

Evaluation of age, sex, and ancestry-related variation in cortical bone and dentine volumes in modern humans, and a preliminary assessment of cortical bone-dentine covariation in later Homo

Supporting Information: Statistical analyses

Mathilde Augoyard^{1*}, Clément Zanolli^{1,2}, Frédéric Santos¹, Anna C. Oetllé³, Ericka N. L'Abbé⁴, Mona Le Luyer^{1,5,6}, Marine Cazenave^{7,8,9}, Thomas Colard^{1,10}, Jakobus Hoffman¹¹, Antonio Profico¹², and Priscilla Bayle^{1*}

¹*Univ. Bordeaux, CNRS, MCC, PACEA, UMR5199, F-33600 Pessac, France.*

²*Department of Maxillofacial and Oral Surgery, Sefako Makgatho Health Sciences University, Ga-Rankuwa, Pretoria, South Africa.*

³*Department of Anatomy and Histology, Sefako Makgatho Health Sciences University, Ga-Rankuwa, Pretoria, South Africa.*

⁴*Forensic Anthropology Research Centre, Department of Anatomy, Faculty of Health Sciences, University of Pretoria, Pretoria 0031, South Africa*

⁵*Center for Genomic Medicine, Massachusetts General Hospital, Boston, USA.*

⁶*Department of Psychiatry, Harvard Medical School, Boston, USA.*

⁷*Division of Anthropology, American Museum of Natural History, New York, NY, USA.*

⁸*Skeletal Biology Research Centre, School of Anthropology and Conservation, University of Kent, Canterbury, UK.*

⁹*Department of Anatomy, Faculty of Health Sciences, University of Pretoria, South Africa*

¹⁰*Department of Oral Radiology, University of Lille, Lille University Hospital, Lille, France.*

¹¹*South African Nuclear Energy Corporation SOC Ltd., Pelindaba, South Africa.*

¹²*Department of Biology, University of Pisa, Pisa, Italy.*

**Corresponding authors.*

March 2, 2023

Contents

| | | |
|----------|--|----------|
| 1 | Configuration | 3 |
| 2 | Load and summarize data | 4 |
| 2.1 | Read the whole dataset | 4 |
| 2.2 | Creation of various subsets of modern humans | 5 |
| 3 | Table 2 | 6 |
| 3.1 | Wilcoxon tests | 6 |
| 3.2 | BCa confidence intervals | 6 |

| | | |
|----------|---|----------|
| 4 | Table 3 | 7 |
| 5 | Resampling experiments for chimeric individuals | 9 |
| 5.1 | Build a sample of chimeric modern individuals | 9 |
| 5.2 | Build a sample of chimeric Neandertals | 9 |

1. Configuration

This document aims to facilitate the replication of a subset of results (tables or figures) presented in our article. As stated in the main text of the article, our data are hosted on Zenodo (Augoyard et al., 2023). All the analyses were performed using R (R Core Team, 2021), and this document has been built with Org mode 9.6.1 for Emacs 28.2 (Schulte, Davison, Dye, & Dominik, 2012).

Along with R version 4.2.2 (2022-10-31) itself, the following R packages are loaded, using their version available on CRAN at a fixed date (2022-11-15), using the {groundhog} R package (Simonsohn & Gruson, 2021) :

```
## Use groundhog to improve the reproducibility:
library(groundhog)
gday <- "2022-11-15"
## Load the following packages:
groundhog.library("boot", date = gday)
groundhog.library("corrplot", date = gday)
groundhog.library("dplyr", date = gday)
groundhog.library("factoextra", date = gday)
groundhog.library("FactoMineR", date = gday)
groundhog.library("forcats", date = gday)
```

Full details about the R session:

```
print(sessionInfo(), locale = FALSE)
```

```
R version 4.2.2 (2022-10-31)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Manjaro Linux

Matrix products: default
BLAS: /usr/lib/libopenblas.so.0.3
LAPACK: /usr/lib/liblapack.so.3.11.0

attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] forcats_0.5.2 FactoMineR_2.6 factoextra_1.0.7 ggplot2_3.4.0
[5] dplyr_1.0.10 corrplot_0.92 boot_1.3-28 groundhog_2.2.0
```

For all technical questions about this document or the R scripts included hereafter, or for any issue in reproducing the results, feel free to send an email to frederic.santos@u-bordeaux.fr.

