Original Paper

Human
Heredity

Hum Hered 2016;82:87–102
DOI: 10.1159/000478897

Received: January 4, 2017
Accepted: June 14, 2017
Published online: September 15, 2017

Consanguinity Rates Predict Long Runs of Homozygosity in Jewish Populations

Jonathan T.L. Kang[a] Amy Goldberg[a] Michael D. Edge[a] Doron M. Behar[b, c]
Noah A. Rosenberg[a]

[a]Department of Biology, Stanford University, Stanford, CA, USA; [b]Estonian Biocentre and Department of Evolutionary Biology, University of Tartu, Tartu, Estonia; [c]Clalit National Personalized Medicine Program, Department of Community Medicine and Epidemiology, Carmel Medical Center, Haifa, Israel

Keywords
Heterozygosity · Identity by descent · Inbreeding coefficient

Abstract
Objectives: Recent studies have highlighted the potential of analyses of genomic sharing to produce insight into the demographic processes affecting human populations. We study runs of homozygosity (ROH) in 18 Jewish populations, examining these groups in relation to 123 non-Jewish populations sampled worldwide. ***Methods:*** By sorting ROH into 3 length classes (short, intermediate, and long), we evaluate the impact of demographic processes on genomic patterns in Jewish populations. ***Results:*** We find that the portion of the genome appearing in long ROH – the length class most directly related to recent consanguinity – closely accords with data gathered from interviews during the 1950s on frequencies of consanguineous unions in various Jewish groups. ***Conclusion:*** The high correlation between 1950s consanguinity levels and coverage by long ROH explains differences across populations in ROH patterns. The dissection of ROH into length classes and the comparison to consanguinity data assist in understanding a number of additional phenomena, including similarities of Jewish populations to Middle Eastern, European, and Central and South Asian non-Jewish populations in short ROH patterns, relative lengths of identity-by-descent tracts in different Jewish groups, and the “population isolate” status of the Ashkenazi Jews.

Introduction

Genome-based analysis of genetic sharing within and between individuals and the use of dense genomic polymorphism data in the direct evaluation of identity by descent (IBD) have provided powerful techniques for enabling advances in human genetics – on problems such as relatedness estimation, inference of population relationships, haplotype phasing and imputation, and various aspects of the mapping of disease-related alleles [1, 2].

Runs of homozygosity (ROH), describing IBD for the two genomic copies possessed by a single diploid individual, represent a particularly informative type of genomic sharing. Because genomic sharing in an individual can result from processes taking place on different time scales, ROH both catalog haplotype homozygosity resulting from shared descent of two parents from the limited number of ancestors who underwent ancient population

KARGER

E-Mail karger@karger.com
www.karger.com/hhe

Jonathan T.L. Kang
Department of Biology, Stanford University
371 Serra Mall
Stanford, CA 94305 (USA)
E-Mail jtlkang@stanford.edu

migrations and record consanguineous unions in the recent ancestors of individuals. ROH studies have been used to measure inbreeding in individuals and populations [3–5], to investigate influences of the features of population history on genetic variation among populations [6–8], as well as to test for influences of genomic homozygosity on phenotypes [9–13].

Levels of homozygosity vary by population as a result of the differing descent of different populations from the ancient migration events that have led to elevated homozygosities. Consequently, Pemberton et al. [8] developed a population-wise method for identifying segments that are sufficiently long to represent ROH. They devised a model-based clustering scheme that partitions the ROH of a population into 3 classes: short ROH, resulting from the pairing of ancient haplotypes; intermediate ROH, largely reflecting cryptic relatedness within populations or groups of populations; and long ROH, indicating recent consanguinity. This subdivision clarifies that multiple forces underlie the observation that high fractions of the genome lie in ROH in a variety of populations. For example, ancient bottlenecks in some Native American populations generate many "short" ROH, and recent consanguinity produces many "long" ROH in some populations of the Middle East. The ternary system of ROH classification has also been employed in analyzing the distribution of deleterious variants among ROH belonging to each of the 3 classes [14] and in detecting ROH of different classes from whole-exome sequencing data [15].

In Jewish populations, studies of genomic sharing, primarily in the form of IBD analyses within and between populations, have produced 3 consistent patterns [16]. First, high levels of IBD sharing between Jewish groups have supported the existence of a component of shared ancestry for Jewish groups in distant locations [17–21]. Second, it has been observed that Jewish groups often have higher levels of within-group IBD sharing than nearby non-Jewish groups [18, 20, 22–24]. Third, studies have noted that Jewish groups vary considerably in their levels of within-group IBD sharing [17, 18, 20, 21].

Here, we investigate ROH in Jewish populations, considering the extra information about consanguinity available from ROH – which examine the two haplotypes of an individual – compared to IBD calculations between individuals or populations. We make use of a remarkable demographic data set on consanguinity collected in the 1950s from many of the groups that we study [25, 26]. By relating ROH to demographic data on consanguinity, we find that the level of consanguinity measured in the populations is predictive of long ROH – both affirming the value of subdividing ROH into length classes and recording genetic evidence of consanguinity practices that existed during the 1950s. The results also contribute insight into the patterns observed in IBD studies in Jewish populations.

Methods

Genotype Data Processing

We assembled a data set of single nucleotide polymorphism (SNP) variants that combines information from two sources. The first is the data of Behar et al. [19] on 1,572 individuals from 89 non-Jewish populations originating from Africa, Asia, and Europe, and 202 individuals from 18 widely dispersed Jewish populations. It contains genotype information at 270,898 SNPs. We obtained a count of 89 non-Jewish populations instead of the 88 reported by Behar et al. [19] as we separate two Bantu populations that they grouped together. The second source consists of the combination of the HGDP-CEPH and HapMap III data sets studied by Verdu et al. [27]. It contains 2,055 non-Jewish individuals (938 HGDP-CEPH and 1,117 HapMap III) from 64 worldwide populations with genotypes at 590,461 SNPs.

We merged the two data sets as follows:

1. First, we identified the 32 populations containing exact duplicates of individuals present in both the Behar et al. [19] and Verdu et al. [27] data sets: 31 HGDP-CEPH populations and the HapMap III Gujarati population. For each duplicate pair, one duplicate was removed.
2. In 2 of the 31 HGDP-CEPH populations with duplicate individuals (Palestinian and Druze), Behar et al. [19] also included individuals that did not originate from HGDP-CEPH. These individuals were retained, but they were treated as belonging to populations separate from the corresponding HGDP-CEPH populations (annotated 1 for Verdu et al. [27], 2 for Behar et al. [19]).
3. Two more populations (Russian and Mongolian) appeared in both Behar et al. [19] and Verdu et al. [27], but with no overlap of individuals across the data sets. In these cases, all individuals were retained, but for each pair of corresponding samples, the two samples were treated as separate (1 for Verdu et al. [27], 2 for Behar et al. [19]).
4. Extensive quality control was performed in assembly of the Behar et al. [19] and Verdu et al. [27] data sets from raw genotype data. We retained the SNPs shared by both sources, discarding SNPs present in only one of the data sets. At 757 SNPs, the data sets had genotypes given for opposite strands, and we converted the Behar et al. [19] genotypes to match those from Verdu et al. [27].

After processing, the merged data set consists of 3,105 individuals from 141 populations, 123 non-Jewish and 18 Jewish, genotyped at 257,091 SNPs. We classified non-Jewish populations into geographic regions: Sub-Saharan Africa, the Middle East (together with North Africa), Europe, the Caucasus region, Central and South Asia, East Asia, Oceania, the Americas, and Admixed, containing African-American and Mexican-American samples (Table 1).

Table 1. Sample sizes and population groupings for the 141 populations in this study

