R documentation

of 'hdpg.Rd'

June 24, 2006

---

hdpg   *Genetic Variation In Human Populations*

Description

This data set gives genotypes variation of 1066 individuals belonging to 52 predefined populations, for 404 microsatellite markers.

Usage

```
data(hdpg)
```

Format

hdpg is a list of 3 components.

tab   is a data frame with the genotypes of 1066 individuals encoded with 6 characters, (individuals in row, locus in column), for example '123098' for a heterozygote carrying alleles '123' and '098', '123123' for a homozygote carrying two alleles '123' and '000000' for a not classified locus (missing data)

ind   is a a data frame with 4 columns containing information about the 1066 individuals hdgpindid containing the Diversity Panel identification number of each individual, and three factors hdgpindsex, hdgpindpopulation and hdgpindregion containing the names of the 52 populations belonging to 7 major geographic regions (see

> **Details**
>
>)
>
> locus is a dataframe containing four columns : hdgp$locus$marknames a vector of names of the microsatellite markers, hdpg$locus$allbyloc a vector containing the number of alleles by loci hdgp$locus$chromosome hdgp$locus$maposition indicating the position of the locus in the chromosome

Details

The rows of `hdgp$pop` are the names of the 52 populations belonging to the geographic regions contained in the rows of hdgp$region. The chosen regions are: America, Asia, Europe, Middle East North Africa, Oceania, Subsaharan AFRICA.

The 52 populations are: Adygei, Balochi, Bantu, Basque, Bedouin, Bergamo, Biaka Pygmies, Brahui, Burusho, Cambodian, Columbian, Dai, Daur, Druze, French, Han, Hazara, Hezhen, Japanese, Kalash, Karitiana, Lahu, Makrani, Mandenka, Maya, Mbuti Pygmies, Melanesian, Miaozu, Mongola, Mozabite, Naxi, NewGuinea, Nilote, Orcadian, Oroqen, Palestinian, Pathan, Pima, Russian, San, Sardinian, She, Sindhi, Surui, Tu, Tujia, Tuscan, Uygur, Xibo, Yakut, Yizu, Yoruba.

`hdgp$freq` is a data frame with 52 rows, corresponding to the 52 populations described above, and 4992 microsatellite markers.

Source

Extract of data prepared by the Human Diversity Panel Genotypes `http://research.marshfieldclinic.org/genetics/Freq/FreqInfo.htm`

prepared by Hinda Haned, from data used in: Noah A. Rosenberg, Jonatahan K. Pritchard, James L. Weber, Howard M. Cabb, Kenneth K. Kidds, Lev A. Zhivotovsky, Marcus W. Feldman (2002) Genetic Structure of human Populations *Science*, **298**, 2381–2385.

Lev A. Zhivotovsky, Noah Rosenberg, and Marcus W. Feldman (2003). Features of Evolution and Expansion of Modern Humans, Inferred from Genomewide Microsatellite Markers *Am. J. Hum. Genet*, **72**, 1171–1186.

Examples

```
## Not run:
library(ade4)
data(hdpg)
freq <- char2genet(hdpg\$tab, hdpg\$ind\$population)
vec <- apply(freq\$tab, 2, function(c) mean(c, na.rm = TRUE))
for (j in 1:4492){
    freq\$tab[is.na(freq\$tab[,j]),j] = vec[j]}
pcatot <- dudi.pca(freq\$tab, center = TRUE, scale = FALSE, scannf = FALSE, nf = 4)
s.label(pcatot\$li, xax = 1, yax = 2, sub = "1-2", lab = freq\$pop.names)
## End(Not run)
```

Index