2. Load and summarize data

2.1. Read the whole dataset

```
## Load CSV file from Zenodo:
dat <- read.csv2(
  file = "https://zenodo.org/record/7691175/files/data.csv?download=1",
  row.names = 2,
  stringsAsFactors = TRUE,
  na.strings = "",
  fileEncoding = "utf-8"
)
## Recode sex factor:
levels(dat$Sex) <- c("F", "M", NA)
## Create age category :
dat$Age_categ <- fct_collapse(
  .f = dat$Age_group,
  "15-19" = "15_19",
  "9-14" = "9_14",
  other_level = "Adult"
)
summary(dat, maxsum = 8)
```

| Taxon | Age_group | Sex | Origin | Ancestry | RL |
|----------------|---------------|---------------|-----------------|------------------|---------------|
| MH :61 | 15_19 : 7 | F :25 | France :29 | African :30 | Min. :162.7 |
| NEAND: 7 | 20_29 :14 | M :25 | South Africa:32 | European:31 | 1st Qu.:217.5 |
| | 20_39 :19 | NA's:18 | NA's : 7 | NA's : 7 | Median :232.0 |
| | 20_49 : 2 | | | | Mean :233.5 |
| | 9_14 : 5 | | | | 3rd Qu.:254.8 |
| | sup_30: 1 | | | | Max. :281.2 |
| | sup_40:13 | | | | NA's :5 |
| | NA's : 7 | | | | |
| log10_RL | Vcor_20_80 | sVcor_20_80 | Vcor_20_30 | log10_Vcor_20_30 | |
| Min. :2.210 | Min. : 4364 | Min. :26.82 | Min. : 721.5 | Min. :2.860 | |
| 1st Qu.:2.340 | 1st Qu.: 9796 | 1st Qu.:44.08 | 1st Qu.:1558.7 | 1st Qu.:3.190 | |
| Median :2.370 | Median :12373 | Median :52.25 | Median :1819.4 | Median :3.260 | |
| Mean :2.366 | Mean :12519 | Mean :52.58 | Mean :1948.5 | Mean :3.269 | |
| 3rd Qu.:2.410 | 3rd Qu.:16080 | 3rd Qu.:61.50 | 3rd Qu.:2434.1 | 3rd Qu.:3.390 | |
| Max. :2.450 | Max. :18861 | Max. :74.02 | Max. :3266.8 | Max. :3.510 | |
| NA's :7 | NA's :7 | NA's :7 | NA's :5 | NA's :7 | |
| sVcor_20_30 | DL | log10_DL | Ve | sVe | |
| Min. : 4.430 | Min. :20.11 | Min. :1.330 | Min. : 67.86 | Min. :2.670 | |
| 1st Qu.: 6.895 | 1st Qu.:24.18 | 1st Qu.:1.390 | 1st Qu.: 94.61 | 1st Qu.:3.720 | |
| Median : 8.220 | Median :25.70 | Median :1.410 | Median :111.27 | Median :4.310 | |
| Mean : 8.224 | Mean :25.83 | Mean :1.416 | Mean :116.23 | Mean :4.425 | |
| 3rd Qu.: 9.395 | 3rd Qu.:27.50 | 3rd Qu.:1.440 | 3rd Qu.:128.59 | 3rd Qu.:5.030 | |
| Max. :12.820 | Max. :32.71 | Max. :1.510 | Max. :200.38 | Max. :7.530 | |
| NA's :5 | NA's :2 | NA's :7 | NA's :7 | NA's :7 | |

| Vp_tot | sVp_tot | Vd_tot | sVd_tot | Vd_50_90 |
|----------------|-----------------|----------------|----------------|----------------|
| Min. : 8.86 | Min. : 0.3700 | Min. : 318.0 | Min. : 13.91 | Min. : 110.1 |
| 1st Qu.: 15.10 | 1st Qu.: 0.6400 | 1st Qu.: 382.5 | 1st Qu.: 15.78 | 1st Qu.: 139.0 |
| Median : 19.94 | Median : 0.7800 | Median : 443.6 | Median : 17.37 | Median : 166.9 |
| Mean : 21.97 | Mean : 0.8303 | Mean : 496.0 | Mean : 18.75 | Mean : 192.4 |
| 3rd Qu.: 28.41 | 3rd Qu.: 1.0500 | 3rd Qu.: 612.2 | 3rd Qu.: 21.97 | 3rd Qu.: 228.5 |
| Max. : 50.79 | Max. : 1.7200 | Max. : 959.1 | Max. : 29.32 | Max. : 446.1 |
| NA's : 7 | NA's : 7 | NA's : 7 | NA's : 7 | NA's : 2 |