a 123 non-Jewish populations

Population	Sample size	Source
Africa		
Bantu (Kenya)	11	[27]
Bantu (S. Africa)	8	[27]
Biaka Pygmy	22	[27]
Ethiopian	19	[19]
Luhya (LWK)	99	[27]
Mandenka	22	[27]
Maasai (MKK)	105	[27]
Mbuti Pygmy	13	[27]
San	5	[27]
Yoruba	21	[27]
Yoruba (YRI)	140	[27]
Middle East		
Bedouin	45	[27]
Cypriot	12	[19]
Druze 1	42	[27]
Druze 2	3	[19]
Egyptian	12	[19]
Iranian	19	[19]
Jordanian	20	[19]
Kurd	6	[19]
Lebanese	8	[19]
Moroccan	10	[19]
Mozabite	27	[27]
Palestinian 1	46	[27]
Palestinian 2	6	[19]
Samaritan	3	[19]
Saudi	20	[19]
Syrian	16	[19]
Turkish	19	[19]
Yemeni	8	[19]
Europe		
Abruzzo	11	[19]
Basque	24	[27]
Belarusian	17	[19]
Bulgarian	13	[19]
Caucasian (CEU)	112	[27]
Chuvash	19	[19]
Croat	24	[19]
Estonian	15	[19]
French	28	[27]
Greek	20	[19]
Hungarian	19	[19]
Italian	12	[27]
Lithuanian	10	[19]
Moldavian	7	[19]
Mordovian	15	[19]
Orcadian	15	[27]
Polish	17	[19]
Romanian	16	[19]
Russian 1	25	[27]
Russian 2	23	[19]
Sardinian	28	[27]
Sicilian	13	[19]
Spanish	12	[19]
Swedish	18	[19]
Tatar	20	[19]
Toscani (TSI)	102	[27]
Tuscan	7	[27]
Ukranian	20	[19]
Caucasus		
Abkhasian	23	[19]
Adygei	17	[27]
Armenian	16	[19]
Azeri	16	[19]
Balkar	22	[19]
Chechen	20	[19]
Georgian	30	[19]
Kabardin	3	[19]
Kumyk	17	[19]
Lezgin	21	[19]
Nogai	16	[19]
North Ossetian	18	[19]
Tabasaran	3	[19]
Central/South Asia		
Balochi	24	[27]
Brahui	25	[27]
Burusho	25	[27]
Gujarati (GIH)	97	[27]
Halakipikki	4	[19]
Hazara	22	[27]
Kalash	23	[27]
Kyrgyz	21	[19]
Makrani	25	[27]
Malayan	2	[19]
North Kannadi	9	[19]
Paniya	4	[19]
Pathan	22	[27]
Sakilli	4	[19]
Sindhi	24	[27]
Tajik	15	[19]
Turkmen	20	[19]
Uyghur	10	[27]
Uzbek	19	[19]
East Asia		
Altaian	13	[19]
Buryat	18	[19]
Cambodian	10	[27]
Chinese (CHD)	106	[27]
Dai	10	[27]
Daur	9	[27]
Han	34	[27]
Han (CHB)	137	[27]
Han (N. China)	10	[27]
Hezhen	9	[27]

DOI: 10.1159/000478897

Table 1 (continued)

Population	Sample size	Source
Japanese	28	[27]
Japanese (JPT)	113	[27]
Lahu	8	[27]
Miao	10	[27]
Mongola 1	10	[27]
Mongolian 2	9	[19]
Naxi	8	[27]
Oroqen	9	[27]
She	10	[27]
Tu	10	[27]
Tujia	10	[27]
Tuvinian	15	[19]
Xibo	9	[27]
Yakut	25	[27]
Yi	10	[27]
Oceania		
Melanesian	11	[27]
Papuan	17	[27]
Americas		
Colombian	7	[27]
Karitiana	13	[27]
Maya	21	[27]
Pima	14	[27]
Surui	8	[27]
Admixed		
African (ASW)	52	[27]
Mexican (MXL)	54	[27]

b 18 Jewish populations

Algerian Jewish	North African	5	[19]
Ashkenazi Jewish	European	29	[19]
Azerbaijani Jewish	Middle Eastern	11	[19]
Cochin Jewish	South Asian	7	[19]
Ethiopian Jewish	Ethiopian	15	[19]
Georgian Jewish	Middle Eastern	7	[19]
Iranian Jewish	Middle Eastern	12	[19]
Iraqi Jewish	Middle Eastern	13	[19]
Italian Jewish	European	10	[19]
Kurdish Jewish	Middle Eastern	10	[19]
Libyan Jewish	North African	6	[19]
Moroccan Jewish	North African	18	[19]
Mumbai Jewish	South Asian	6	[19]
Sephardi Jewish	European	22	[19]
Syrian Jewish	Middle Eastern	2	[19]
Tunisian Jewish	North African	6	[19]
Uzbekistani Jewish	Middle Eastern	5	[19]
Yemenite Jewish	Yemenite	18	[19]

We classified the 18 Jewish populations into 6 regional groups, following Behar et al. [19]:

1. European (Ashkenazi, Italian, and Sephardi);
2. Middle Eastern (Azerbaijani, Georgian, Iranian, Iraqi, Kurdish, Syrian, and Uzbekistani);
3. North African (Algerian, Libyan, Moroccan, and Tunisian);
4. South Asian (Cochin and Mumbai);
5. Ethiopian;
6. Yemenite.

The Middle Eastern Jewish group accords with the group termed "Mizrahi" or "Oriental" elsewhere. Note that the regional groups for the Jewish populations do not necessarily map onto single geographic regions among those used for the non-Jewish populations.

Identification of ROH

Within individual genomes, we identified ROH and classified them by size according to the procedure of Pemberton et al. [8]. For each population, we estimated the allele frequencies at each SNP by sampling 40 alleles without replacement, calculating the allele frequencies from the sampled alleles. This resampling procedure is performed to account for sample size differences across populations (Table 1).

Next, to identify ROH, we employed a likelihood approach from Wang et al. [28] adapted by Pemberton et al. [8]. This approach considers a sliding window of n SNPs that moves along the chromosome with an increment of m SNPs. Because our SNP density was approximately half that of Pemberton et al. [8] (257,091 compared to 577,489), we chose $(n, m) = (30, 1)$, in contrast to (60, 1) in Pemberton et al. [8]. By halving n, we arrange for the windows to contain comparably many base pairs to those used by Pemberton et al. [8].

Following Pemberton et al. [8], the strength of autozygosity for a window is quantified by a log-likelihood (LOD) score comparing the hypothesis that the segment is autozygous to the hypothesis that it is non-autozygous, allowing for an error term that accommodates genotyping error or mutation within autozygous regions. As in Pemberton et al. [8], we set the error parameter to 0.001. For each population, we obtained the LOD score distribution across all windows in all individuals, using the "density" function in R with a Gaussian kernel and default nrd0 bandwidth.

As in Pemberton et al. [8], the LOD score distributions have two modes. The locations of these modes differ by population, and for each population, we followed Pemberton et al. [8] in using the local minimum between the modes as the ROH threshold. All windows whose LOD score exceeded the population-specific threshold were taken to be homozygous, with contiguous windows joined and considered as part of a single ROH.

Size Classification of ROH

The length of each SNP window determined to be an ROH was recorded as the length of the interval between its two most extreme SNPs, including the endpoints. Again following Pemberton et al. [8], separately in each population, we modeled the ROH length distribution as a mixture of 3 Gaussian distributions representing 3 ROH classes: (A) short ROH measuring tens of kb, (B) intermediate ROH measuring hundreds of kb to a few Mb, and (C) long ROH measuring multiple Mb. Unsupervised 3-component Gaussian fitting was performed population-wise, using the Mclust function from the mclust package in R, and allowing component proportions, means, and variances to be free variables.

For each population, let A_{min} and A_{max} be minimum and maximum ROH lengths classified as belonging to Class A, and define B_{min}, B_{max}, C_{min}, and C_{max} analogously. The boundary between Classes A and B is given by $(A_{max} + B_{min})/2$, and the boundary between Classes B and C by $(B_{max} + C_{min})/2$. Across all populations, the A–B boundaries lie in the range [421,410.5 bp, 686,103 bp], with mean 504,952 bp, and standard deviation 37,451 bp. The B–C boundaries lie in the range [1,343,237 bp, 2,325,452 bp], with mean 1,711,184 bp, and standard deviation 159,590 bp. Thus, the class boundaries vary across populations, but with all A–B boundaries strictly below all B–C boundaries, so that the classes are clearly delineated.

Demographic Data on Jewish Patterns of Consanguinity

We use demographic data reported by Goldschmidt et al. [25] on the rate of consanguineous unions in different Jewish populations in Israel during 1955–1957. Goldschmidt et al. [25] surveyed 11,424 mothers of newborn babies in maternity wards of 8 hospitals in Haifa, Jerusalem, and Tel Aviv, recording data on the unions represented by the parents of the newborns. Among unions classified as consanguineous, 3 further subdivisions were employed: "first cousins," "uncle–niece," and "more distant relationships."

Nine Jewish populations appear in both our genotype data and the demographic data from Goldschmidt et al. [25]: Ashkenazi, Iranian, Iraqi, Libyan, Moroccan, Sephardi, Syrian, Tunisian, and Yemenite. The Jewish population labeled by Behar et al. [19] as "Iranian" corresponds to the Persian population of Goldschmidt et al. [25]. We treated the "Sephardi" population (Behar et al. [19]) as commensurable with the Turkish population of Goldschmidt et al. [25], as the Sephardi sample in Behar et al. [19] was largely from the Turkish Jewish population.

For each Jewish group, we estimated the overall inbreeding coefficient by weighting the percentages of the population in each of the 3 consanguinity classes by their associated inbreeding coefficients. For first cousins, this inbreeding coefficient is 1/16; for uncle–niece unions, it is 1/8. For consanguineous unions that are more distant than first cousins, we assigned a value of 1/32. For non-consanguineous unions, we assigned a value of 0.

Results

Jewish ROH Lengths in the Context of Worldwide Populations

We first examined the ROH in Jewish populations in relation to those seen in other populations. Summing ROH lengths across the genome, we evaluated, within individuals, the total length of all ROH and the total length of ROH in each length class.

