

Population-Genetic Influences on Genomic Estimates of the Inbreeding Coefficient: A Global Perspective

Trevor J. Pemberton^a Noah A. Rosenberg^b

^aDepartment of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Man., Canada;

^bDepartment of Biology, Stanford University, Stanford, Calif., USA

Key Words

Consanguinity · Homozygosity · Identity by descent · Inbreeding · Short tandem repeats

Abstract

Background/Aims: Culturally driven marital practices provide a key instance of an interaction between social and genetic processes in shaping patterns of human genetic variation, producing, for example, increased identity by descent through consanguineous marriage. A commonly used measure to quantify identity by descent in an individual is the inbreeding coefficient, a quantity that reflects not only consanguinity, but also other aspects of kinship in the population to which the individual belongs. Here, in populations worldwide, we examine the relationship between genomic estimates of the inbreeding coefficient and population patterns in genetic variation. **Methods:** Using genotypes at 645 microsatellites, we compare inbreeding coefficients from 5,043 individuals representing 237 populations worldwide to demographic consanguinity frequency estimates available for 26 populations as well as to other quantities that can illuminate population-genetic influences on inbreeding coefficients. **Results:** We observe higher inbreeding coefficient

estimates in populations and geographic regions with known high levels of consanguinity or genetic isolation and in populations with an increased effect of genetic drift and decreased genetic diversity with increasing distance from Africa. For the small number of populations with specific consanguinity estimates, we find a correlation between inbreeding coefficients and consanguinity frequency ($r = 0.349$, $p = 0.040$). **Conclusions:** The results emphasize the importance of both consanguinity and population-genetic factors in influencing variation in inbreeding coefficients, and they provide insight into factors useful for assessing the effect of consanguinity on genomic patterns in different populations.

© 2014 S. Karger AG, Basel

Introduction

‘Inbreeding’ refers to the production of offspring by the mating of related individuals, often via consanguinity – intra-familial unions of individuals related as second cousins or closer. A large body of ethnographic evidence supports the commonplace occurrence of consanguineous unions in many traditional

human societies, with 353 of the 763 societies listed in the ethnographic tabulations of Murdock either permitting or favoring marriage between first or second cousins [1, 2].

In addition to its occurrence through consanguinity, inbreeding often takes place in human populations ‘cryptically’ as a consequence of background relatedness – recent but unknown kinship among mating pairs. It has been estimated that even in a large randomly mating population of 1 million individuals, at least one shared ancestor likely exists for any given pair of individuals within the last 11 generations [3]; this value decreases to 6 generations for a population of size 1,000. Compared with ostensibly outbred populations, groups that are more inbred can have a higher prevalence of rare recessive monogenic disorders [4, 5]. Further, within-population comparisons have observed a higher prevalence of these disorders [6–8], and in many cases, common multifactorial disorders [9–14] and even infectious diseases [15–17], among inbred individuals compared with more outbred individuals. Thus, understanding worldwide patterns of inbreeding has important consequences for human genetic disease investigations, requiring knowledge both of sociocultural factors that promote overt inbreeding through consanguinity and of population-genetic processes that underlie cryptic inbreeding through background relatedness.

A commonly used measure to quantify the level of inbreeding that exists in an individual’s lineage is the inbreeding coefficient. This measure seeks to estimate the proportion of a genome that is ‘autozygous’ – homozygous for alleles inherited identically by descent from a common ancestor. Originally calculated theoretically from known pedigree structures and Mendel’s law of segregation [18], with the availability of genomic data the inbreeding coefficient can now be estimated empirically from levels of genomic homozygosity [19–29] or inferred autozygosity [30]. In the absence of known pedigrees, genomic inbreeding coefficient estimates in an individual provide an empirical measure of the probability of identity by descent in the two parental copies at a randomly chosen locus. It has been found using simulations that genomic estimators estimate true empirical inbreeding coefficients more accurately than the pedigree-based estimator [3]. This observation can be explained as a consequence of stochasticity in Mendelian segregation ratios [31, 32] and variation in cryptic relatedness [33] that pedigree-based inbreeding coefficient estimates ignore, but that are naturally taken into account by genomic estimates.

A number of sociogenetic and population-genetic processes can influence genomic estimates of the inbreeding coefficient, even in the absence of consanguinity. For example, cultural traditions that promote endogamy – the restriction of marriage to within a predefined preferred group of individuals – have the potential over many generations to inflate the inbreeding coefficient. Such traditions reduce the overall pool of mates, increasing relatedness within populations, and in turn, increasing the probability that cryptic inbreeding will arise through background relatedness. In other populations, historical bottlenecks, geographic isolation, or small population size can result in both high levels of relatedness among members of a population and reduced levels of genetic diversity; consequently, genomic inbreeding coefficients might be high in such populations due to high rates of cryptic inbreeding.

Here, to investigate population patterns in the inbreeding coefficient, we compute genomic inbreeding coefficient estimates in 5,043 individuals from 237 human populations worldwide, using genotypes at 645 microsatellite loci. We compare these patterns to demographic estimates of the frequency of consanguinity available for 26 of the populations as well as to various population-genetic statistics that can assist in interpreting the potential for cryptic inbreeding. The results contribute to an understanding of the factors that influence variation in the inbreeding coefficient in populations worldwide.

Material and Methods

Genotype Data

We used the MS5255 subset of the Pemberton et al. [34] human-only data set containing 645 autosomal microsatellite loci, restricting our analysis to 5,043 unrelated individuals (online suppl. table S1; see www.karger.com/doi/10.1159/000362878 for all online suppl. material) from 237 human populations worldwide (online suppl. table S2) that had both a sample size of at least 10 individuals as well as available geographic coordinates. Geographic region assignments for populations follow Pemberton et al. [34], as do geographic distances of the 220 non-admixed populations from Addis Ababa, Ethiopia.

Heterozygosity

We evaluated the level of genetic variation in each population using observed and expected heterozygosity (H_{obs} and H_{exp} , respectively) averaged across the 645 microsatellite loci in the data set. For population p , we calculated $H_{exp,p}$ with a sample-size-corrected estimator [35], and $H_{obs,p}$ as the mean across loci of the proportion of observed heterozygous genotypes, considering in the calculation at each locus only those individuals from population p with non-missing genotypes. Similarly, for individual i , we determined $H_{obs,i}$ as the proportion of heterozygous genotypes in the

genotype data of individual i , considering only those loci with non-missing genotypes. $H_{obs,i}$ estimates for each individual are given in online suppl. table S1, and population $H_{obs,p}$ and $H_{exp,p}$ estimates are given in online suppl. table S2.

Population Differentiation

We obtained genetic distances between all possible pairs among the 237 populations by computing F_{st} on the 645 microsatellite loci using Arlequin (v.3.5.1.3) [36]. This approach employs the F_{st} estimator of Weir and Cockerham [37]. Separately for each population, we calculated mean pairwise F_{st} across the intra-geographic-region comparisons that included that population (online suppl. table S2).

