Tooth variability in Pleistocene and Recent dhole, *Cuon alpinus*

**(Carnivora, Mammalia)**

Gennady F. Baryshnikov1 and Andrey Yu. Puzachenko2

1Zoological Institute of Russian Academy of Sciences, St. Petersburg, Russia;

2Institute of Geography of Russian Academy of Sciences, Moscow, Russia;

**Supplementary Material**

**Contents**

[**1. Supplementary Material for the section “Materials and methods”** 3](#_Toc172213270)

[Table S1. Number of teeth from extant and fossil *C. alpinus* used in the study 3](#_Toc172213271)

[Table S2 Number of teeth of the extant subspecies of *C. alpinus* of known sex (according to the labels of the museums) 4](#_Toc172213272)

[**2. Supplementary Material for the section “Results”** 5](#_Toc172213273)

[Table S3 Correlations of extant *C. alpinus* tooth measurements and indices with descriptive model coordinates (E1, E2, K1–K3) and multiple *r*2 for multivariate linear regression model y = a1E1+a2E2+a3K1+a4K2+a5K3 + a0 (backward stepwise method) 5](#_Toc172213274)

[Table S4. Mean (M), standard error (m) and min-max of tooth measurements and indices *in C. a. alpinus* and *C. a. hespericus* and statistical significance of differences between subspecies using Mann-Whitney test (Bonferroni corrected *p*) 6](#_Toc172213275)

[Table S5. Correlations of *C. a. alpinus* and *C. a. hesperius* tooth measurements and indices with coordinates of SZM and SHM descriptive models (E1–E4, K1–K3) and multiple *r*2 for multivariate linear regression model y = a1E1+a2E2...+... a6K2+a7K3 + a0 (backward stepwise method) 6](#_Toc172213276)

[Fig. S1. Projections onto the first two coordinates of descriptive models of variation in tooth size (left) and shape (right) in *C. a. alpinus* and *C. a. hesperius* 8](#_Toc172213277)

[Table S6. Correlations of recent C. alpinus ssp. except C. a. alpinus and *C. a. hesperius*, tooth measurements and indices with coordinates of SZM and SHM descriptive models (E1, E4, K1, K2) and multiple *r*2 for multivariate linear regression model y = a1E1+a2E2+a3K1+a4K2 + a0 (backward stepwise method) 8](#_Toc172213278)

[Fig. S2. Descriptive models of variability in tooth size (a) and shape (b) in the extant “southern” group of subspecies of *C. alpinus*; c, d - variability in P4L and p4L/m1L index. 9](#_Toc172213279)

[**3. Supplementary Material for “Assessing the "quality" of the a priori taxonomic partitioning of the sample”** 10](#_Toc172213280)

[**References** 10](#_Toc172213281)

[**List of libraries of R packages used:** 10](#_Toc172213282)

[**Abbreviation:** 10](#_Toc172213283)

[**R code:** 10](#_Toc172213284)

[**##Package for partitioning around medoids with estimation of number of clusters (library(fpc))** 11](#_Toc172213285)

[**##NbClust - Package for determining the best number of clusters** (**library(NbClust))** 11](#_Toc172213286)

[**##Gaussian finite mixture model fitted by EM algorithm (library(Mclust))** 11](#_Toc172213287)

[**Protocol** 12](#_Toc172213288)

[**1) Upper canine** 12](#_Toc172213289)

[**2) Upper premolar P4** 15](#_Toc172213290)

[**3) Lower canine c** 19](#_Toc172213291)

[**4) Lower premolar p2** 22](#_Toc172213292)

[**5) Lower premolar p4** 25](#_Toc172213293)

[**5) Lower molar m1** 27](#_Toc172213294)