Across all ROH, the worldwide pattern refines the pattern found in Pemberton et al. [8], with an increase in individual-level total ROH length with increasing distance of populations from Sub-Saharan Africa (Fig. 1D). The Jewish populations have similar total ROH lengths to non-Jewish populations from the Middle East, Europe, the Caucasus, and Central and South Asia. The high variability across individuals in the total ROH length seen within Jewish populations is also observed elsewhere, most frequently in the Middle Eastern, Central and South Asian, and Native American populations.

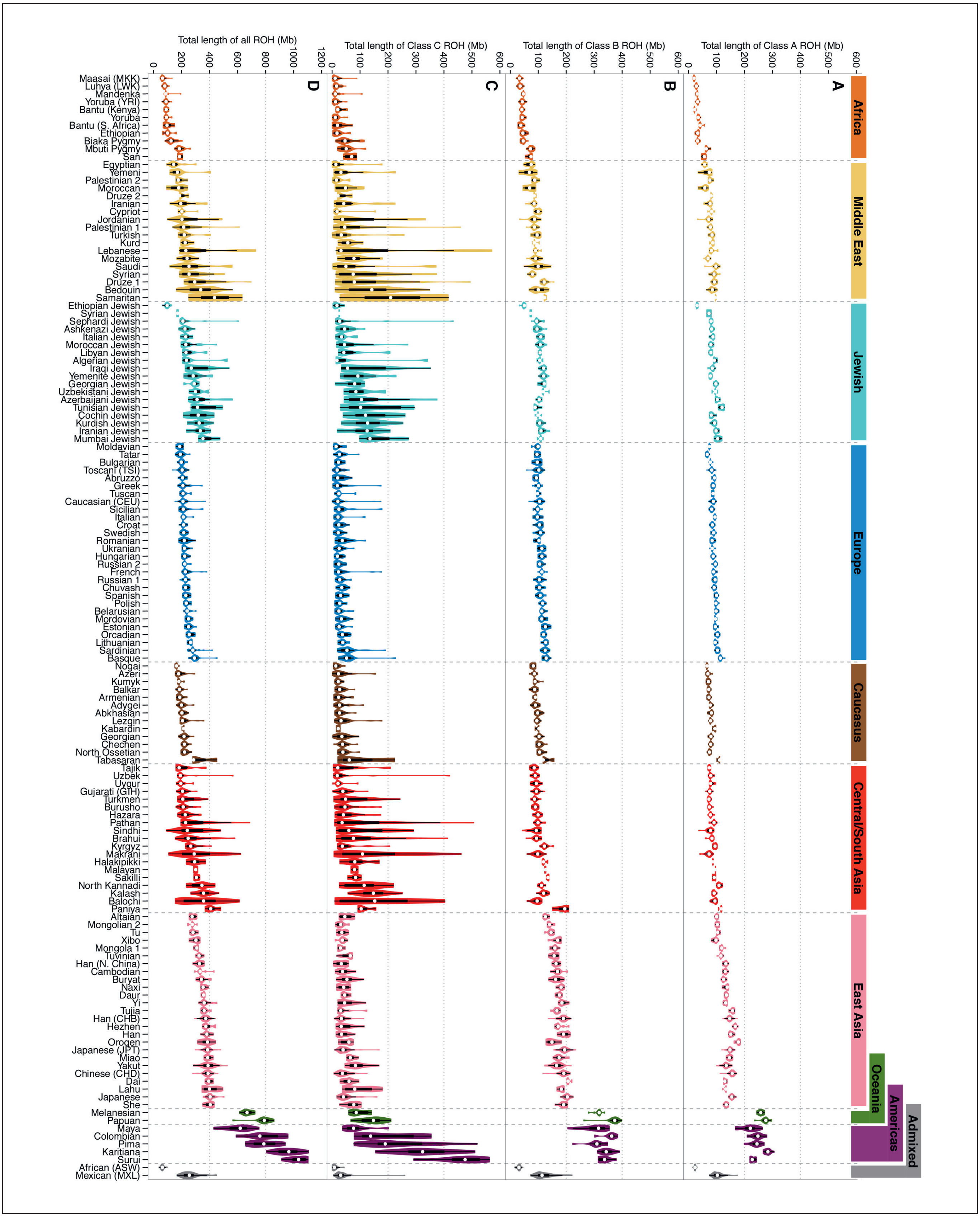
As in Pemberton et al. [8], the median length in an individual's genome that lies in the shorter Class A and Class B ROH increases stepwise with distance from Africa in successive continental groups (Fig. 1A, B). For Class A ROH in particular, Jewish populations have distributions comparable to the Middle East, Europe, the Caucasus, and Central and South Asia (Fig. 1A, Fig. 2A). Permutation tests for a difference between a pair of population groups in the median across populations of the median ROH length across individuals – permuting group memberships and recomputing the absolute difference between group medians – confirm this observation, as low p values, indicating a significant absolute difference from the Jewish populations in Class A ROH length, do not occur for these regions (Table 2).

Unlike Class A and Class B ROH, which largely follow distance from Africa, Class C ROH lengths in non-Jewish populations have the highest values in the Middle East, Central and South Asia, and the Americas (Fig. 1C, Fig. 2C). As was noted by Pemberton et al. [8], individuals from these regions often possess high degrees of recent parental relatedness. After 2 Native American populations, the highly consanguineous Samaritan population isolate [29] has the highest median Class C ROH length. A number of Jewish populations, including the Mumbai, Kurdish, Iranian, Cochin, and Azerbaijani groups, have particularly long Class C ROH. Considerable variability in the pattern of Class C ROH exists across Jewish populations, with comparable variation across populations to that seen in non-Jewish populations of the Middle East and Central and South Asia (Fig. 1C).

ROH Lengths among Jewish Populations

To compare ROH patterns across Jewish populations in more detail, we considered only the Jewish samples, reporting in Figure 3 the relationship between ROH lengths in pairs of classes. For ROH lengths in Classes A and B, Figure 3A suggests that, for Jewish groups, a correlation is largely due to the Ethiopian Jewish population, the only Jewish group with substantial recent Sub-Saharan African ancestry ($r = 0.646$, $p < 2 \times 10^{-6}$ including Ethiopian Jews; $r = 0.051$, $p = 0.486$ excluding them). At the worldwide level, the major factor that drives strong correlations between Class A and Class B ROH (Fig. 4) is high variability across continental regions in the residual signal of ancient migrations outward from Sub-Saharan

1



(For legend see next page.)

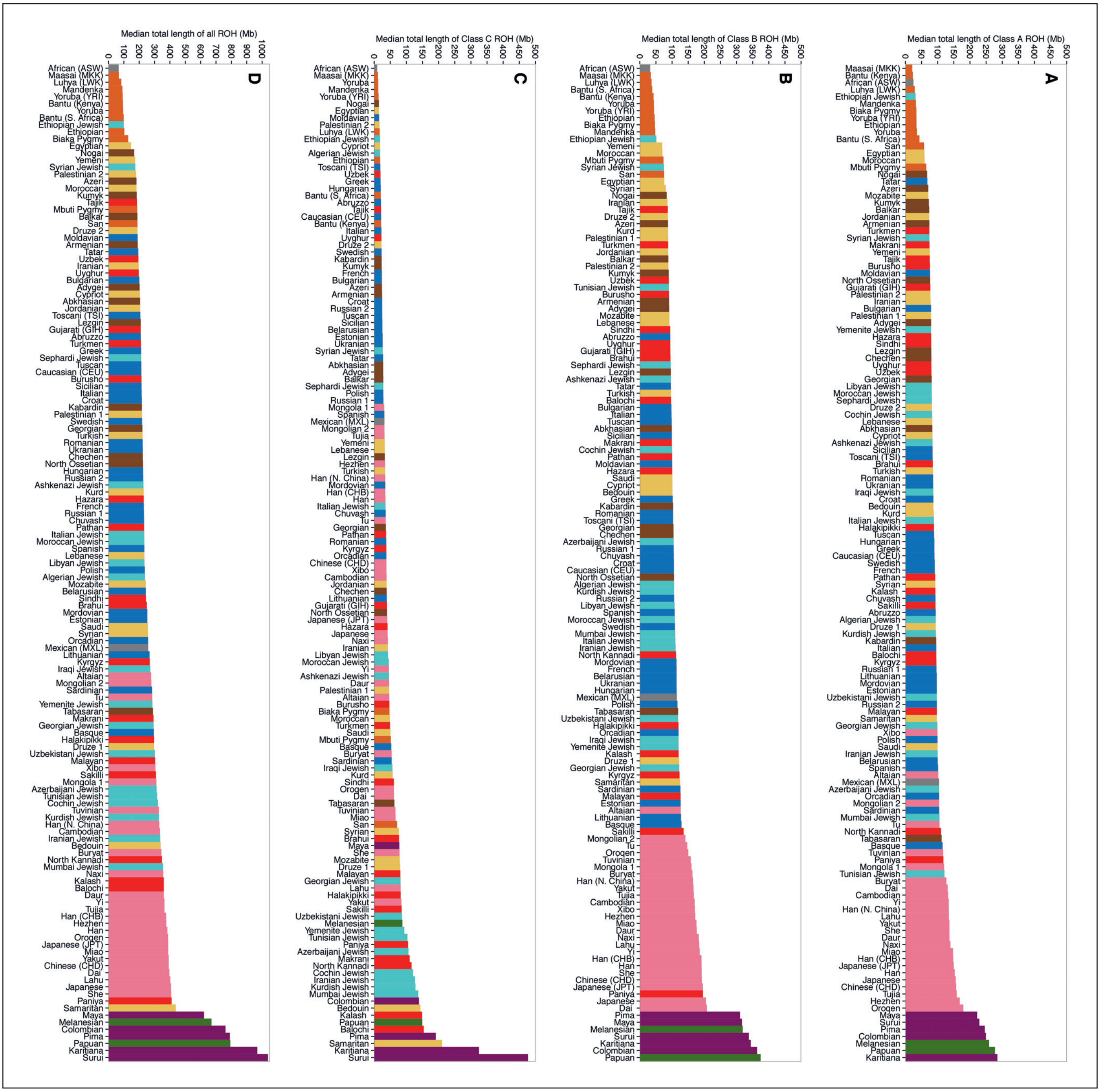


Fig. 2. Population-specific median ROH lengths across individuals. **A** Class A ROH. **B** Class B ROH. **C** Class C ROH. **D** All ROH. Populations are listed in increasing order of their values. Bars are colored according to regional groupings, following Figure 1. The figure shows an alternative visualization of the median from Figure 1.