Genomic Inbreeding Coefficient Estimates

Individual-level genomic inbreeding coefficient estimates, F_{hbd} , were defined as the proportion of an individual genome inferred to be autozygous – or homozygous by descent (HBD) – by a maximum likelihood method that uses a hidden Markov model to infer HBD status along the genome. The method uses genotypes, allele frequencies, and a genomic map to probabilistically identify transitions along the genome between autozygous and non-autozygous regions and to estimate a parameter describing the probability that a genome is autozygous at a random location [30]. We calculated F_{hbd} using the program *FEst* [38] with population-specific count estimates of allele frequencies, genetic map positions taken from Pemberton et al. [39], and a genotyping error rate of 0.008 based on the HGDP-CEPH component of the data set considered here [40]. Default parameter values were used for the number of Markov chain iterations (8,000) and for the initial values for the estimate of the inbreeding parameter (0.05) and the identity-by-descent rate of change (0.05). Individual F_{hbd} estimates are given in online suppl. table S1, and per-population means and standard deviations are given in online suppl. table S2.

Demographic Estimates of Consanguinity Frequency

Demographic estimates of the frequency of consanguinity – intra-familial unions between couples related as second cousins or closer – in 26 populations were taken from the Global Consanguinity Database (www.consang.net; last updated June 13th, 2012; online suppl. table S2). This database contains consanguinity frequencies reported in the peer-reviewed literature. In instances in which a population spanned multiple countries, consanguinity frequency estimates were included only if they were obtained in the same country in which population-genetic sampling took place. In situations in which more than one estimate was available in the database for a population, the mean was used in our analyses.

Statistical Analyses

Pairwise comparisons were performed using R (v.3.0.0) [41]. Pearson's product moment correlation coefficient r was calculated using *cor.test* and Wilcoxon signed-rank and rank-sum tests using *wilcox.test*. To account for effects of population-genetic quantities on correlations of genomic inbreeding coefficients with geographic distance from East Africa, we performed partial-correlation analyses using *pcor.test* [42], calculating a Pearson's partial correlation coefficient r_{pc} while including each population-genetic quantity as a covariate in separate computations.

Results and Discussion

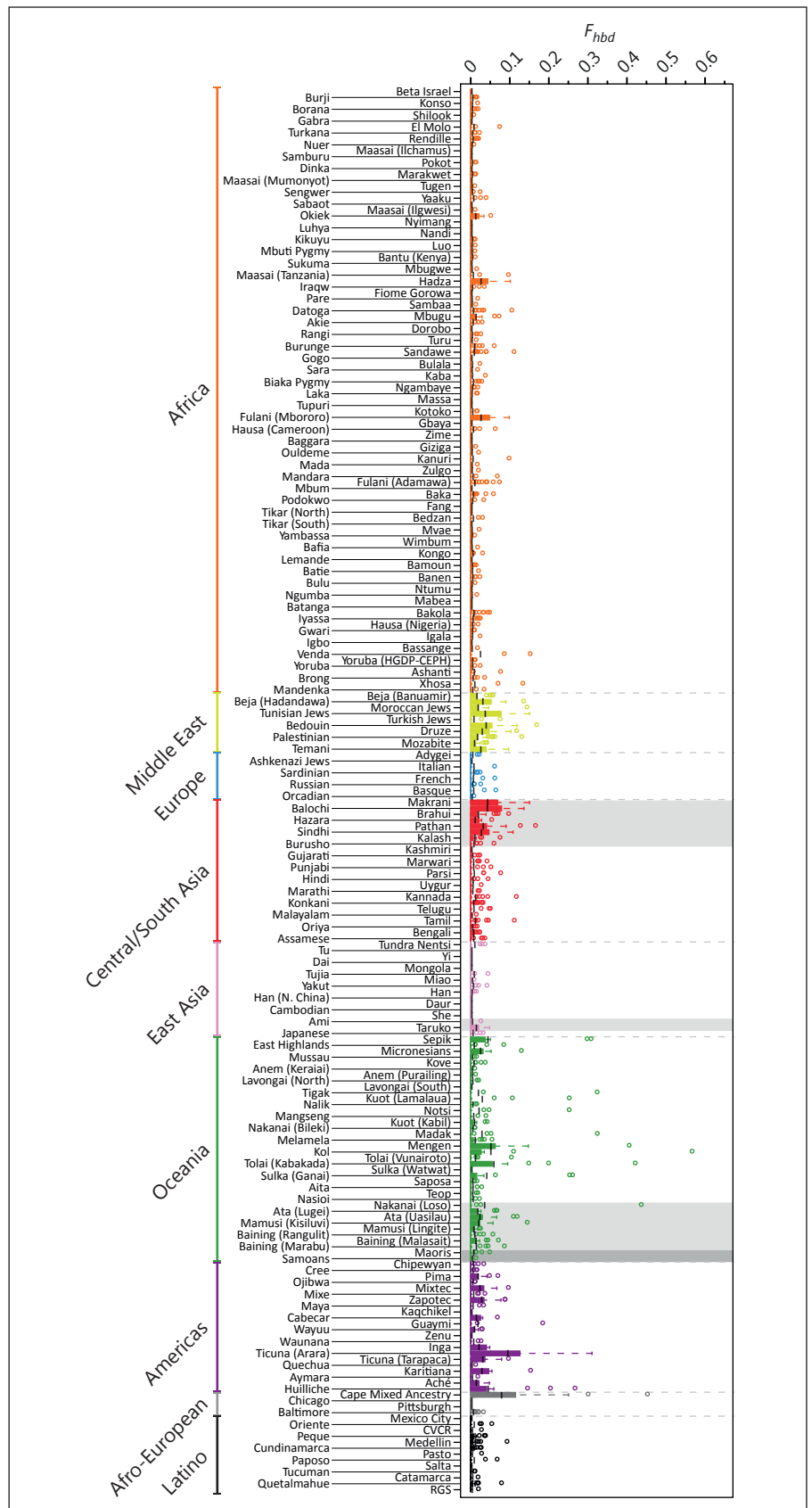
Worldwide Patterns in Genomic Inbreeding Coefficients

Figure 1 plots F_{hbd} for each population, showing that individuals with non-zero F_{hbd} appear in all geographic regions. Values of F_{hbd} generally increase with geographic distance from East Africa ($r = 0.358$, $p = 4.64 \times 10^{-8}$), with higher and more variable values occurring in most populations from the Middle East, Central/South Asia, Oceania, and the Americas than in most populations from Sub-Saharan Africa, Europe, and East Asia. Thus, a larger fraction of individuals from the Middle East, Central/South Asia, Oceania, and the Americas tend to have higher levels of parental relatedness, in accord with demographic estimates of high levels of consanguineous marriage [43, 44] (Middle East, Central/South Asia) or the existence of large numbers of small or isolated populations (Oceania, Americas). The result is compatible with the expectation that homozygosity in these regions is more likely to reflect autozygosity than in Sub-Saharan Africa, Europe, and East Asia, where consanguinity is generally less frequent than in the Middle East and Central/South Asia and populations have not experienced the same bottleneck or isolation events as in Oceania and the Americas.