**1. Supplementary Material for the section “Materials and methods”**

Table S1. Number of teeth from extant and fossil *C. alpinus* used in the study

| **Tooth** | **Subspecies or locality** | **Extant (E) or fossil (F)** | **N** |
| --- | --- | --- | --- |
| **C** | *C. a. hesperius* | E | 7 |
| *C. a. alpinus* | E | 12 |
| *C. a.* *javanicus* | E | 1 |
| *C. a. caucasicus* | F | 8 |
| *C. a. europaeus* | F | 3 |
| Bliznetz Cave | F | 1 |
| **P4** | *C. a. hesperius* | E | 10 |
| *C. a. alpinus* | E | 13 |
| *C. a. laniger* | E | 2 |
| *C. a. primaevus* | E | 10 |
| *C. a. dukhunensis* | E | 20 |
| *C. a. adustus* | E | 8 |
| *C. a. infuscus*/*C. a. sumatrensis* | E | 3 |
| *C. a. javanicus* | E | 7 |
| *C. a. javanicus* (*antiquus*) | F | 2 |
| *C. a. caucasicus* | F | 4 |
| *C. a. europaeus* | F | 13 |
| *C. a. priscus* | F | 4 |
| Bliznetz Cave | F | 2 |
| **c** | *C. a. hesperius* | E | 7 |
| *C. a. alpinus* | E | 11 |
| *C. a.* *javanicus* | E | 1 |
| *C. a. caucasicus* | F | 6 |
| *C. a. europaeus* | F | 8 |
| Bliznetz Cave | F | 1 |
| *C. a. fossilis* | F | 1 |
| **p2** | *C. a. hesperius* | E | 7 |
| *C. a. alpinus* | E | 12 |
| *C. a. laniger* | E | 1 |
| *C. a. primaevus* | E | 10 |
| *C. a. dukhunensis* | E | 20 |
| *C. a. adustus* | E | 7 |
| *C. a. infuscus*/*C. a. sumatrensis* | E | 2 |
| *C. a.* *javanicus* | E | 5 |
| *C. a. caucasicus* | F | 2 |
| *C. a. europaeus* | F | 17 |
| *C. a. nizhneudensis* | F | 1 |
| *C. a. fossilis* | F | 3 |
| San Josechito Cave | F | 2 |
| Bliznetz Cave | F | 1 |
| Mocun Cave | F | 1 |
| **p4** | *C. a. hesperius* | E | 9 |
| *C. a. alpinus* | E | 12 |
| *C. a. laniger* | E | 1 |
| *C. a. primaevus* | E | 10 |
| *C. a. dukhunensis* | E | 20 |
| *C. a. adustus* | E | 8 |
| *C. a. infuscus*/*C. a. sumatrensis* | E | 3 |
| *C. a.* *javanicus* | E | 7 |
| *C. a. javanicus* (*antiquus*) | F | 1 |
| *C. a. caucasicus* | F | 6 |
| *C. a. europaeus* | F | 23 |
| *C. a. nizhneudensis* | F | 2 |
| *C. a. fossilis* | F | 4 |
| Geographicheskogo Obshestva Cave (Geographical Society Cave) | F | 1 |
| San Josechito Cave | F | 2 |
| **m1L** | *C. a. hesperius* | E | 9 |
| *C. a. alpinus* | E | 12 |
| *C. a. laniger* | E | 1 |
| *C. a. primaevus* | E | 10 |
| *C. a. dukhunensis* | E | 20 |
| *C. a. adustus* | E | 8 |
| *C. a. rutilans* | E | 1 |
| *C. a. infuscus*/*C. a. sumatrensis* | E | 3 |
| *C. a.* *javanicus* | E | 9 |
| *C. a. javanicus*(*antiquus*) | F | 1 |
| *C. a. caucasicus* | F | 6 |
| *C. a. europaeus* | F | 40 |
| *C. a. priscus* | F | 2 |
| *C. a. fossilis* | F | 11 |
| *C. a. crassidens* | F | 1 |
| Geographicheskogo obshestva Cave | F | 1 |
| San Josechito Cave\* | F | 2 |
| **Total** | | | 502 |

Comment. \* – We used the minimum and maximum values of the m1 measurements according to (Tedford et al. 2009).

Table S2 Number of teeth of the extant subspecies of *C. alpinus* of known sex (according to the labels of the museums)

|  |  |  |  |
| --- | --- | --- | --- |
| **Tooth** | **Subspecies** | **Females** | **Males** |
| **C** | *C. a. hesperius* | 1 | - |
| *C. a. alpinus* | 4 | 3 |
| **P4** | *C. a. hesperius* | 1 | - |
| *C. a. alpinus* | 4 | 4 |
| *C. a. primaevus* | 4 | 4 |
| *C. a. dukhunensis* | 4 | 9 |
| *C. a. adustus* | - | 3 |
| *C. a. infuscus*/ *C. a. sumatrensis* | - | 2 |
| *C. a.* *javanicus* | 1 | 4 |
| **c** | *C. a. hesperius* | 1 | - |
| *C. a. alpinus* | 4 | 3 |
| **p2** | *C. a. hesperius* | 1 | 1 |
| *C. a. alpinus* | 4 | 4 |
| *C. a. primaevus* | 2 | 4 |
| *C. a. dukhunensis* | 4 | 9 |
| *C. a. adustus* | - | 3 |
| *C. a. infuscus*/ *C. a. sumatrensis* | - | 1 |
| *C. a.* *javanicus* | 1 | 3 |
| **p4** | *C. a. hesperius* | 1 | - |
| *C. a. alpinus* | 4 | 4 |
| *C. a. primaevus* | 2 | 4 |
| *C. a. dukhunensis* | 4 | 9 |
| *C. a. adustus* | - | 3 |
| *C. a. infuscus*/ *C. a. sumatrensis* | - | 2 |
| *C. a.* *javanicus* | 1 | 4 |
| **m1L** | *C. a. hesperius* | 1 | 1 |
| *C. a. alpinus* | 4 | 4 |
| *C. a. primaevus* | 2 | 4 |
| *C. a. dukhunensis* | 4 | 9 |
| *C. a. infuscus*/ *C. a. sumatrensis* | - | 2 |
| *C. a. adustus* | - | 3 |
| *C. a.* *javanicus* | 1 | 4 |