Fig. 1. Population-specific distributions of ROH length across individuals. **A** Class A ROH. **B** Class B ROH. **C** Class C ROH. **D** All ROH. Each distribution is shown as a violin plot, with the width depicting a kernel density trace and its reflection. A box plot is embedded in each violin plot. The white dot is the median of the distribution. Populations are ordered by regional groupings, and within groups by median total ROH length.

Table 2. *p* values from permutation tests of equality of the median ROH lengths between Jewish and non-Jewish populations

Population group	ROH			
	Class A	Class B	Class C	All
Africa	0.00011	0.00002	0.00370	0.00005
Middle East	0.11594	0.00076	0.04278	0.01085
Europe	0.51785	0.70336	<0.00001	0.00005
Caucasus	0.05814	0.01505	0.00289	0.00182
Central/South Asia	0.79118	0.24431	0.72541	0.62450
East Asia	<0.00001	0.00004	0.00109	0.00023
Oceania	0.00566	0.00566	0.25843	0.00566
Americas	0.00189	0.00172	0.00500	0.00188
Admixed	0.11054	0.08186	0.13437	0.04625

For each ROH class and each non-Jewish population group, we determined the median ROH length across individuals for each population in the group. The absolute difference of the median of these values across populations from the corresponding median ROH length across Jewish populations was then calculated. The Jewish/non-Jewish labels were permuted among the populations, and the number of permutations for which the permuted absolute difference was greater than or equal to the unpermuted absolute difference was tabulated. The Ethiopian Jewish population was excluded from the Jewish sample for these computations. The number of permutations was 100,000.

Africa, whose effects contribute similarly to both classes [8]. With the exception of the Ethiopian Jews, Jewish populations trace to regions at comparable continental locations in terms of distance from Africa, so that continental differences that give rise to the correlation between Class A and Class B are largely absent.

Figure 3B, C show that for Jewish samples, neither Class A nor Class B ROH is strongly correlated with Class C ROH ($r = 0.243$ and $r = 0.183$ for the correlations of A and C and of B and C with the Ethiopian Jewish population included, $r = 0.099$ and $r = 0.006$ for the corresponding calculations excluding it), in the same way that patterns in Class C ROH differ from those seen for Class A and Class B ROH worldwide (see Figure 1). The Jewish samples with elevated Class C ROH lengths originate mostly from the Middle Eastern and South Asian regional groups, where nearby non-Jewish Middle Eastern and Central and South Asian populations often have relatively high levels of Class C ROH as well.

Excluding the Ethiopian Jewish population, observed heterozygosity is strongly negatively correlated with the total length of all ROH (Fig. 5D, $r = -0.962$, $p < 2 \times 10^{-16}$). Unlike in a worldwide analysis, in which the relationship with observed heterozygosity of total length in all ROH is more tightly connected to Class A and Class B ROH than to Class C (Fig. 6), in the Jewish samples, this correlation is driven primarily by Class C (Fig. 5C, $r = -0.961$, $p < 2 \times 10^{-16}$). The magnitudes of the correlations with observed heterozygosity are lower for Class A and Class B ROH lengths (Fig. 5A, B, $r = -0.137$ and $r = -0.114$, respectively). The pattern further indicates that other than for the Ethiopian Jews, differing ROH patterns across Jewish populations are attributable mainly to differences in Class C ROH lengths – and hence, to underlying consanguinity differences – rather than to differences in ROH of Classes A and B.

ROH Lengths and Consanguinity in Jewish Populations

For the Jewish populations, the data of Goldschmidt et al. [25] provide direct measurements of consanguinity. Therefore, with the aim of studying the relationship between demographic and genetic measures of consanguinity, we examined ROH lengths in Jewish populations in relation to the consanguinity rates reported by Goldschmidt et al. [25]. From demographic consanguinity data, we estimated a population inbreeding coefficient for the 9 Jewish populations that are also present in our genotype data. Consanguinity rates from Goldschmidt et al. [25] are reproduced in Table 3, which also includes the associated inbreeding coefficients.

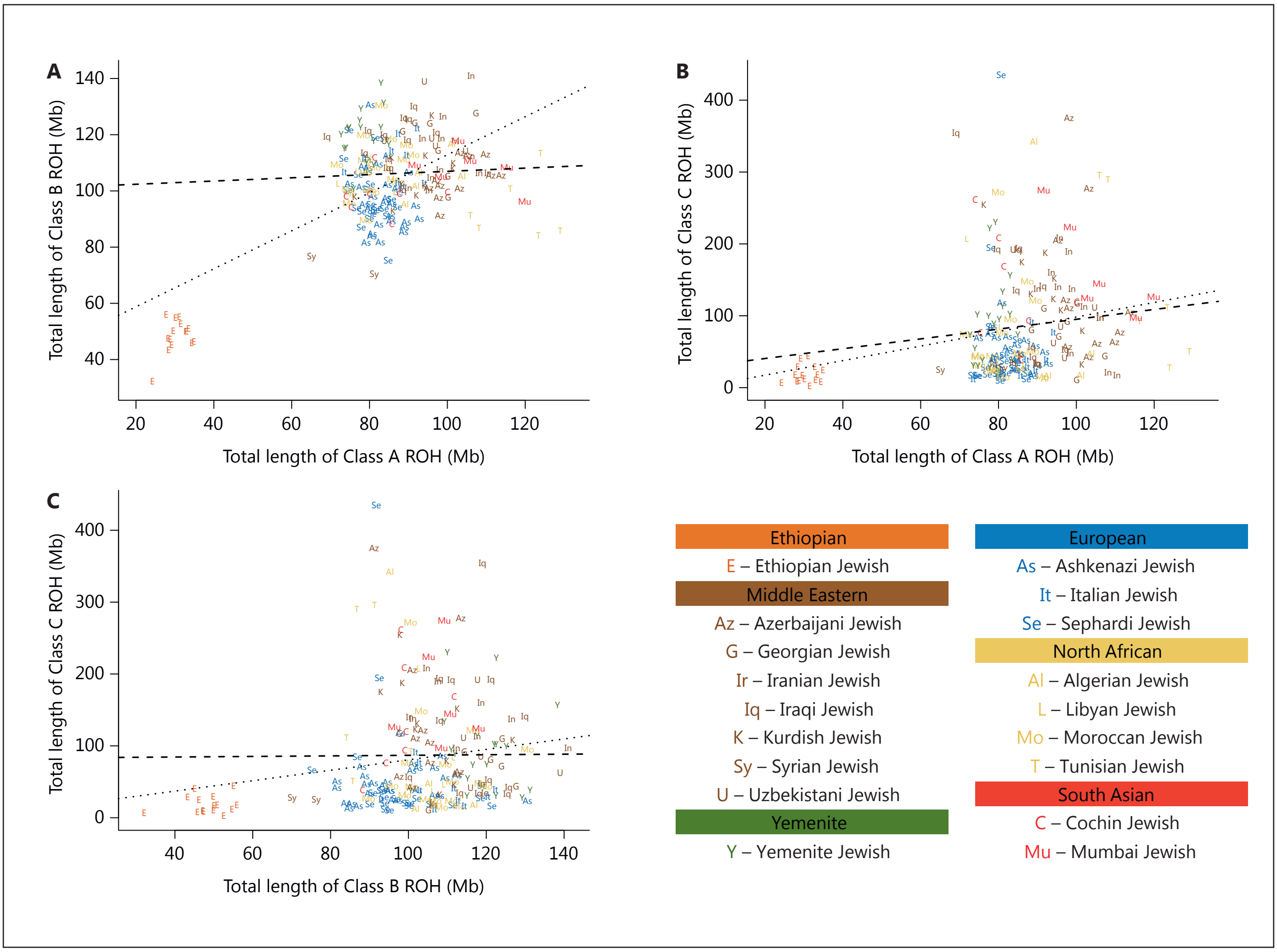


Fig. 3. Individual-level ROH lengths in pairs of classes, considering individuals in Jewish populations. **A** Class B versus Class A. **B** Class C versus Class A. **C** Class C versus Class B. The dotted regression lines include the Ethiopian Jewish samples, and the dashed regression lines exclude them.