Most Sub-Saharan African populations contain some individuals with non-zero F_{hbd} , with noticeably higher values occurring in such groups as the Hadza, Fulani (Mbororo), and Venda populations. The Hadza are a group of hunter-gatherers whose small population size and isolation [45–47] would be expected to create high levels of background relatedness, while their acceptance of consanguineous unions – only first-degree relatives are reportedly prohibited from marrying [48] – would be expected to produce non-trivial levels of overt inbreeding. The Fulani (Mbororo) – also known as Wodaabe – are a small nomadic subgroup of the Fulani ethnic group who actively practice consanguinity through arranged marriages between first cousins of the same lineage [49]. By contrast, the Venda are a large clan-based society within which arranged marriages occur frequently between members of the same clan and, as a consequence, are largely consanguineous [50, 51]. The higher F_{hbd} values observed in these populations therefore reflect their traditional practices of kin marriage.

Considerable variation exists in values of F_{hbd} among Middle Eastern populations, with the highest values occurring in the Bedouin – consistent with the exceptionally high frequency of consanguineous unions reported in

Fig. 1. Population-specific distributions of genomic inbreeding coefficients. Box-and-whisker plots represent distributions across individuals in each population for F_{hbd} . Populations are ordered from top to bottom by geographic region and within regions by increasing geographic distance from Addis Ababa, Ethiopia. Oceanian populations are separated into coastal, inland (highlighted by a light gray shade), and Polynesian (highlighted by a dark gray shade) groups, again ordering populations within groups by distance from Addis Ababa. Among the Central/South Asian populations, the 8 Pakistani populations are indicated by a light gray shade, as are the 2 Taiwanese populations (Ami and Taruko) among East Asians.



this traditionally nomadic tribal population [52, 53]. Some of the lowest values are found in the Mozabite subgroup of the Berber ethnic group, in contrast to their reported preference for consanguineous marriage [54]. F_{hbd} values in European populations are almost exclusively below those observed in the Middle East. Curiously, the lowest values occur in the Ashkenazi Jewish group, well-known to have high levels of identity by descent [55–59], and the highest occur in the French population, despite the low frequency of consanguinity reported among its Roman Catholic component [60].

Among Central/South Asians, several Pakistani non-tribal populations (Makrani, Balochi, Brahui, Pathan, and Sindhi) have noticeably higher and more variable values of F_{hbd} than those of the other populations – which are predominantly from India – with the highest values occurring in the Makrani and Balochi populations. These results are compatible with the higher rates of consanguinity reported among Pakistani populations than among Indian populations [43, 44] and with documented preferences of the Makrani [61] and Balochi [62, 63] populations for consanguineous marriage. Interestingly, whereas high levels of consanguinity have been reported among the Hazara in Afghanistan [64, 65], F_{hbd} values in the Pakistani Hazara population in our data set are among the lowest observed in Central/South Asians.

Worldwide, East Asians have the lowest F_{hbd} values, with only 8 of 16 populations containing individuals with non-zero F_{hbd} values: Tundra Nentsi, Tujia, Miao, Yakut, Han, Ami, Taruko, and Japanese. Whereas the Han population, sampled in Southern China [66], contains individuals with non-zero F_{hbd} values, the Han (Northern China) population does not, compatible with the higher reported frequency of consanguinity among Southern compared with Northern Han [67–69].

Populations from Oceania and the Americas contain individuals with the highest inbreeding coefficients observed worldwide. In Oceania, many of the individuals with exceptionally high F_{hbd} occur in coastal rather than inland Melanesian populations – in contrast with the increased isolation documented for inland compared with coastal populations [70–72]. Polynesian groups – Samoans and Maoris – have noticeably lower F_{hbd} than most other Oceanian populations. These observations are consistent with Samoan and Maori prohibition of incestuous unions – considered as first cousins or closer among Samoans [73] and second cousins or closer among Maori [74] – and the small size of many Melanesian populations [72, 75].

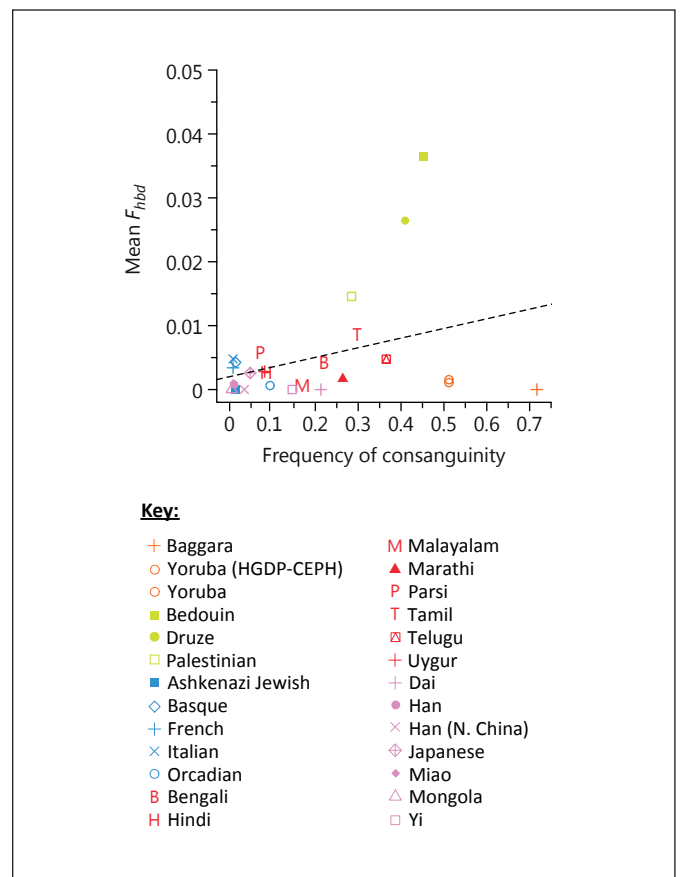


Fig. 2. Relationship between genomic inbreeding coefficients and the frequency of consanguinity. The scatter plot shows the relationship between demographic estimates of the frequency of consanguinity in 26 populations worldwide and mean F_{hbd} across individuals ($r = 0.349$, $p = 0.040$). The black dashed line depicts the regression line.

In the Americas, which contain many small and isolated populations, several groups, including the Amazonian Ticuna (Arara) [75, 76] and Andean Huilliche [75], have high F_{hbd} values. However, many of the smallest groups do not have high F_{hbd} , including the Aché, a small population of hunter-gatherers who prohibit marriage at the level of first cousins or closer [75, 77], the Waunana, a forest group that has increasingly intermarried with non-Waunana over the past 70 years [78], and the Pima, a relatively small and isolated population in which consanguinity is traditionally avoided [79].