**2. Supplementary Material for the section “Results”**

Table S3 Correlations of extant *C. alpinus* tooth measurements and indices with descriptive model coordinates (E1, E2, K1–K3) and multiple *r*2 for multivariate linear regression model y = a1E1+a2E2+a3K1+a4K2+a5K3 + a0 (backward stepwise method)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measu-rements** | **E1** | **E2** | **K1** | **K2** | **K3** | ***r*2** | **Index** | **E1** | **E2** | **K1** | **K2** | **K3** |
| **CL** | 0.72 | 0.30 | 0.06 | 0.37 | -0.28 | 0.51 | **CW/CL** | -0.07 | 0.13 | -0.21 | -0.17 | 0.26 |
| **CW** | 0.37 | 0.32 | -0.18 | 0.05 | 0.11 | 0.0 | **P4W/P4L** | 0.69 | -0.68 | 0.89 | 0.41 | 0.18 |
| **P4L** | 0.70 | 0.48 | -0.06 | 0.30 | 0.21 | 0.92 | **P4lp/P4L** | -0.42 | 0.24 | -0.48 | -0.07 | -0.58 |
| **P4lp** | 0.39 | 0.70 | -0.46 | 0.25 | -0.26 | 0.71 | **cW/cL** | -0.28 | -0.10 | 0.02 | -0.34 | 0.37 |
| **P4W** | 0.84 | -0.47 | 0.78 | 0.46 | 0.24 | 0.97 | **p2L/m1L** | 0.72 | -0.64 | 0.92 | 0.45 | 0.08 |
| **cL** | 0.43 | 0.50 | -0.24 | 0.10 | -0.38 | 0.45 | **p4L/m1L** | 0.73 | 0.00 | 0.30 | 0.73 | -0.18 |
| **cW** | 0.21 | 0.58 | -0.34 | -0.36 | -0.02 | 0.33 | **m1W/m1L** | -0.16 | 0.17 | -0.30 | -0.58 | 0.31 |
| **p2L** | 0.85 | -0.47 | 0.85 | 0.48 | -0.04 | 0.96 | **m1tal/m1L** | -0.23 | -0.32 | 0.21 | -0.16 | -0.09 |
| **p4L** | 0.87 | 0.25 | 0.21 | 0.61 | -0.35 | 0.93 |  |  |  |  |  |  |
| **m1L** | 0.76 | 0.51 | 0.02 | 0.25 | -0.43 | 0.89 |  |  |  |  |  |  |
| **m1tal** | 0.18 | -0.06 | 0.23 | -0.02 | -0.31 | 0.0 |  |  |  |  |  |  |
| **m1W** | 0.63 | 0.63 | -0.19 | -0.17 | -0.22 | 0.90 |  |  |  |  |  |  |