Figure 7 examines the relationship between ROH lengths in the 9 Jewish populations and consanguinity-based inbreeding coefficients. We observe a positive correlation between the inbreeding coefficient of a population and the mean total ROH length of its constituent individuals (Fig. 7D). A regression slope of 92.06 indicates that each 1% increase in the inbreeding coefficient contributes 92.06 Mb to the total ROH length, and the high correlation coefficient of $r = 0.762$ between the mean total ROH length and the inbreeding coefficient has $p = 0.017$. In considering ROH classes separately, we see that Class C ROH length is the most important contributor to this relationship, with both the greatest slope and the largest correlation coefficient (Fig. 7C, slope = 61.42, $r = 0.765$, $p = 0.016$); values for Class A (Fig. 7A, slope = 12.10, $r = 0.418$, $p = 0.263$) and Class B (Fig. 7B, slope = 18.54, $r = 0.621$, $p = 0.074$) are positive for both the slope and the correlation coefficient but smaller. The stronger relationship between Class C ROH and the inbreeding coefficient, which is compatible with the view of Class C ROH as reflecting recent consanguinity, is robust to different assumptions regarding the appropriate choice of inbreeding coefficient for relationships more distant than first cousins (Table 4). Note that the intercepts in Figure 7 are non-zero as even with no consanguinity in recent generations, inbreeding in earlier generations produces ROH. Because recent consanguinity is connected primarily to Class C ROH, the intercept is lowest for Class C.

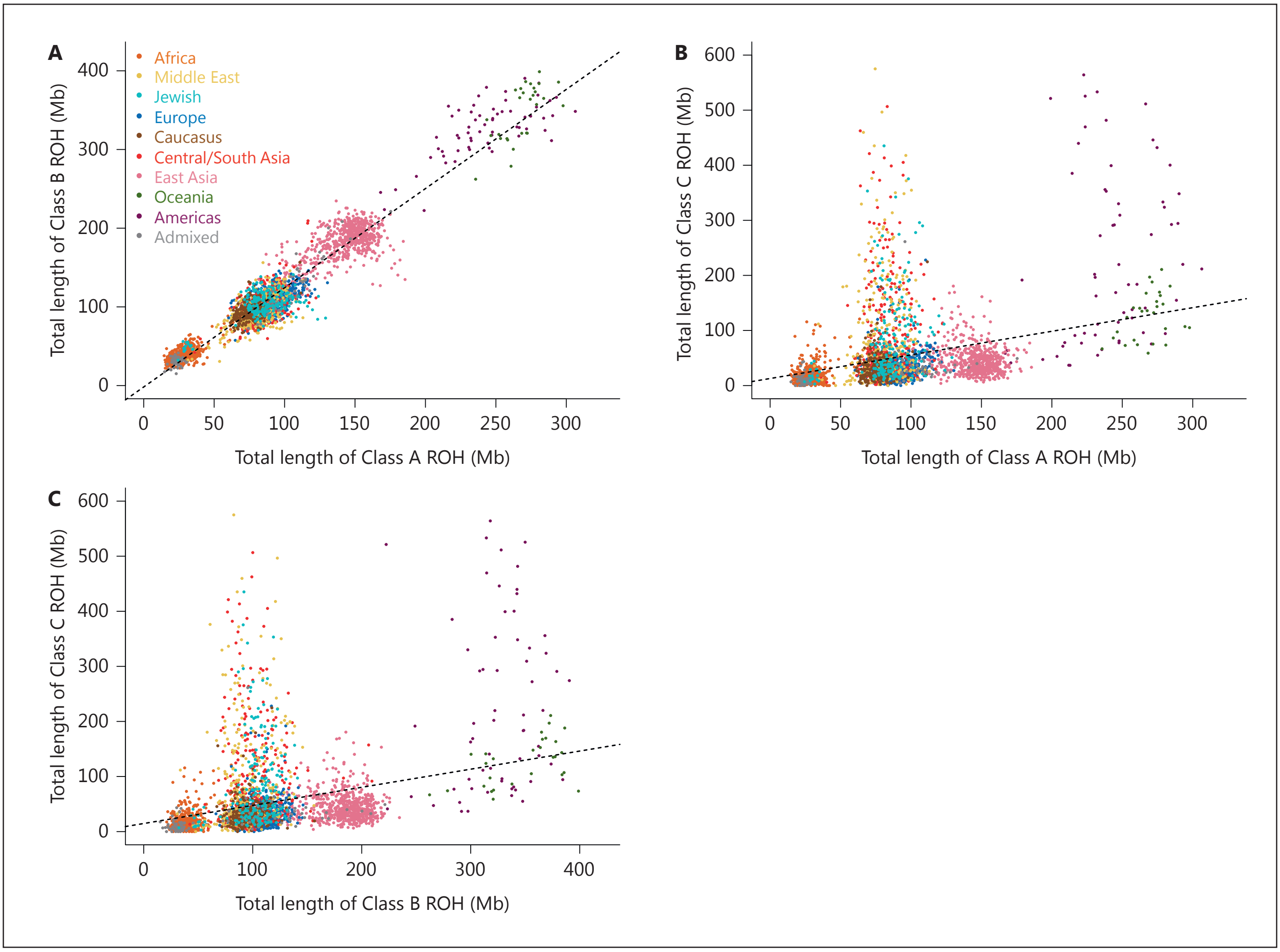


Fig. 4. Individual-level ROH lengths in pairs of classes. **A** Class B versus Class A. **B** Class C versus Class A. **C** Class C versus Class B. The lines represent regression lines. Colors follow those in Figure 1. Across individuals, ROH lengths of Classes A and B are highly correlated with each other ($r = 0.963$, $p < 2 \times 10^{-16}$), but they are less correlated with Class C ($r = 0.297$ for Classes A and C, $p < 2 \times 10^{-16}$; $r = 0.299$ for Classes B and C, $p < 2 \times 10^{-16}$). This pattern of correlation reflects the fact that ROH of Classes A and B have been produced by a different process from the process that generates Class C ROH: the former by population-level linkage disequilibrium on longer evolutionary time scales, and the latter by recent consanguinity [8].

Discussion

We have analyzed ROH in Jewish populations in relation to ROH in other populations. Short and intermediate ROH in Jewish groups, largely representing autozygosity for haplotypes that trace to ancient migration events, follow patterns seen in other groups from Europe and regions of Asia historically inhabited by Jewish populations (Fig. 1A, B). Long ROH, however, indicating recent parental relatedness, occupy more of the genome in the Jewish populations than in most groups (Fig. 1C), and they drive the differences among Jewish populations in total ROH levels (Fig. 3, 5). Many Jewish populations, including the Azerbaijani, Cochin, Georgian, Iranian, Mumbai, Tunisian, Uzbekistani, and Yemenite populations, have some of the highest proportions of their genomes in long ROH among populations worldwide, comparable to many non-Jewish populations of the Middle East and Central and South Asia, and exceeding non-Jewish European and African groups (Fig. 2C). These high proportions of long ROH accord with demographic data that also identify high consanguinity levels in various Jewish populations (Fig. 7).