Among Afro-European admixed populations, the Cape Mixed Ancestry population – in contrast to a low reported frequency of consanguinity [80] – has higher and more variable estimates than the three investigated

Table 1. Correlation between per-population means across individuals of F_{hbd} and population-genetic variables in each geographic region

Geographic region	Number of populations	$H_{obs,p}$		Mean pairwise F_{st}	
		r	p	r	p
World	237	-0.563	<10 ⁻¹⁶	0.331	5.16 × 10 ⁻⁷
Africa	102	-0.731	<10 ⁻¹⁶	0.465	8.46 × 10 ⁻⁷
Middle East	10	-0.820	0.004	0.017	0.960
Europe	8	-0.570	0.140	-0.506	0.201
Central/South Asia	24	-0.846	1.89 × 10⁻⁷	0.076	0.724
East Asia	16	-0.640	0.008	0.758	6.68 × 10⁻⁴
Oceania	38	-0.510	0.001	-0.137	0.411
Americas	22	-0.594	0.004	0.279	0.208
Afro-European	4	-0.999	0.001	0.968	0.032
Latino	13	-0.253	0.405	0.408	0.166

p values <0.05 are shown in bold.

African-American populations, a result that we might hypothesize reflects greater historical isolation or fewer founders [81] compared to African Americans. Latino populations, descended primarily from European and Native American groups, share similar F_{hbd} values that in most cases more closely resemble those found in Europeans than those of Native Americans.

Influence of Consanguinity Frequency on Genomic Inbreeding Coefficients

We have observed higher and more variable inbreeding coefficients in geographic regions with known high levels of consanguineous marriage [43, 44] – particularly the Middle East and Central/South Asia. These results are consistent with the expectation that genomic estimators of the inbreeding coefficient reflect consanguinity levels in a population. To further investigate the relationship between inbreeding coefficients and consanguinity, we examined the correlation of F_{hbd} with demographic estimates of the frequency of consanguinity, which was available for 26 of the 237 populations (online suppl. table S2). In accord with the prior expectation, we observe a significant positive correlation of this variable with the mean F_{hbd} values across individuals in a population (fig. 2; $r = 0.349$, $p = 0.040$).

Population-Genetic Influences on Genomic Inbreeding Coefficients

The high F_{hbd} values in many of the populations might also partly reflect the influence of historical bottlenecks or genetic isolation on levels of genetic variability – processes that act separately from overt preferences for consanguineous marriage in inflating inbreeding coefficients. To investigate such influences, we examined rela-

tionships of genomic inbreeding coefficients with heterozygosity and F_{st} .

Heterozygosity

F_{hbd} measures excess homozygosity in individual genomes [30]; consequently, we might expect to observe negative correlations between this quantity and heterozygosity. Consistent with this expectation, F_{hbd} is negatively correlated with observed heterozygosity at both the population level ($H_{obs,p}$; fig. 3a, table 1) and the individual level ($H_{obs,i}$; online suppl. fig. S1, online suppl. table S3). If we separately compare F_{hbd} with H_{obs} for each geographic region, we still observe negative correlations at both the population (table 1) and individual (online suppl. table S3) levels, confirming the pattern seen at a worldwide scale. Some of the strongest negative correlations occur in regions with known high levels of consanguinity [43, 44], particularly the Middle East and Central/South Asia (table 1, online table S3).

Consistent with previous investigations of $H_{exp,p}$ [34, 82–84], which is highly correlated with $H_{obs,p}$ (online suppl. fig. S2; $r = 0.975$, $p < 10^{-16}$), we observe $H_{obs,i}$ to decrease with distance from Africa (fig. 4a; $r = -0.905$, $p < 10^{-16}$). Taken together with the negative correlation observed between genomic inbreeding coefficients and H_{obs} , the observed decrease in H_{obs} with distance from Africa might at least in part explain the observed increase in F_{hbd} with distance from Africa (fig. 1). In this view, decreasing heterozygosity inflates identity by descent and, in turn, the inbreeding coefficient. If we compute the partial correlation of F_{hbd} with geographic distance from Africa, accounting for the effect of H_{obs} , we observe a significant negative correlation ($r_{pc} = -0.431$, $p = 1.98 \times 10^{-12}$). This