Table S4. Mean (M), standard error (m) and min-max of tooth measurements and indices *in C. a. alpinus* and *C. a. hesperius* and statistical significance of differences between subspecies using Mann-Whitney test (Bonferroni corrected *p*)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Measurement** | ***C. a. alpinus*** | | **N** | ***C. a. hesperius*** | | **N** | ***p*** |
| **M±m** | **Min-Max** | **M±m** | **Min-Max** |
| **CL** | 10.1±0.1 | 9.5-10.7 | 12 | 9.5±0.15 | 9.0-10.1 | 7 | 0.06 |
| **CW** | 5.7±0.11 | 5.0-6.3 | 12 | 5.8±0.24 | 4.9-6.9 | 7 | 0.55 |
| **P4L** | 20.5±0.16 | 19.7-21.5 | 13 | 20.5±0.18 | 19.8-21.3 | 10 | 0.97 |
| **P4Lp** | 12.8±0.11 | 12.2-13.7 | 12 | 12.7±0.11 | 12.2-13.2 | 8 | 0.93 |
| **P4W** | 9.7±0.27 | 7.0-10.8 | 13 | 9.1±0.38 | 6.7-10.5 | 10 | 0.14 |
| **cL** | 10.1±0.21 | 8.9-11.1 | 11 | 9.3±0.17 | 8.7-9.9 | 7 | 0.09 |
| **cW** | 6.6±0.1 | 6.0-7.0 | 11 | 6.7±0.12 | 6.1-7.1 | 7 | 0.42 |
| **p2L** | 9.0±0.35 | 5.3-9.9 | 12 | 7.6±0.6 | 4.9-9.3 | 7 | 0.009 |
| **p4L** | 13.2±0.27 | 10.7-14.5 | 12 | 12.5±0.2 | 11.6-13.7 | 9 | 0.02 |
| **m1L** | 22.7±0.16 | 21.7-23.7 | 12 | 21.8±0.28 | 21.0-23.2 | 9 | 0.03 |
| **m1tal** | 6.4±0.1 | 5.8-6.9 | 11 | 6.0±0.29 | 5.0-6.8 | 6 | 0.09 |
| **m1W** | 8.6±0.07 | 8.3-9.1 | 12 | 8.5±0.13 | 7.7-9.2 | 9 | 0.47 |
| **Index** | **M±m** | **Min-Max** | **N** | **M±m** | **Min-Max** | **N** | ***p*** |
| **CW/CL** | 0.56±0.01 | 0.49-0.61 | 12 | 0.61±0.026 | 0.52-0.71 | 7 | 0.09 |
| **P4W/P4L** | 0.47±0.011 | 0.36-0.51 | 13 | 0.44±0.018 | 0.34-0.50 | 10 | 0.08 |
| **P4lp/P4L** | 0.62±0.006 | 0.59-0.65 | 12 | 0.62±0.005 | 0.60-0.64 | 8 | 0.97 |
| **cW/cL** | 0.66±0.013 | 0.58-0.71 | 11 | 0.72±0.015 | 0.68-0.77 | 7 | 0.05 |
| **p2L/m1L** | 0.40±0.015 | 0.24-0.43 | 12 | 0.35±0.027 | 0.23-0.40 | 7 | 0.01 |
| **p4L/m1L** | 0.58±0.01 | 0.48-0.61 | 12 | 0.57±0.007 | 0.54-0.59 | 9 | 0.07 |
| **m1W/m1L** | 0.38±0.003 | 0.36-0.40 | 12 | 0.39±0.004 | 0.36-0.40 | 9 | 0.10 |
| **m1tal/m1L** | 0.28±0.004 | 0.26-0.3 | 11 | 0.27±0.014 | 0.23-0.31 | 6 | 0.73 |

Table S5. Correlations of *C. a. alpinus* and *C. a. hesperius* tooth measurements and indices with coordinates of SZM and SHM descriptive models (E1–E4, K1–K3) and multiple *r*2 for multivariate linear regression model y = a1E1+a2E2...+... a6K2+a7K3 + a0 (backward stepwise method)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measurements** | **E1** | **E2** | **E3** | **E4** | **K1** | **K2** | **K3** | ***r*2** |
| **CL** | 0.90 | -0.01 | 0.12 | -0.21 | -0.07 | 0.52 | 0.72 | 0.89 |
| **CW** | 0.40 | 0.29 | -0.44 | 0.41 | -0.10 | -0.22 | 0.19 | 0.89 |
| **P4L** | 0.78 | 0.05 | -0.50 | -0.15 | -0.09 | -0.11 | 0.78 | 0.87 |
| **P4lp** | 0.63 | 0.58 | 0.08 | -0.27 | -0.53 | 0.50 | 0.35 | 0.74 |
| **P4W** | 0.54 | -0.74 | -0.32 | 0.02 | 0.62 | -0.06 | 0.70 | 0.93 |
| **cL** | 0.78 | 0.37 | 0.28 | 0.05 | -0.30 | 0.54 | 0.33 | 0.86 |
| **cW** | 0.30 | 0.55 | -0.27 | 0.53 | -0.65 | -0.26 | 0.15 | 0.87 |
| **p2L** | 0.49 | -0.77 | -0.11 | 0.06 | 0.63 | 0.08 | 0.65 | 0.82 |
| **p4L** | 0.85 | -0.29 | 0.20 | -0.15 | 0.27 | 0.55 | 0.62 | 0.88 |
| **m1L** | 0.81 | -0.13 | 0.32 | 0.12 | 0.08 | 0.50 | 0.50 | 0.78 |
| **m1tal** | 0.03 | -0.46 | 0.65 | 0.46 | 0.57 | 0.38 | -0.28 | 0.82 |
| **m1W** | 0.75 | 0.24 | 0.11 | 0.41 | -0.33 | 0.23 | 0.42 | 0.88 |
| **Index** | **E1** | **E2** | **E3** | **E4** | **K1** | **K2** | **K3** | ***r*2** |
| **CW/CL** | -0.17 | 0.27 | -0.48 | 0.50 | -0.05 | -0.51 | 0.26 | 0.75 |
| **P4W/P4L** | 0.40 | -0.82 | -0.22 | 0.06 | 0.70 | -0.03 | -0.56 | 0.88 |
| **P4lp/P4L** | -0.13 | 0.54 | 0.57 | -0.14 | -0.46 | 0.61 | 0.40 | 0.79 |
| **cW/cL** | -0.60 | -0.04 | -0.45 | 0.30 | -0.10 | -0.70 | 0.24 | 0.85 |
| **p2L/m1L** | 0.36 | -0.80 | -0.20 | 0.04 | 0.66 | -0.04 | -0.59 | 0.83 |
| **p4L/m1L** | 0.59 | -0.32 | 0.03 | -0.30 | 0.32 | 0.39 | -0.51 | 0.57 |
| **m1W/m1L** | -0.02 | 0.45 | -0.24 | 0.37 | -0.51 | -0.30 | 0.06 | 0.47 |
| **m1tal/m1L** | -0.35 | -0.40 | 0.50 | 0.41 | 0.53 | 0.15 | 0.51 | 0.72 |