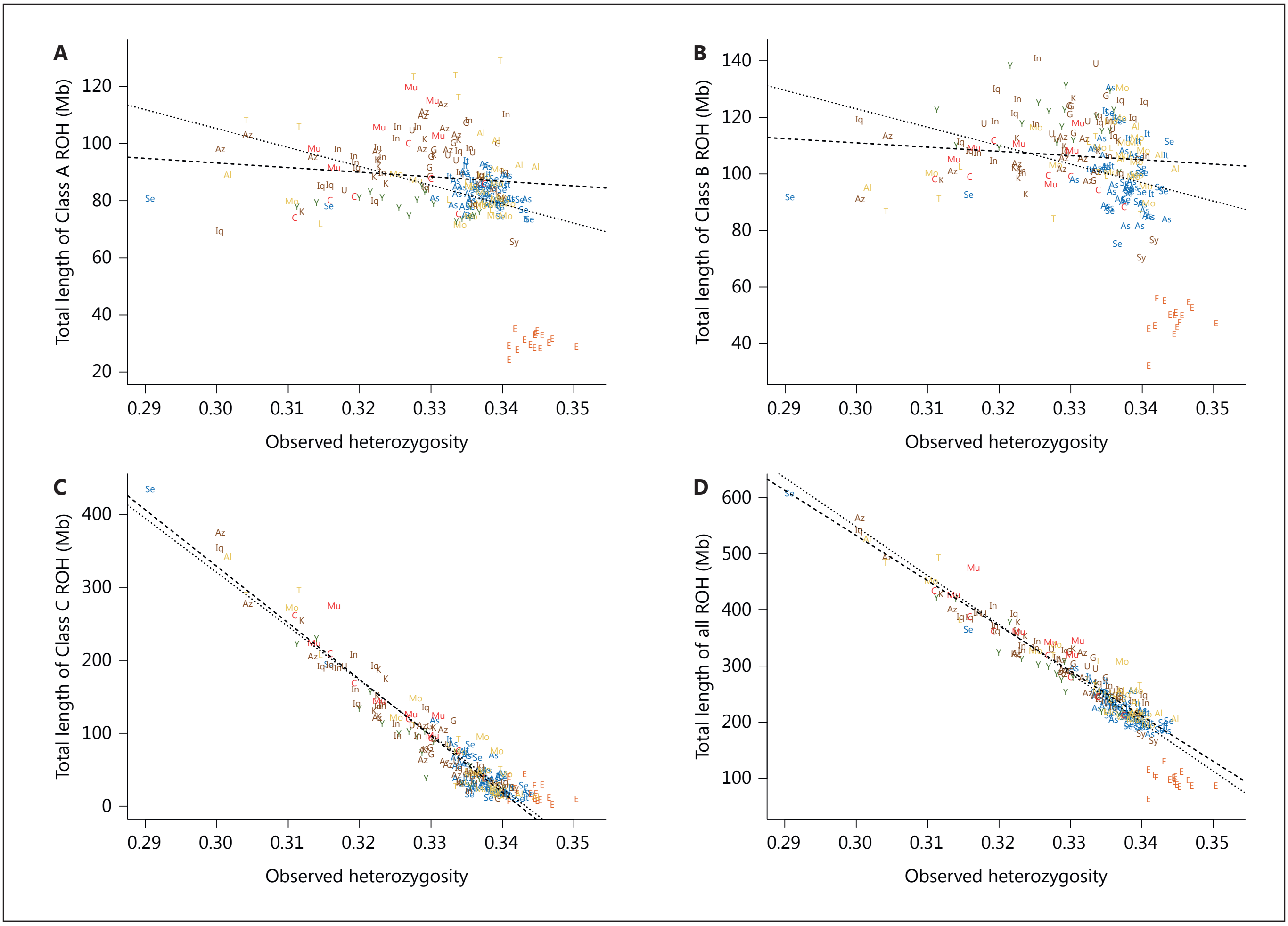


Fig. 5. Individual-level ROH lengths and observed heterozygosities for Jewish populations. **A** Class A ROH. **B** Class B ROH. **C** Class C ROH. **D** All ROH. The dotted regression lines include the Ethiopian Jewish samples, and the dashed regression lines exclude them. The legend follows that in Figure 3.

Our ROH patterns generally agree with past data on genomic sharing in Jewish and non-Jewish populations. The ROH signals at the level of larger geographic regions add to the work of Pemberton et al. [8], with short and intermediate ROH lengths increasing outward from Africa, and with long ROH occurring frequently in the Middle East, Central and South Asia, and the Americas. The ranking of Jewish populations by total ROH length largely accords with that of Waldman et al. [21], obtained in a separate sample of individuals, with ROH detected using a fixed minimum length threshold for ROH identification rather than employing population-specific thresholds and ROH length classes. Among Jewish populations that overlap between our study and that of Waldman et al. [21], Waldman et al. reported that in decreasing order, the Mumbai, Georgian, Cochin, Libyan, Iranian, Tunisian, Iraqi, Yemenite, Algerian, Moroccan, Italian, Syrian, Ashkenazi, and Turkish populations had the longest median total ROH lengths. The corresponding order in our study is Mumbai, Iranian, Cochin, Tunisian, Georgian, Yemenite, Iraqi, Algerian, Libyan, Moroccan, Italian, Ashkenazi, Sephardi (largely Turkish), and Syrian (Fig. 1D). Although the specific rankings differ in several positions, both studies find that ROH values are generally higher in most Middle Eastern, North African, and South Asian Jewish populations than in the Ashkenazi and Sephardi populations.

More generally, studies of IBD levels across individuals within Jewish populations have detected similar patterns to those we have seen for ROH, typically with analogous

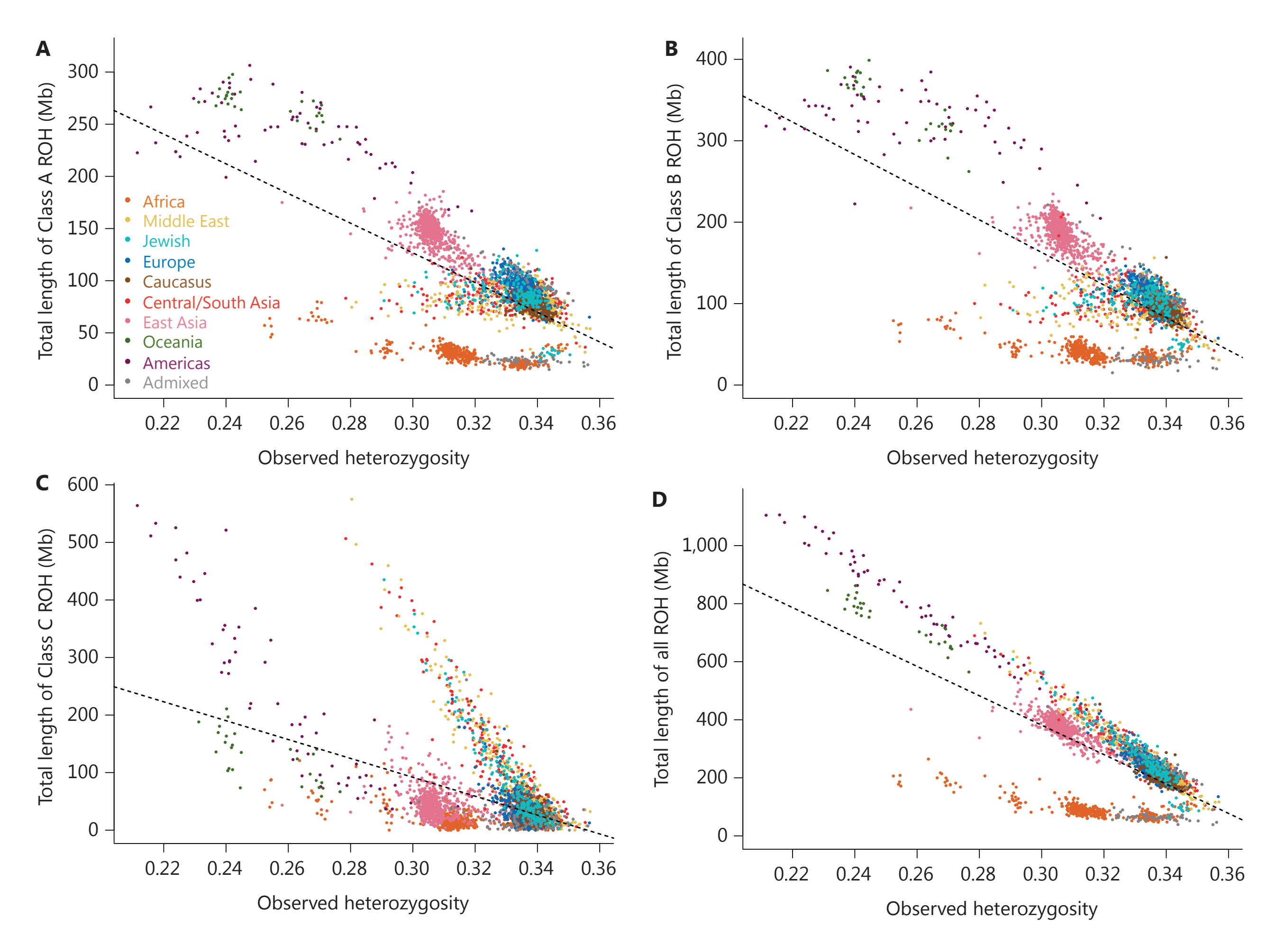


Fig. 6. Individual-level ROH lengths and observed heterozygosities. **A** Class A ROH. **B** Class B ROH. **C** Class C ROH. **D** All ROH. For an individual, observed heterozygosity is the proportion of SNP positions that are heterozygous, excluding positions with missing data. The lines represent regression lines. Colors follow those in Figure 1. Considering all sampled individuals, the negative correlation between total ROH length and observed heterozygosity is strong ($r = -0.696$). Class A and Class B ROH show similar patterns of correlation with observed heterozygosity ($r = -0.597$ and $r = -0.641$, respectively), whereas the relationship is somewhat weaker for Class C ROH ($r = -0.477$).

higher IBD levels in Middle Eastern, North African, and South Asian Jewish populations than in the Ashkenazi and Sephardi groups [17, 18, 20, 21]. By subdividing ROH into classes, we have found that owing to their similar positions in relation to out-of-Africa migrations, the Jewish groups are relatively similar in their short and intermediate ROH, and ROH variability across populations lies primarily in the long ROH. It is possible that increased consanguinity rates that underlie an increase in long ROH can inflate IBD sharing not only for the two haplotypes of the offspring of a consanguineous union, but also for pairs of haplotypes in the population more generally. If consanguinity were to increase IBD sharing in this manner, then variability across Jewish populations in within-population IBD sharing might result in part from the differences among the populations in consanguinity rates. This argument is supported by an observation that in a population pedigree model, an increase in consanguinity decreases the mean time to the most recent common ancestor of a pair of lineages sampled from different individuals (solving eqs. 1–3 for V in Campbell [30]). The reduced time to the most recent common ancestor from increased consan-