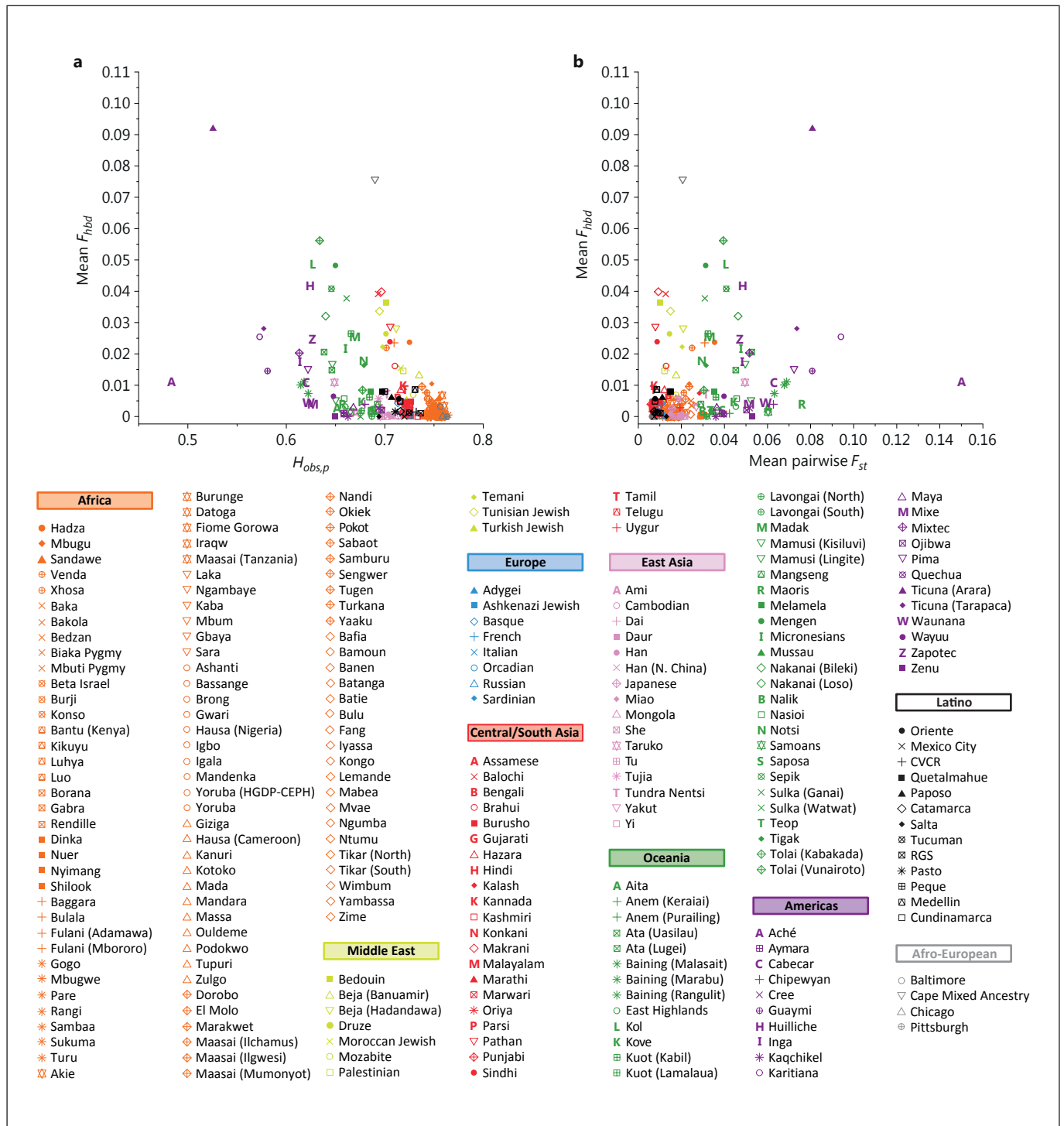
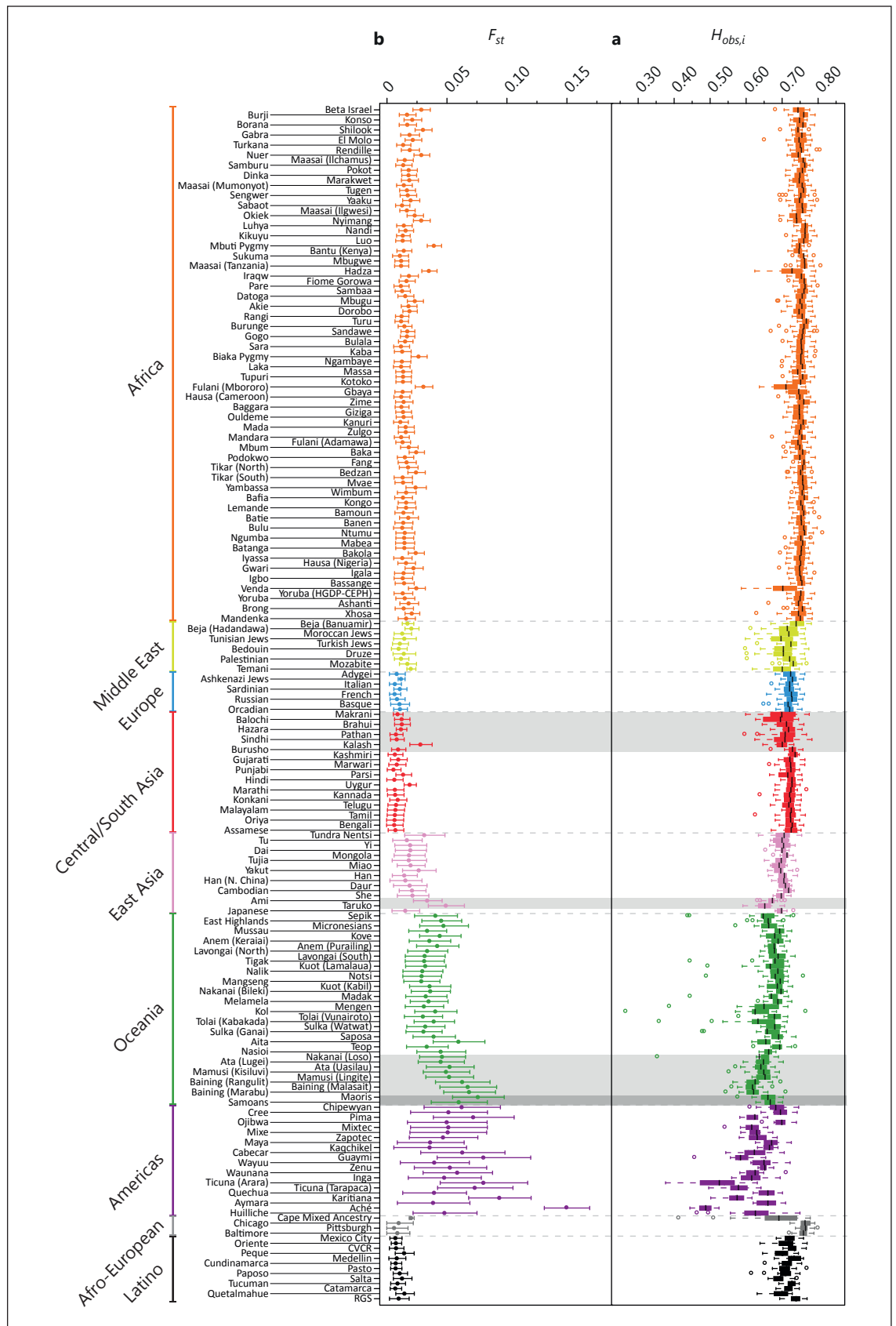


Fig. 3. Relationship between genomic inbreeding coefficients and population-genetic quantities. Scatter plots show the relationships between per-population mean F_{hbd} across individuals and $H_{obs,p}$ (**a**) and mean pairwise F_{st} across intra-geographic-region comparisons that include a population (**b**). African populations are as-

signed the same symbol if they had similar cluster memberships in the $K = 14$ *Structure* analysis by Tishkoff et al. [110]. Pacific Islander populations from the same tribe [72] are assigned the same symbol. Pearson correlation coefficients for each comparison appear in table 1.



4

(For legend see next page.)

negative rather than positive correlation supports the view that much of the signal of higher F_{hbd} with increasing distance from Africa is attributable to the general patterns of decreasing genetic variability with distance from Africa.

Mean Pairwise F_{st}

Pairwise F_{st} measures population differentiation, producing higher values when two populations have large between-population differences but small within-population differences. We expect larger values in comparisons involving genetically isolated populations, in whom we might also expect higher levels of cryptic inbreeding due to reduced mate choice. We can therefore predict mean pairwise F_{st} – calculated across intra-geographic-region pairwise comparisons that include a population – to be positively correlated with mean F_{hbd} across individuals. Consistent with this expectation, we observe a positive correlation between mean pairwise F_{st} and mean F_{hbd} (fig. 3b, table 1). When we consider the relationship of mean pairwise F_{st} with mean F_{hbd} separately in each geographic region, we observe significant positive correlations in Africa and East Asia (table 1); all other non-admixed regions had $p > 0.2$. Similar results are obtained when mean pairwise F_{st} is instead calculated across all 236 pairwise comparisons worldwide that include a population (not shown). This result suggests that the positive correlation at the worldwide level is driven primarily by differences across regions in the regional levels of F_{st} and mean F_{hbd} , rather than by differences among populations within regions.

In accord with the observed positive correlations with mean F_{hbd} , mean pairwise F_{st} generally increases with a population's distance from East Africa (fig. 4b; $r = 0.768$, $p < 10^{-16}$). Taken together with the positive correlation between genomic inbreeding coefficients and mean pairwise F_{st} , the increase in mean pairwise F_{st} with distance from Africa might partly explain the parallel increases in F_{hbd} (fig. 1). The hypothesis is that genetic isolation, as measured by F_{st} , inflates identity by descent and the inbreeding coefficient. If we compute the partial correlation

Fig. 4. Population-specific distributions of population-genetic quantities. **a** Box-and-whisker plot representing the distribution of $H_{obs,i}$ across individuals in each population. **b** Per-population means (dots) and standard deviations (horizontal bars) of pairwise F_{st} across the intra-geographic-region comparisons that contain the population. The figure follows the same format as figure 1.

of F_{hbd} with geographic distance from Africa, accounting for the effect of mean pairwise F_{st} , we observe a weaker positive correlation ($r_{pc} = 0.172$, $p = 0.010$) than in the unadjusted comparison. These results support the view that increasing F_{st} with increasing distance from Africa contributes to the signal of higher F_{hbd} with distance from Africa, though to a lesser extent than is observed in the corresponding computation with heterozygosity.