| log10_Vd_50_90 | sVd_50_90 | Vd_crown | log10_Vd_crown | sVd_crown |
|----------------|----------------|----------------|----------------|----------------|
| Min. : 2.04 | Min. : 4.750 | Min. : 110.6 | Min. : 2.040 | Min. : 4.560 |
| 1st Qu.: 2.14 | 1st Qu.: 5.730 | 1st Qu.: 135.6 | 1st Qu.: 2.130 | 1st Qu.: 5.473 |
| Median : 2.21 | Median : 6.450 | Median : 156.2 | Median : 2.190 | Median : 6.390 |
| Mean : 2.24 | Mean : 7.441 | Mean : 178.3 | Mean : 2.218 | Mean : 6.900 |
| 3rd Qu.: 2.33 | 3rd Qu.: 8.300 | 3rd Qu.: 214.1 | 3rd Qu.: 2.320 | 3rd Qu.: 7.595 |
| Max. : 2.58 | Max. : 18.590 | Max. : 338.9 | Max. : 2.450 | Max. : 14.350 |
| NA's : 7 | NA's : 2 | NA's : 2 | NA's : 7 | NA's : 2 |

| sVd_50_90.sVcor_20_30 | sVd_crown.sVcor_20_30 | Age_categ |
|-----------------------|-----------------------|-----------|
| Min. : 0.4800 | Min. : 0.5100 | 15-19: 7 |
| 1st Qu.: 0.7400 | 1st Qu.: 0.6900 | 9-14 : 5 |
| Median : 0.8400 | Median : 0.7800 | Adult: 49 |
| Mean : 0.8546 | Mean : 0.8102 | NA's : 7 |
| 3rd Qu.: 0.9500 | 3rd Qu.: 0.9100 | |
| Max. : 1.3300 | Max. : 1.2700 | |
| NA's : 7 | NA's : 7 | |

2.2. Creation of various subsets of modern humans

From the whole dataframe, we extract various subsets one can use in subsequent analyses:

```
## Dataset of modern humans only:
mh <- subset(dat, Taxon == "MH")
## Dataset of complete individuals (without any NA value):
complete <- na.omit(mh)
## Subset of adult individuals:
adults <- droplevels(subset(mh, Age_categ == "Adult"))
## Subset of individuals of known sex:
sex <- droplevels(subset(mh, ! is.na(Sex)))
## Subset of immature individuals:
immat <- droplevels(subset(mh, Age_categ != "Adult"))
## Subset to compare Adults vs. [9-14[ y.o. :
comp1 <- droplevels(subset(mh, Age_categ != "15-19"))
## Subset to compare Adults vs. [15-19[ y.o. :
comp2 <- droplevels(subset(dat, Age_categ != "9-14"))
```

3. Table 2

We provide here the R code used for the computation of Wilcoxon tests, and of bias-corrected and accelerated (BCa) bootstrap confidence intervals for the difference between means (Zi-
effler, Harring, & Long, 2011). We illustrate this in the particular case of the comparison
between female and male individuals for the variable `sVcor_20_30`; the code can easily be
adapted for all other comparisons presented in Table 2.

3.1. Wilcoxon tests

```
## Comparison female vs. males:  
wilcox.test(sVcor_20_30 ~ Sex, data = sex)
```

```
Wilcoxon rank sum exact test  
  
data: sVcor_20_30 by Sex  
W = 26, p-value = 1.856e-10  
alternative hypothesis: true location shift is not equal to 0
```

3.2. BCa confidence intervals

We first define a custom function for the bootstrap procedure:

```
diff_means <- function(data, indices) {  
  data <- data[indices, ]  
  group1 <- subset(data, Sex == "F")  
  group2 <- subset(data, Sex == "M")  
  return(mean(group1$sVcor_20_30) - mean(group2$sVcor_20_30))  
}
```

