B B M R I • N L

Population Structure of the Netherlands

Abdel Abdellaoui, Jouke-Jan Hottenga, Peter de Knijff, Michel G. Nivard, Xiangjun Xiao, Paul Scheet, Andrew Brooks, Erik E. Ehli, Yueshan Hu, Gareth E. Davies, James Hudziak, Patrick F. Sullivan, Toos van Beijsterveldt, Gonneke Willemsen, Eco J. de Geus, Brenda W.J.H. Penninx, Dorret I. Boomsma

Nederlands Tweelingen Register

Nederlandse Studie naar Depressie en Angst

Background

Population genetic studies are of great value for detecting population substructure and making inferences about human history regarding migrations and human evolution. The genetic variation in a population can be summarized by uncorrelated principal components (PCs) through PCA on genome-wide data. The PCs explaining most variation often show striking correlations with geography, a consequence of the decreasing genetic similarity (i.e., shared ancestry) between individuals as geographic distance increases. We investigated how to capture the genetic population structure of the Netherlands and examined whether the relatively recent genetic differentiation observed within the country may have been influenced by migration and/or selection.

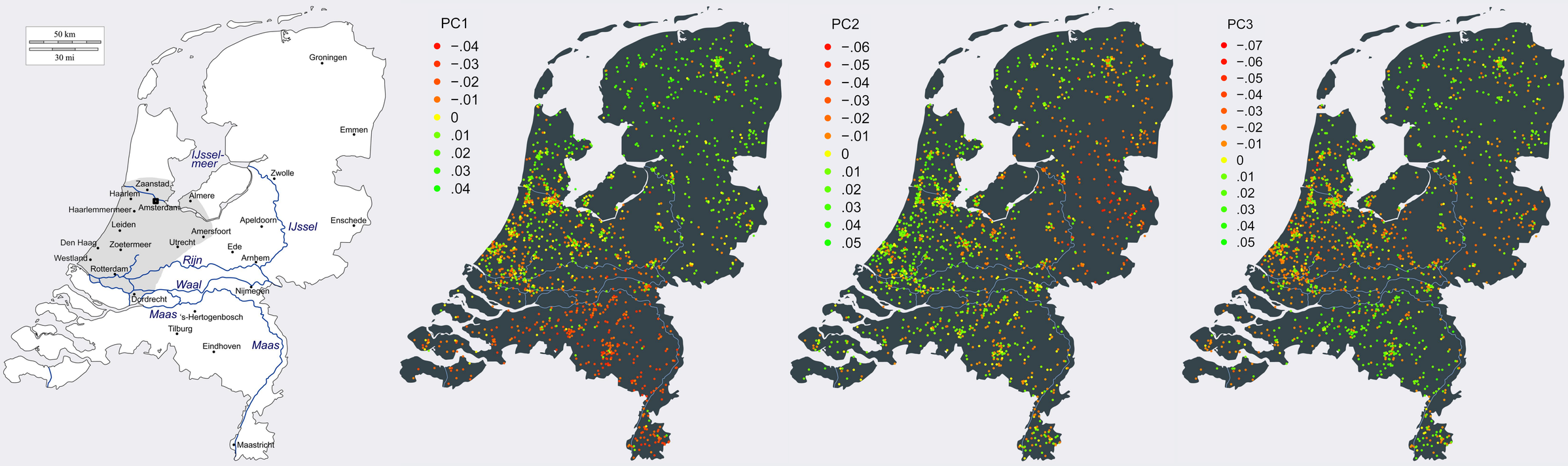


Figure 1: Subjects are plotted on their current living address, colored by their PC value. The PCs come from the SNP set with minimized LD (excluding 24 long-range LD regions and pruned for LD: ~130k SNPs). Correlations with latitude (⇕) and longitude (↔) coordinates based on the current living address: $r_{PC1,\updownarrow}$=.603, $r_{PC2,\leftrightarrow}$=.378, $r_{PC3,\leftrightarrow}$ = .162 (*p*-values < .001).

Methods

PCAs: PCs were computed using EIGENSTRAT on a sample of 4,441 unrelated individuals with Dutch ancestry, using several SNP sets (Affy 6) varying in LD. Correlations between PCs and geography will determine which SNP set captures variation due to ancestry differences best and will be examined further.

Selection pressures: F_{st} values were computed for ~500k SNPs by comparing the top and bottom 1000 individuals of each PC. F_{st} coefficients are decomposed into two components: a population-specific component (β), shared by all loci, and a locus-specific component (α), shared by both populations. If α differs significantly from 0, the locus was likely under selection.

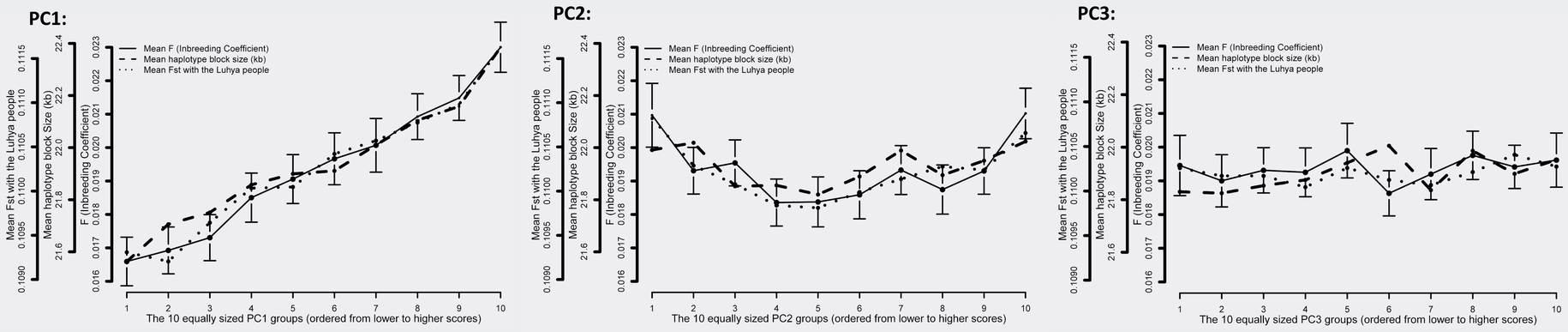
Results

PCAs: The SNP set with minimized LD showed significantly higher correlations between PCs and geography, and was the only one to produce 3 instead of 2 PCs reflecting ancestry (Fig 1).

PC1 showed additional correlations also present in the North-South cline of Europe: F (*r*=.245: likely a serial founder effect due to northwards migration: Fig 2), and height (♂: *r*=.142, ♀: *r*=.153).

Selection pressures: All three PCs showed significant signals (545 SNPs) for diversifying selection through extreme F_{st} values (the majority within 184 genes). The strongest signal was observed between North and South (PC1) in the functional SNP with the largest effect on human blue/brown eye color.

Figure 2: The serial founder effect was observed previously through increasing homozygosity and LD with distance from the human origin: Ethiopia. To illustrate whether this effect is present, subjects were divided into ten groups based on PC score (low to high). For each of the ten groups three related measures were plotted: the mean F (genome-wide homozygosity; with 95% confidence interval), the genome-wide average haplotype block size, and the mean F_{st} with the Luhya people (Kenya).



Conclusions

Traces left by migration and adaptation are detectable in the human genome of a relatively small country. Further research is needed to identify the other functional variants in the regions showing selection pressures. These are likely to influence traits that increased fitness and/or reproductive success. Our results also confirm the importance of considering stratification in association studies, even within smaller supposedly homogeneous populations.

Contact information: a.abdellaoui@vu.nl

VRIJE UNIVERSITEIT AMSTERDAM