Conclusions

Genomic estimates of the inbreeding coefficient are used frequently in measuring parental relatedness, owing to their relative ease of estimation from genetic data. Many estimators rely on the relationship between autozygosity and parental relatedness [28–30]. However, when predicting the proportion of the genome that is autozygous, these estimators can fluctuate with levels of homozygosity that vary as a consequence of population processes unconnected to consanguinity [85, 86]. We have examined the relationship with different population-genetic quantities of a genomic inbreeding coefficient estimator based on inferred autozygosity [30] in genome-wide data from a large number of populations [34].

Inbreeding coefficients increased with the geographic distance of a population from a proxy for the origin of the out-of-Africa migration of anatomically modern humans and decreased with increasing heterozygosity. These results are compatible with a migration model outward from East Africa that predicts increases in genetic drift and identity by descent and decreases in heterozygosity with distance from Africa [83, 84, 87]. The positive correlation between inbreeding coefficients and distance from Africa was eliminated if heterozygosity was controlled, indicating that the increase in inbreeding coefficient is a manifestation of the same reduction of ancestral population size outward from Africa that contributes to decreasing heterozygosity. We detected a positive correlation between inbreeding coefficients and mean pairwise F_{st} , a proxy for genetic isolation; in a partial correlation computation, this variable was also seen to explain part of the signal of increasing inbreeding coefficients with distance from Africa. These results suggest that the increase in inbreeding coefficients with distance from Africa reflects the reduction in genetic variability and increase in genetic isolation outward from Africa during the history of human migrations, rather than increases in cultural practices of consanguinity.

In agreement with a recent study that produced related results on genomic inbreeding coefficients in a smaller set of populations [88], superimposed on the signal of the out-of-Africa migration, we observed high inbreeding coefficients in populations from geographic regions with known high levels of consanguinity [43, 44] – particularly the Middle East and Central/South Asia. This result reflects the importance of consanguinity in addition to the cryptic inbreeding generated by the out-of-Africa migration in giving rise to high inbreeding coefficients, and we found a significant positive correlation between demographic estimates of the frequency of consanguinity and inbreeding coefficients in populations for which consanguinity frequency estimates were available. It is noteworthy that variability in the criteria used in determining the frequency of consanguinity for the various populations [52, 67–69, 80, 89–109] likely introduced stochastic variation among these estimates. Given the suggestive nature of the result and the small number of populations examined, it will be worthwhile to revisit this analysis as standardized data on the frequency of consanguinity become available for additional populations.

We note that although the population samples examined have not been sampled by inbreeding status, they generally are not entirely random, and some may lie at

one or another extreme in terms of levels of consanguinity. For example, if a population has many communities and only one was sampled, heterozygosity might be underestimated and inbreeding overestimated. Consequently, the observed inbreeding coefficient patterns might not accurately reflect true patterns in such populations, providing a possible explanation for instances in which genomic inbreeding coefficient estimates disagreed with consanguinity estimates. Nevertheless, because F_{hbd} had sensible patterns of geographic structure and correlations with consanguinity frequencies and population-genetic statistics, and because we have no reason to suspect systematic differences by geography in the oversampling or undersampling of inbreeding, it is likely that any non-random sampling of inbreeding levels within populations is insufficient to obscure underlying geographic and population-genetic influences that we have detected on inbreeding coefficients.

Acknowledgements

We thank the three reviewers for their helpful comments. This investigation was supported by National Institutes of Health grant HG005855 and by the Burroughs Wellcome Fund (N.A.R.).

References

- Murdock GP: Ethnographic Atlas. Pittsburgh, University of Pittsburgh Press, 1967.
- Bittles AH: Consanguinity in Context. Cambridge, Cambridge University Press, 2012.
- 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.
- Modell B, Darr A: Science and society: genetic counselling and customary consanguineous marriage. *Nat Rev Genet* 2002;3:225–229.
- Habeb AM, Flanagan SE, Deeb A, Al-Alwan I, Alawneh H, Balafrej AAL, Mutair A, Hattersley AT, Hussain K, Ellard S: Permanent neonatal diabetes: different aetiology in Arabs compared to Europeans. *Arch Dis Child* 2012;97:721–723.
- Maatouk F, Laamiri D, Argoubi K, Ghedira H: Dental manifestations of inbreeding. *J Clin Pediatr Dent* 1995;19:305–306.
- Stoll C, Alembik Y, Roth MP, Dott B: Parental consanguinity as a cause for increased incidence of birth defects in a study of 238,942 consecutive births. *Ann Genet* 1999;42:133–139.
- Zlotogora J: Genetic disorders among Palestinian Arabs: 1. Effects of consanguinity. *Am J Med Genet* 1997;68:472–475.
- Morton NE: Effect of inbreeding on IQ and mental retardation. *Proc Natl Acad Sci USA* 1978;75:3906–3908.
- Shami SA, Qaisar R, Bittles AH: Consanguinity and adult morbidity in Pakistan. *Lancet* 1991;338:954.
- Rudan I: Inbreeding and cancer incidence in human isolates. *Hum Biol* 1999;71:173–187.
- Rudan I, Rudan D, Campbell H, Carothers A, Wright A, Smolej-Narancic N, Janicijevic B, Jin L, Chakraborty R, Deka R, Rudan P: Inbreeding and risk of late onset complex disease. *J Med Genet* 2003;40:925–932.
- Kanaan ZM, Mahfouz R, Tamim H: The prevalence of consanguineous marriages in an underserved area in Lebanon and its association with congenital anomalies. *Genet Test* 2008;12:367–372.
- Shieh JTC, Bittles AH, Hudgins L: Consanguinity and the risk of congenital heart disease. *Am J Med Genet A* 2012;158A:1236–1241.
- el-Orfi AHA, Singh M, Giasuddin ASM: Conjugal leprosy among Libyan patients. *Dermatology* 1998;196:271–272.
- Lyons EJ, Amos W, Berkley JA, Mwangi I, Shafi M, Williams TN, Newton CR, Peshu N, Marsh K, Scott JAG, Hill AVS: Homozygosity and risk of childhood death due to invasive bacterial disease. *BMC Med Genet* 2009;10:55.
- Lyons EJ, Frodsham AJ, Zhang L, Hill AVS, Amos W: Consanguinity and susceptibility to infectious diseases in humans. *Biol Lett* 2009;5:574–576.
- Wright S: Coefficients of inbreeding and relationship. *Am Nat* 1922;56:330–338.
- Wright S: Isolation by distance. *Genetics* 1943;28:114–138.
- Li CC, Horvitz DG: Some methods of estimating the inbreeding coefficient. *Am J Hum Genet* 1953;5:107–117.
- Haldane JBS: An exact test for randomness of mating. *J Genet* 1954;52:631–635.
- Nei M, Chesser RK: Estimation of fixation indices and gene diversities. *Ann Hum Genet* 1983;47:253–259.