Table 3. Demographic estimates of consanguinity rates in the 1950s, and associated inbreeding coefficients

Jewish population	Regional group	Sample size	Relationship			Inbreeding coefficient, F (%)
			First cousins	Uncle–niece	More distant	
Iranian	Middle Eastern	427	68	7	37	1.471
Iraqi	Middle Eastern	1,450	238	16	162	1.513
Syrian	Middle Eastern	406	15	2	16	0.416
Yemenite	Yemenite	628	50	4	61	0.881
Ashkenazi	European	4,734	64	2	50	0.123
Sephardi	European	607	19	2	27	0.376
Libyan	North African	298	18	2	12	0.587
Moroccan	North African	504	26	10	18	0.682
Tunisian	North African	149	16	2	2	0.881

Each data point represents a birth, and the consanguinity of the parents is tabulated, as reported by Goldschmidt et al. [25]. The population inbreeding coefficient is computed by summing 1/16 of the fraction of first-cousin unions, 1/8 of the fraction of uncle–niece unions, and 1/32 of the fraction of more distant unions of relatives.

Table 4. Slope, correlation coefficient, and p value for the regression of population mean ROH lengths on inbreeding coefficients, using different values for the inbreeding coefficient value F_{distant} for relationships more distant than first cousins

ROH class	$F_{\text{distant}} = 1/16$			$F_{\text{distant}} = 1/32$			$F_{\text{distant}} = 1/64$		
	slope	r	p	slope	r	p	slope	r	p
Class A	7.67	0.317	0.406	12.10	0.418	0.263	15.28	0.481	0.190
Class B	16.40	0.657	0.055	18.54	0.621	0.074	19.42	0.593	0.093
Class C	47.81	0.712	0.031	61.42	0.765	0.016	69.95	0.794	0.011
All	71.88	0.711	0.032	92.06	0.762	0.017	104.65	0.789	0.012

guinity generates an increase in IBD sharing as a consequence, because less time has transpired on average since the occurrence of recombinations that break down IBD segments. We note, however, that the model [30] that underlies this reasoning uses sib mating; further analysis of consanguinity models suited to human populations will be required for clarifying the relationship between consanguinity and within-population IBD sharing.

In the context of European and European-American populations, the Ashkenazi Jewish population has been seen to have relatively high ROH and between-individual IBD levels [22–24, 31]. With a larger number of European populations tested, this pattern is somewhat supported in our study, as ROH levels in Ashkenazi Jews exceed those in many, though not all, European populations (Fig. 1, 2). Both according to ROH in our study and that of Waldman et al. [21], and by within-population IBD [17, 18, 20, 21], however, Ashkenazi Jews are not among the Jewish populations with the highest levels of genomic sharing. This result is observed for shorter Class A and Class B ROH as well as for the longer Class C. The status of the Ashkenazi Jewish group as a relatively homogeneous population isolate in relation to Europeans and European Americans contrasts with its shorter ROH and IBD segments in relation to many Middle Eastern, North African, and South Asian Jewish populations. This contrast also extends to consideration of Mendelian diseases whose prevalences are amplified by consanguinity, as Ashkenazi Jewish populations can be regarded as having a high Mendelian disease burden in the context of European and European-American populations, but not necessarily in relation to other Jewish populations [32].

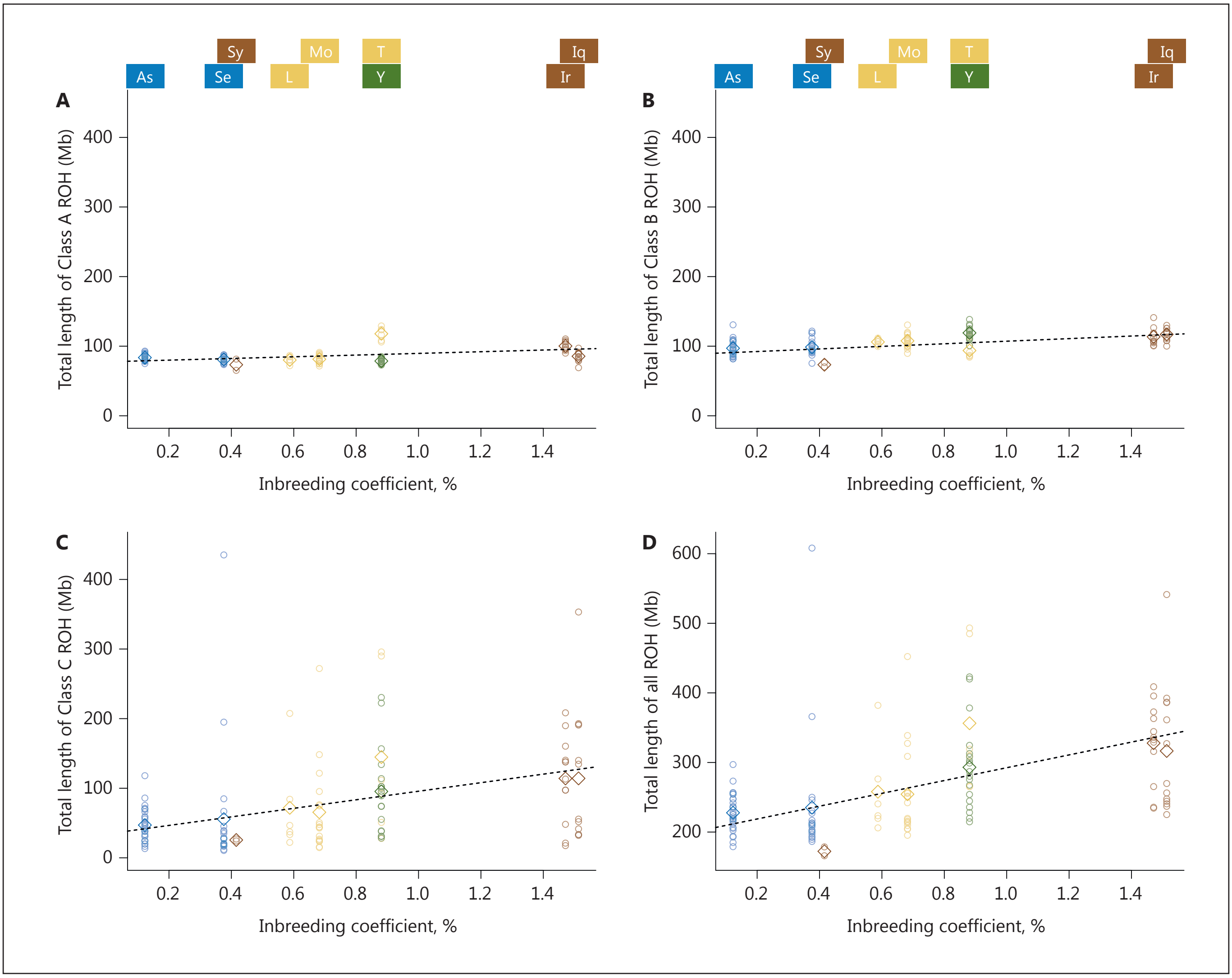


Fig. 7. ROH lengths and inbreeding coefficients from demographic data for Jewish populations. **A** Class A ROH. **B** Class B ROH. **C** Class C ROH. **D** All ROH. The plots consider the 9 Jewish populations for which demographic data were available. Panels **A–C** are plotted on the same scale, and in panel **D**, the *y*-axis represents the same height, so that the slopes of all 4 regression lines are visually comparable. Each point represents an individual, with its *x*-axis value being the estimated inbreeding coefficient of the population to which it belongs. The diamonds indicate the mean total ROH length of a particular class over all individuals in a population, and the dashed lines represent regression lines for the mean ROH lengths. Population abbreviations, which appear above the figures in alignment with their associated inbreeding coefficients, follow those in Figure 3.

Because the total lengths of long ROH have a close conceptual relationship with the demographic consanguinity measure – closer than corresponding relationships involving the short and intermediate ROH lengths – the accord of the level of long ROH observed in individual genomes with demographic measures of consanguinity illustrates the perspective of Pemberton et al. [8] that long ROH reflect recent parental relatedness, whereas short and intermediate ROH reflect more ancient migration events. Among Jewish populations, both the lengths of long ROH and consanguinity levels are greatest in geographic regions where the non-Jewish populations also have high total lengths for long ROH. Many factors underlie the historical consanguinity practices of the various Jewish populations, including their levels of isolation from other Jewish populations, their interpretations of Jewish texts favoring consanguinity, and cultural and economic factors [25, 32]. The geographic overlap of high-consanguinity populations suggests that historically, consanguinity in the Jewish populations might have been influenced in some cases by factors similar to those that have contributed to consanguinity in neighboring non-Jewish populations.