Then we perform the resampling experiments, and finally compute the bootstrap confidence interval for the difference between means. The resampling experiments are done using $R = 4999$ bootstrap replicates.

```
## Perform R=4999 resampling experiments:  
res_boot <- boot(  
  data = sex,  
  strata = sex$Sex,  
  R = 4999,  
  statistic = diff_means,  
  stype = "i"  
)  
  
## Compute a 95% BCa confidence interval:  
boot.ci(res_boot, type = "bca", conf = 0.95)
```

```

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 4999 bootstrap replicates

CALL :
boot.ci(boot.out = res_boot, conf = 0.95, type = "bca")

Intervals :
Level      BCa
95%      (-3.110, -1.937 )
Calculations and Intervals on Original Scale

```

Since we did not set explicitly the random seed in the R code above, the result may vary slightly from one execution to another, and in particular from the results presented in Table 2 in the main text. These small variations let however the global interpretation unchanged.

4. Table 3

We provide here one of the six correlation matrices presented in Table 3, namely the correlation matrix for all adults ($n = 49$). The code can easily be adapted to the other situations.

```

## Compute the Spearman correlation matrix:
input <- select(
  .data = adults,
  "sVcor_20_30", "sVd_50_90", "sVd_crown"
)
mat <- cor(input, method = "spearman")

## Perform correlation tests:
pmat <- cor.mtest(input, method = "spearman")

## Graphical representation of the correlation matrix:
corrplot(
  corr = mat,
  p.mat = pmat$p,
  sig.level = 0.05,
  addCoef.col = "white",
  number.cex = 1.2,
  type = "upper"
)

```

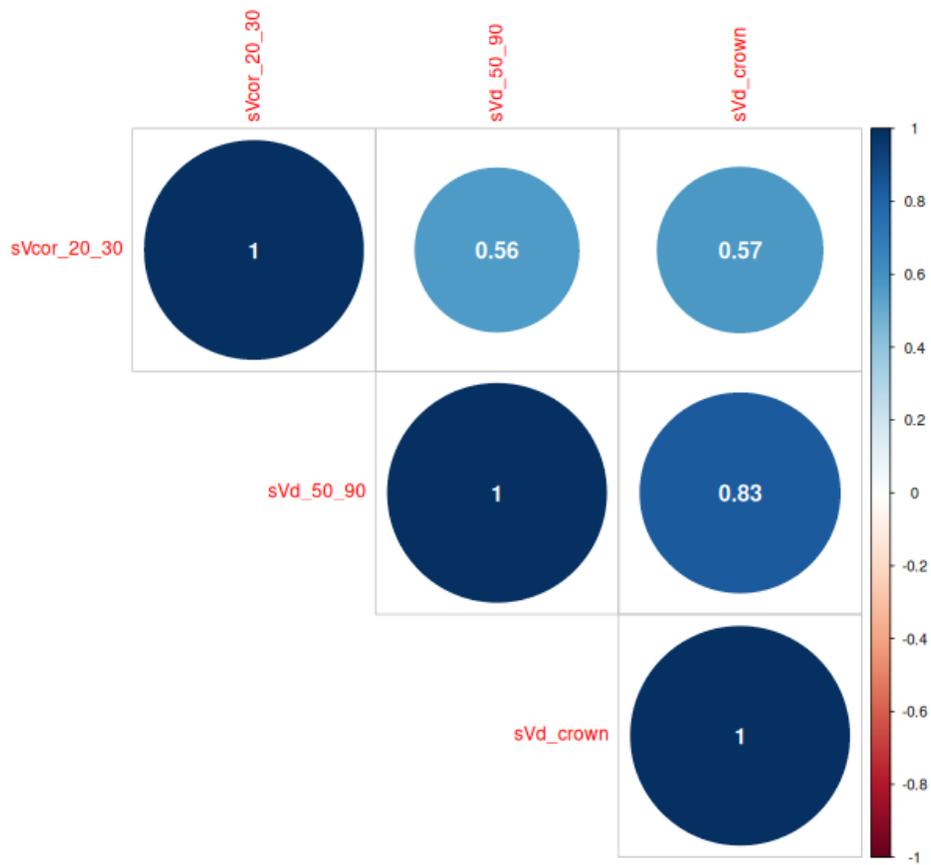


Figure 1: Spearman correlation matrix for all adults. Pairs of variables whose correlation coefficient is not significantly different from 0 (if any) are indicated by a black cross.