- 23 Robertson A, Hill WG: Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics* 1984;107:703–718.
- 24 Ayres KL, Balding DJ: Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity* 1998;80:769–777.
- 25 Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, Guinness FE, Clutton-Brock TH: Microsatellites reveal heterosis in red deer. *Proc Biol Sci* 1998;265:489–495.
- 26 Coltman DW, Pilkington JG, Smith JA, Pemberton JM: Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* 1999;53:1259–1267.
- 27 Amos W, Worthington Wilmer J, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T: The influence of parental relatedness on reproductive success. *Proc Biol Sci* 2001;268:2021–2027.
- 28 Carothers AD, Rudan I, Kolcic I, Polasek O, Hayward C, Wright AF, Campbell H, Teague P, Hastie ND, Weber JL: Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Ann Hum Genet* 2006;70:666–676.
- 29 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
- 30 Leutenegger AL, Prum B, Genin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson EA: Estimation of the inbreeding coefficient through use of genomic data. *Am J Hum Genet* 2003;73:516–523.
- 31 Franklin IR: The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theor Popul Biol* 1977;11:60–80.
- 32 Hill WG, Weir BS: Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet Res (Camb)* 2011;93:47–64.
- 33 Hill WG, Weir BS: Variation in actual relationship among descendants of inbred individuals. *Genet Res (Camb)* 2012;94:267–274.
- 34 Pemberton TJ, Degiorgio M, Rosenberg NA: Population structure in a comprehensive genomic data set on human microsatellite variation. *G3 (Bethesda)* 2013;3:891–907.
- 35 Nei M: *Molecular Evolutionary Genetics*. New York, Columbia University Press, 1987.
- 36 Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005;1:47–50.
- 37 Weir B, Cockerham C: Estimating F-statistics for the analysis of population structure. *Evolution* 1984;38:1358–1370.
- 38 Leutenegger AL, Labalme A, Genin E, Toutain A, Steichen E, Clerget-Darpoux F, Edery P: Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to Taybi-Linder syndrome. *Am J Hum Genet* 2006;79:62–66.
- 39 Pemberton TJ, Sandefur CI, Jakobsson M, Rosenberg NA: Sequence determinants of human microsatellite variability. *BMC Genomics* 2009;10:612.
- 40 Rosenberg NA: Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet* 2006;70:841–847.
- 41 R Development Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, R Foundation for Statistical Computing, 2013.
- 42 Kim SH, Yi SV: Understanding relationship between sequence and functional evolution in yeast proteins. *Genetica* 2007;131:151–156.
- 43 Bittles AH, Black ML: Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci USA* 2010;107(suppl 1):1779–1786.
- 44 Bittles AH, Black ML: Consanguineous marriage and human evolution. *Annu Rev Anthropol* 2010;39:193–207.
- 45 Marlowe FW: Mate preferences among Hadza hunter-gatherers. *Hum Nature* 2004;15:365–376.
- 46 Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernando-pulle N, Lema G, Nyambo TB, Ramakrishnan U, Reed FA, Mountain JL: History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol* 2007;24:2180–2195.
- 47 Marlowe FW: *The Hadza: Hunter-Gatherers of Tanzania*. Berkeley, University of California Press, 2010.
- 48 Marlowe FW: Why the Hadza are still hunter-gatherers; in Kent S (ed): *Ethnicity, Hunter-Gatherers, and the 'Other': Association or Assimilation in Africa*. Washington, Smithsonian Institution Press, 2002, pp 247–275.
- 49 Beckwith C: Niger's Wodaabe: 'people of the taboo'. *Natl Geogr Mag* 1983;164:483–509.
- 50 Lund P, Maluleke TG, Gaigher I, Gaigher R: Variations in the frequency of albinism among clans in the Venda regions of South Africa. *J Med Genet* 2003;40:S39.
- 51 Raphaelalani TD, Musehane NM: Arranged marriage practices of the Vhavenda community of the Vhembe district, Limpopo province, South Africa. *J Lang Cult* 2013;4:18–22.
- 52 Vardi-Saliternik R, Friedlander Y, Cohen T: Consanguinity in a population sample of Israeli Muslim Arabs, Christian Arabs and Druze. *Ann Hum Biol* 2002;29:422–431.
- 53 Joseph SE: 'Kissing cousins': consanguineous marriage and early mortality in a reproductive isolate. *Curr Anthropol* 2007;48:756–764.
- 54 Brett M, Fentress E: *The Berbers*. Malden, Blackwell Publishing, 1996.
- 55 Atzmon G, Hao L, Pe'er I, Velez C, Pearlman A, Palamara PF, Morrow B, Friedman E, Oddoux C, Burns E, Ostrer H: 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.
- 56 Campbell CL, Palamara PF, Dubrovsky M, Botigue LR, Fellous M, Atzmon G, Oddoux C, Pearlman A, Hao L, Henn BM, Burns E, Bustamante CD, Comas D, Friedman E, Pe'er I, Ostrer H: North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. *Proc Natl Acad Sci USA* 2012;109:13865–13870.
- 57 Gusev A, Palamara PF, Aponte G, Zhuang Z, Darvasi A, Gregersen P, Pe'er I: The architecture of long-range haplotypes shared within and across populations. *Mol Biol Evol* 2012;29:473–486.
- 58 Palamara PF, Lencz T, Darvasi A, Pe'er I: Length distributions of identity by descent reveal fine-scale demographic history. *Am J Hum Genet* 2012;91:809–822.
- 59 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.
- 60 Sutter J, Tabah L: Fréquence et nature des anomalies dans les familles consanguines. *Population* 1954;9:425–450.
- 61 Pastner S, Pastner CM: Agriculture, kinship and politics in Southern Baluchistan. *Man* 1972;7:128–136.
- 62 Afzal M, Ali SM, Siyal HB: Consanguineous marriages in Pakistan. *Pak Dev Rev* 1994;33:663–674.
- 63 Dawson KP, Frossard PM: A hypothesis regarding the origin and spread of the cystic fibrosis mutation deltaF508. *QJM* 2000;93:313–315.
- 64 Saify K, Saadat M: Consanguineous marriages in Afghanistan. *J Biosoc Sci* 2012;44:73–81.
- 65 Saadat M, Tajbakhsh K: Prevalence of consanguineous marriages in west and south of Afghanistan. *J Biosoc Sci* 2012:1–7.
- 66 Cavalli-Sforza LL: *The Human Genome Diversity Project: past, present and future*. *Nat Rev Genet* 2005;6:333–340.
- 67 Du RF: Percentages and types of consanguineous marriage in different nationalities of China (in Chinese). *Zhonghua Yi Xue Za Zhi* 1981;61:723–728.
- 68 Wu L: Investigation of consanguineous marriages among 30 Chinese ethnic groups. *Hered Dis* 1987;4:163–166.
- 69 Zhang JX: Effects of consanguineous marriages on hereditary diseases: a study of the Han ethnic group in different geographic districts of Zhejiang Province (in Chinese). *Zhonghua Yi Xue Za Zhi* 1992;72:674–676, 703.
- 70 Long JC, Naidu JM, Mohrenweiser HW, Gershowitz H, Johnson PL, Wood JW, Smouse PE: Genetic characterization of Gainj- and Kalam-speaking peoples of Papua New Guinea. *Am J Phys Anthropol* 1986;70:75–96.
- 71 Friedlaender JS: Introduction; in Friedlaender JS (ed): *Genes, Language, and Culture History in the Southwest Pacific*. New York, Oxford University Press, 2007, pp 1–9.