We found that each 1% increase in the consanguinity-based inbreeding coefficient predicted an increase of 92.06 Mb in the total length of ROH. This value corresponds to ~3% of the human genome, a larger increase in ROH length than the 1% expected from a 1% inbreeding coefficient increase. Because the consanguinity-based measure is based only on the most recent generation, it does not capture effects of consanguinity in previous generations. Nontrivial consanguinity rates might have persisted over many generations, and a single-generation computation might substantially underestimate the true inbreeding coefficient.

Both for the Jewish populations and for other populations with high median values for the total length of long ROH, the variability across individuals of the total Class C ROH length was particularly high. It is possible that like the median, a high variance is also an indicator of high consanguinity levels. In this view, within a population, some individuals might descend from multiple generations of consanguineous union, whereas other family lineages – perhaps even most lineages – might not participate in a cultural preference for consanguinity at all. That such preferences vary among families within populations is seen in the aggregated Ashkenazi Jewish population, for which consanguinity rates vary by country of origin [25, 33]; other variables such as religiosity and education level that can be intergenerationally correlated within families have associations with endogamy levels [33]. Further investigation of intergenerational patterns might shed light on the information possessed by ROH variances regarding consanguinity practices.

The demographic consanguinity data we have used were collected for births that occurred during the 1950s. More recent studies have documented substantial decreases in consanguinity for the Jewish populations [33, 34]. Even when the 1950s data were collected, consanguinity rates were decreasing; although the births for which consanguinity was measured took place during 1955–1957, marriages had a range of dates, and in most populations, more recent marriages had lower consanguinity rates [25, 26]. Our samples, collected from adult volunteers prior to the study of Behar et al. [19], reflect a wide range of ages, and for 54 individuals among the 126 from the 9 populations for which consanguinity data were available, we were able to extrapolate from the age at the time of sampling to obtain approximate birth dates. This computation suggests a mean and median birth date of 1963, close to the time at which the consanguinity data were measured, with standard deviation 13 and a range from 1929 to 1989. It is interesting that although consanguinity rates in the populations have undergone considerable change, ROH evaluated in comparatively small samples of volunteers in the genomic era have recovered the signature of population-level consanguinity patterns measured near the time of their births, 6 decades ago. Note, however, that the modern sample might not be entirely independent of the 1950s data: because the 1950s data, with a sample size of >10,000, represent a substantial fraction of all newborns during 1955–1957 [35] in a population of <2,000,000 [36], some of the participants in our study might very well be among those whose mothers were interviewed when they were born.

Genome-wide data sets provide new opportunities for comparing demographic and pedigree-based measures of consanguinity and relatedness with direct measurements of genomic sharing, with increasingly many applications favoring genomic values and even finding that non-genomic values can be unnecessary [37, 38]. We have seen that for understanding population history, including the history of consanguinity, genomic aspects of ROH are highly informative. However, significant additional information was obtained by considering ROH together with demographic data on consanguinity; we expect that studies will have increasing potential to capitalize on combinations of multiple forms of data in studying the recent history of mating practices and their genomic consequences.

Acknowledgements

We thank T. Pemberton for technical assistance, and NIH grant R01 HG005855 and the Stanford Center for Computational, Evolutionary, and Human Genomics for financial support.

References

1 Browning SR, Browining BL: Identity by descent between distant relatives: detection and applications. Annu Rev Genet 2012;46:617–633.
2 Thompson EA: Identity by descent: variation in meiosis, across genomes, and in populations. Genetics 2013;194:301–326.
3 McQuillan R, Leutenegger AL, Abdel-Rahman R, et al: Runs of homozygosity in European populations. Am J Hum Genet 2008;83:359–372.

4 Keller MC, Visscher PM, Goddard ME: Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 2011;189:237–249.

5 Rehder CW, David KL, Hirsch B, Toriello HV, Wilson CM, Kearney HM: American College of Medical Genetics and Genomics: standards and guidelines for documenting suspected consanguinity as an incidental finding of genomic testing. Genet Med 2013; 15:150–152.

6 Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF: Genomic runs of homozygosity record population history and consanguinity. PLoS One 2010;5:e13996.

7 Nothnagel M, Lu TT, Kayser M, Krawczak, M: Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum Mol Genet 2010;19:2927–2935.

8 Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ: Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet 2012;91:275–292.

9 Campbell H, Carothers AD, Rudan I, et al: Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. Hum Mol Genet 2007;16:233–241.

10 Keller MC, Simonson MA, Ripke S, et al: Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLoS Genet 2012; 8:e1002656.

11 McQuillan R, Eklund N, Pirastu N, et al: Evidence of inbreeding depression on human height. PLoS Genet 2012;8:e1002655.

12 Ghani M, Reitz C, Cheng R, et al: Association of long runs of homozygosity with Alzheimer disease among African American individuals. JAMA Neurol 2015;72:1313–1323.

13 Joshi PK, Esko T, Mattsson H, et al: Directional dominance on stature and cognition in diverse human populations. Nature 2015;523: 459–462.

14 Szpiech ZA, Xu J, Pemberton TJ, Peng W, Zöllner S, Rosenberg NA, Li JZ: Long runs of homozygosity are enriched for deleterious variation. Am J Hum Genet 2013;93:90–102.

15 Pippucci T, Magi A, Gialluisi A, Romeo G: Detection of runs of homozygosity from whole exome sequencing data: state of the art and perspectives for clinical, population and epidemiological studies. Hum Hered 2014;77: 63–72.

16 Ostrer H, Skorecki K: The population genetics of the Jewish people. Hum Genet 2013;132: 119–127.

17 Atzmon G, Hao L, Pe'er I, et al: Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern ancestry. Am J Hum Genet 2010;86:850–859.

18 Campbell CL, Palamara PF, Dubrovsky M, et al: North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. Proc Natl Acad Sci USA 2012;109:13865–13870.

19 Behar DM, Metspalu M, Baran Y, et al: No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. Hum Biol 2013; 85:859–900.

20 Waldman YY, Biddanda A, Davidson NR, et al: The genetics of Bene Israel from India reveals both substantial Jewish and Indian ancestry. PLoS One 2016;11:e0152056.

21 Waldman YY, Biddanda A, Dubrovsky M, Campbell CL, Oddoux C, Friedman E, Atzmon G, Halperin E, Ostrer H, Keinan A: The genetic history of Cochin Jews from India. Hum Genet 2016;135:1127–1143.

22 Bray SM, Mulle JG, Dodd AF, Pulver AE, Wooding S, Warren ST: Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. Proc Natl Acad Sci USA 2010;107:16222–16227.

23 Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, Mountain JL: Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. PLoS One 2012;7:e34267.

24 Carmi S, Hui KY, Kochav E, et al: Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. Nat Commun 2014;5:4835.

25 Goldschmidt E, Ronen A, Ronen I: Changing marriage systems in the Jewish communities of Israel. Ann Hum Genet 1960;24:191–204.

26 Ronen A, Ronen I, Goldschmidt E: Marriage systems; in Goldschmidt E (ed): The Genetics of Migrant and Isolate Populations. Baltimore, Williams and Wilkins, 1963, pp 340–343.

27 Verdu P, Pemberton TJ, Laurent R, et al: Patterns of admixture and population structure in native populations of northwest North America. PLoS Genet 2014;10:e1004530.

28 Wang S, Haynes C, Barany F, Ott J: Genome-wide autozygosity mapping in human populations. Genet Epidemiol 2009;33:172–180.

29 Bonné-Tamir B: The Samaritans: a living ancient isolate; in Eriksson AW, Forsius HR, Nevallina HR, Workman PL, Norio RK (eds): Population Structure and Genetic Disorders. London, Academic Press, 1980, pp 27–41.

30 Campbell RB: The effect of inbreeding constraints and offspring distribution on time to the most recent common ancestor. J Theor Biol 2015;382:74–80.

31 Olshen AB, Gold B, Lohmueller KE, et al: Analysis of genetic variation in Ashkenazi Jews by high density SNP genotyping. BMC Genet 2008;9:14.

32 Goodman RM: Genetic Disorders among the Jewish People. Baltimore, Johns Hopkins University Press, 1979.

33 Cohen T, Vardi-Saliternik R, Friedlander Y: Consanguinity, intracommunity and intercommunity marriages in a population sample of Israeli Jews. Ann Hum Biol 2004;31:38–48.

34 Tsafrir J, Halbrecht I: Consanguinity and marriage systems in the Jewish community in Israel. Ann Hum Genet 1972;35:343–347.

35 Israel Central Bureau of Statistics: Total livebirths, http://statil.org/Population/252/3714, 2012.

36 Israel Central Bureau of Statistics: Population, http://statil.org/Population/263/3700, 2012.

37 Speed D, Balding DJ: Relatedness in the post-genomic era: is it still useful? Nat Rev Genet 2015;16:33–44.

38 Cussens J, Sheehan NA: Special issue on new developments in relatedness and relationship estimation. Theor Popul Biol 2016;107:1–3.