5. Resampling experiments for chimeric individuals

5.1. Build a sample of chimeric modern individuals

The R code below selects a subset of 12 modern humans, and then build a dataframe of 132 (= 12 × 11) “chimeric” (i.e., composite) individuals as follows. Each of these 132 chimeric individuals is formed using two bones-related variables of one modern individual, and four dentine-related variables of another modern individual.

```
## Modern individuals used for the resampling experiment:
inds <- c("SEP_181", "SEP_264", "SEP_571", "SEP_157", "SEP_509",
          "1943", "2060", "1278", "1825", "6177", "6290", "6338")
## Subset of variables relative to cortical bone:
vcort <- c("Vcor_20_30", "sVcor_20_30")
## Subset of variables relative to dentine:
vdent <- c("Vd_crown", "sVd_crown", "Vd_50_90", "sVd_50_90")
## Build composite/chimeric sample :
pairs <- expand.grid(inds, inds) |>
  as.data.frame() |>
  mutate(across(everything(), as.character)) |>
  subset(Var1 != Var2)
hchim <- data.frame(
  Taxon = factor(rep("MH", nrow(pairs))),
  dat[pairs[, 1], vcort], # bone part
  dat[pairs[, 2], vdent] # dentine part
)
rownames(hchim) <- paste(pairs[, 1], pairs[, 2], sep = "/")
```

5.2. Build a sample of chimeric Neandertals

We also build (manually this time) five chimeric Neandertals, as reported in the main text.

```
## Build chimeric Neandertals:
neandbone <- c(rep("REG1", 3), rep("SpyII", 2))
neanddent <- c("Kr36", "Kr37", "Kr76", "Vi12_5", "Palomas35")
nchim <- data.frame(
  Taxon = factor(c(rep("NeandMIS5", 3), rep("NeandMIS3", 2))),
  dat[neandbone, vcort],
  dat[neanddent, vdent]
)
rownames(nchim) <- paste(neandbone, neanddent, sep = "/")
print(nchim)
```

| | Taxon | Vcor_20_30 | sVcor_20_30 | Vd_crown | sVd_crown | Vd_50_90 |
|--------------|-----------|------------|-------------|----------|-----------|----------|
| REG1/Kr36 | NeandMIS5 | 2290.85 | 9.87 | 338.89 | 14.12 | 446.13 |
| REG1/Kr37 | NeandMIS5 | 2290.85 | 9.87 | 288.68 | 14.35 | 312.70 |
| REG1/Kr76 | NeandMIS5 | 2290.85 | 9.87 | 242.03 | 10.44 | 340.26 |
| SpyII/Vi12_5 | NeandMIS3 | 1708.99 | 8.31 | 306.43 | 14.03 | 348.82 |

| | | | | | | |
|-----------------|-----------|---------|------|--------|------|--------|
| SpyII/Palomas35 | NeandMIS3 | 1708.99 | 8.31 | 149.94 | 7.05 | 187.34 |
| | sVd_50_90 | | | | | |
| REG1/Kr36 | | 18.59 | | | | |
| REG1/Kr37 | | 15.55 | | | | |
| REG1/Kr76 | | 14.68 | | | | |
| SpyII/Vi12_5 | | 15.97 | | | | |
| SpyII/Palomas35 | | 8.81 | | | | |

References

- Augoyard, M., Zanolli, C., Santos, F., Oetllé, A., L'Abbé, E. N., Le Luyer, M., ... Priscilla, P. (2023). *Data file for Augoyard et al.'s article, "Evaluation of age, sex, and ancestry-related variation in cortical bone and dentine volumes in modern humans, and a preliminary assessment of cortical bone-dentine covariation in later Homo"*. Zenodo. doi:10.5281/ZENODO.6341546
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Schulte, E., Davison, D., Dye, T., & Dominik, C. (2012). A Multi-Language Computing Environment for Literate Programming and Reproducible Research. *Journal of Statistical Software*, 46(1), 1–24. doi:10.18637/jss.v046.i03
- Simonsohn, U., & Gruson, H. (2021). *Groundhog: Reproducible Scripts via Version-Specific Package Loading*.
- Zieffler, A., Harring, J., & Long, J. D. (2011). *Comparing groups: Randomization and bootstrap methods using R*. Hoboken, N.J: Wiley.