- 72 Friedlaender JS, Friedlaender FR, Reed FA, Kidd KK, Kidd JR, Chambers GK, Lea RA, Loo JH, Koki G, Hodgson JA, Merriwether DA, Weber JL: The genetic structure of Pacific Islanders. *PLoS Genet* 2008;4:e19.
- 73 Shore B: Incest prohibitions and the logic of power in Samoa. *J Polyn Soc* 1976;85:275–296.
- 74 Papakura M: *The Old-Time Maori*; Penniman TK (ed). London, Victor Gollancz Ltd, 1938.
- 75 Paul LM, Simons GF, Fennig CD (eds): *Ethnologue: Languages of the World*, ed 17. Dallas, SIL International, 2013. <http://www.ethnologue.com>.
- 76 Nimuendajú C: *The Tukuna*. University of California Publications in American Archaeology and Ethnology, vol 45. Berkeley, University of California Press, 1952.
- 77 Hill KR, Hurtado AM: *Aché Life History: The Ecology and Demography of a Foraging People*. Hawthorne, Aldine de Gruyter, 1996.
- 78 Olson JS: *The Indians of Central and South America: An Ethnohistorical Dictionary*. Westport, Greenwood Publishing Group, 1991.
- 79 Ezell PH: History of the Pima; in Sturtevant WC (ed): *Handbook of North American Indians*. Washington, Smithsonian Institution, 1983, vol 10, pp 149–160.
- 80 Stevenson AC, Johnston HA, Stewart MI, Golding DR: Congenital malformations. A report of a study of series of consecutive births in 24 centres. *Bull World Health Organ* 1966; 34(suppl):9–127.
- 81 de Wit E, Delpont W, Rugamika CE, Meintjes A, Moller M, van Helden PD, Seoighe C, Hoal EG: Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Hum Genet* 2010;128:145–153.
- 82 Prugnolle F, Manica A, Balloux F: Geography predicts neutral genetic diversity of human populations. *Curr Biol* 2005;15:R159–R160.
- 83 Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL: Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 2005;102:15942–15947.
- 84 DeGiorgio M, Jakobsson M, Rosenberg NA: Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proc Natl Acad Sci USA* 2009;106: 16057–16062.
- 85 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.
- 86 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.
- 87 DeGiorgio M, Degnan JH, Rosenberg NA: Coalescence-time distributions in a serial founder model of human evolutionary history. *Genetics* 2011;189:579–593.
- 88 Leutenegger AL, Sahbatou M, Gazal S, Cann H, Genin E: Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? *Eur J Hum Genet* 2011;19:583–587.
- 89 Fraccaro M: Consanguineous marriages in Italy: a note. *Biodemogr Soc Biol* 1957;4:36–39.
- 90 Goldschmidt E, Ronen A, Ronen I: Changing marriage systems in the Jewish communities of Israel. *Ann Hum Genet* 1960;24:191–204.
- 91 Sutter J, Goux JM: Évolution de la consanguinité en France de 1926 à 1958 avec des données récentes détaillées. *Population* 1962; 17:683–702.
- 92 Dronamraju KR, Khan PM: The frequency and effects of consanguineous marriages in Andhra Pradesh. *J Genet* 1963;58:387–401.
- 93 Sanghvi LD: Inbreeding in India. *Eugen Q* 1966;13:291–301.
- 94 Kumar S, Pai RA, Swaminathan MS: Consanguineous marriages and the genetic load due to lethal genes in Kerala. *Ann Hum Genet* 1967;31:141–147.
- 95 Ali SGM: Inbreeding and endogamy in Kerala (India). *Acta Genet* 1968;18:369–379.
- 96 Goswami HK: Frequency of consanguineous marriages in Madhya Pradesh. *Acta Genet Med Gemellol (Roma)* 1970;19:486–490.
- 97 Sundar Rao PS, Inbaraj SG, Kaliaperumal VG: An epidemiological study of consanguinity in a large south Indian town. *Indian J Med Res* 1971;59:294–301.
- 98 Scott-Emuakpor AB: The mutation load in an African population. 1. An analysis of consanguineous marriages in Nigeria. *Am J Hum Genet* 1974;26:674–682.
- 99 Imaizumi Y, Shinozaki N, Aoki A: Inbreeding in Japan: results of a nation-wide study. *Jinrui Idengaku Zasshi* 1975;20:91–107.
- 100 Huq F: Consanguinity and inbreeding among the Muslims of Murshidabad and Birbhum Districts of West Bengal. *J Indian Anthropol Soc* 1976;11:21–25.
- 101 Undevia JV, Balakrishnan V: Temporal changes in consanguinity among the Parsi and Irani communities of Bombay; in Verma IC (ed): *Medical Genetics in India*. Pondicherry, Auroma Enterprises, 1978, vol 2, pp 145–150.
- 102 Malhotra KC: Inbreeding among Dhangar castes of Maharashtra, India. *J Biosoc Sci* 1979;11:397–409.
- 103 Roberts DF, Roberts MJ, Cowie JA: Inbreeding levels in Orkney Islanders. *J Biosoc Sci* 1979;11:391–395.
- 104 Romeo G, Menozzi P, Mastella G, Giunta A, Lodi G, Constantini D: Studio genetico ed epidemiologico della fibrosi cistica in Italia. *Riv Ital Pediatr* 1981;7:201–209.
- 105 Ai Q, Haligamu, Ke Q, Tong J, Zhang B, Wang Z, Jiang Y: A survey of five minority nationalities' consanguineous marriage in Yili, Xinjiang. *Acta Anthropol Sinica* 1985; 4:242–249.
- 106 Bayoumi RA, Taha TSM, Saha N: A study of some genetic characteristics of the Fur and Baggara tribes of the Sudan. *Am J Phys Anthropol* 1985;67:363–370.
- 107 Imaizumi Y: A recent survey of consanguineous marriages in Japan. *Clin Genet* 1986;30: 230–233.
- 108 Calderon R, Pena JA, Morales B, Guevara JI: Inbreeding patterns in the Basque Country (Alava Province, 1831–1980). *Hum Biol* 1993;65:743–770.
- 109 Assaf S, Khawaja M: Consanguinity trends and correlates in the Palestinian Territories. *J Biosoc Sci* 2009;41:107–124.
- 110 Tishkoff SA, Reed FA, Friedlaender FR, et al: The genetic structure and history of Africans and African Americans. *Science* 2009;324: 1035–1044.