archive/KAPPOA.14.pdf>.

Keinan, A. and D. Reich. 2010. 'Human population differentiation is strongly correlated with local recombination rate', *PLoS Genetics* 6: e1000886 1–12.

Kuhn, T. S. [1962] 1970. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press.

Levins, R. and R. C. Lewontin. 1985. *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Lewontin, R. C. 1972. 'The apportionment of human diversity', *Evolutionary Biology* 6: 381–98.

Lewontin, R. C. and J. L. Hubby. 1966. 'A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*', *PubMed* 54: 595–609.

Li, W.-H. and L. A. Sadler. 1991. 'Low nucleotide diversity in man', *Genetics* 129: 513–23.

Livingstone, F. B. and T. Dobzshansky. 1962. 'On the non-existence of human races', *Current Anthropology* 3: 279–81.

López-Beltrán, C. (ed.). 2011. *Genes & mestizos: genómica y raza en la biomedicina Mexican*a. Mexico City: UNAM.

Love A. 2009. 'Typology Reconfigured: From the Metaphysics of Essentialism to the Epistemology of Representation', *Acta Biotheoretica* 57: 51-75.

Lowe, A., S. Harris and P. Ashton. 2004. *Ecological Genetics: Design, Analysis, and Application*. Oxford: Blackwell.

Mallon, R. 2006. 'Race: Normative, Not Metaphysical or Semantic', *Ethics* 116(3): 525–51.

Mills, C. 1998, *Blackness Visible: Essays on Philosophy and Race*, Ithaca, NY: Cornell University Press.

Mountain, J. L. and L. L. Cavalli-Sforza. 1997. 'Multilocus genotypes, a tree of individuals, and human evolutionary history', *American Journal of Human Genetics* 61: 705–18.

Nei, M. and A. K. Roychoudhury. 1972. 'Gene differences between Caucasian, Negro, and Japanese populations', *Science* 177: 434–6.

Nei, M. and A. K. Roychoudhury. 1993. 'Evolutionary relationships of human populations on a global scale', *Mol. Biol. Evol.* 10: 927–43.

Nelson, G., Platnick, N. 1981. *Systematics and Biogeography: Cladistics and Vicariance*. New York: Columbia University Press.

Oyama, S. 2000. *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd ed. Durham: Duke University Press.

Oyama, S., P. E. Griffiths and R. D. Gray. 2001. *Cycles of Contingency: Developmental Systems and Evolution.* Cambridge, MA: MIT Press.

Patterson, N., D. J. Richter, S. Gnerre, E. S. Lander and D. Reich. 2006. 'Genetic evidence for complex speciation of humans and chimpanzees', *Nature* 441: 1103–8.
Pigliucci, M. and J. Kaplan. 2003. 'On the concept of biological race and its applicability
to humans', *Philosophy of Science* 70: 1161–72.
Pritchard, J. K., M. Stephens and P. Donnelly. 2000. 'Inference of population structure using multilocus genotype data, *Genetics* 155: 945–59.
Relethford, J. H. 1998. 'Genetics of modern human origins and diversity', *Annual Review of Anthropology* 27: 1–23.
Rieppel, O. 1994. Homology, topology, and typology: the history of modern debates. In: Hall B.K. (ed) *Homology. The hierarchical basis of comparative biology*. Academic Press, San Diego: 63–100.
Risch, N., E. Burchard, E. Ziv and H. Tang. 2002. 'Categorization of humans in biomedical research: Genes, race and, disease', *Genome Biology* 3: 2007–12.
Rose, S., R. C. Lewontin and L. J. Kamin. 1984. *Not in Our Genes: Biology, Ideology, and Human Nature.* Harmondsworth, U.K.: Pelican Books.
Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky and M. A. Feldman. 2002. 'Genetic structure of human populations', *Science* 298: 2381–5.
Sarich, V. M. and A. C. Wilson. 1967. 'Immunological time scale for hominid evolution', *Science* 158: 1200–3.
Scharpf, C. 2000. Politics, Science, and the Fate of the Alabama Sturgeon. *American Currents* 26: 6-14.
Sesardic, N. 2010. 'Race: A social destruction of a biological concept', *Biology and Philosophy* 25(2):143–62.
Shapiro, M. D., Z. Kronenberg, C. Li, E. T. Domyan, H. Pan, M. Campbell, H. Tan, C. D. Huff, H. Hu, A. I. Vickrey, S. C. A. Nielsen, S. A. Stringham, H. Hu, E. Willerslev, M. T. P. Gilbert, M. Yandell, G. Zhang and J. Wang. 2013. 'Genomic Diversity and Evolution of the Head Crest in the Rock Pigeon', *Science* doi: 10.1126/science.1230422
Silva-Zolezzi, I., A. Hidalgo-Miranda, J. Estrada-Gil, J. C. Fernandez-Lopez, L. Uribe-Figueroa, A. Contreras, E. Balam-Ortiz, L. del Bosque-Plata, D. Velazquez Fernandez, C. Lara, R. Goya, E. Hernandez-Lemus, C. Davila, E. Barrientos, S. March and G. Jimenez-Sanchez. 2009. 'Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico', *Proceedings of the National Academy of Sciences* 106(21): 8611–6.
Smith, H. M., D. Chiszar and R. R. Montanucci. 1997. 'Subspecies and Classification', *Herpetological Review* 28: 13–16.
Sober, E. 2004. *Evolution and Evidence: The Logic Behind the Science*. Cambridge University Press.
Stringham, S. A., E. E. Mulroy, J. Xing, D. Record, M. W. Guernsey, J. T. Aldenhoven, E. J. Osborne and M. D. Shapiro. 2012. 'Genomic diversity and evolution of the head crest in the rock pigeon', *Current Biology* 22: 1–7.
Templeton, A. R. 1997. 'Out of Africa? What do genes tell us?' *Current Opinion in Genetics and Development* 7: 841–7.
Templeton, A. R. 1999. 'Human races: A genetic and evolutionary perspective', *American Anthropologist* 100(3): 632–50.

Teng, Y., Y. M. Luo, C. Y. Huang, J. Long, Z. G. Li and P. Christie. 2008. 'Tolerance of Grasses to Heavy Metals and Microbial Functional Diversity in Soils Contaminated with Copper Mine Tailings', *Pedosphere* 18(3): 363–70.

van Oven, M, Kayser, M. 2009. 'Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation', *Human Mutation* 20 (2): E386-94.

Weersing, K. and R. J. Toonen. 2009. 'Population genetics, larval dispersal, and connectivity in marine systems', *Marine Ecology Progress Series* 393: 1–12.

Wells, S. 2003. *The Journey of Man: A Genetic Odyssey*. New York: Random House.

Winther, R. G. 2009. 'Character analysis in cladistics: Abstraction, reification, and the search for objectivity', *Acta Biotheoretica* 57: 129–62.

Winther, R. G. 2011. '¿La cosificación genética de la "raza"? Un análisis crítico,' in C. López-Beltrán (ed.), *Genes & mestizos: genómica y raza en la biomedicina Mexican*a. Mexico City: UNAM, 237–58. <http://philpapers.org/rec/WINLCG>.

Winther, R. G. 2012. 'Interweaving categories: Styles, paradigms, and models', *Studies in History and Philosophy of Science, Part A* 43: 628–39.

Winther, R. G. 2013. 'Evo-Devo as a Trading Zone', in A. Love (ed.), *Conceptual Change in Biology: Scientific and Philosophical Perspectives on Evolution and Development*. Springer Verlag (Boston Studies in the Philosophy of Science); 20 pp. http://philpapers.org/rec/WINEAA-4

Yamamichi, M., J. Gojobori and H. Innan. 2012. 'An autosomal analysis gives no genetic evidence for complex speciation of humans and chimpanzees', *Molecular Biology and Evolution* 29: 145–56.

Yu, N., F.-C. Chen, S. Ota, L. B. Jorde, P. Pamilo, L. Patthy, M. Ramsay, T. Jenkins, S.-K. Shyue and W.-H. Li. 2002. 'Larger genetic differences within Africans than between Africans and Eurasians', *Genetics* 161: 269–74.

Zack, N. 2002. *Philosophy of Science and Race*. New York: Routledge.

Zerubavel, E. 1996. 'Lumping and splitting: Notes on social classification', *Sociological Forum* 11: 421–33.

Zuckerkandl, E. and L. Pauling. 1965. 'Molecules as documents of evolutionary history', *Journal of Theoretical Biology* 8: 357–360.