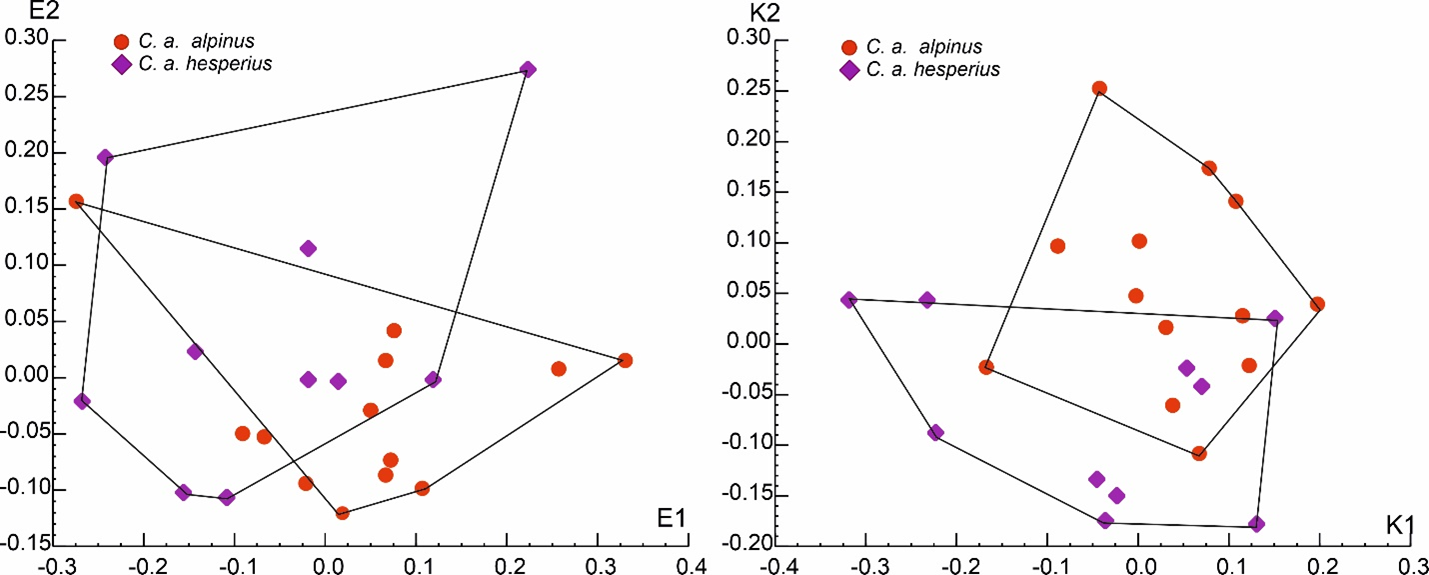


Fig. S1. Projections onto the first two coordinates of descriptive models of variation in tooth size (left) and shape (right) in *C. a. alpinus* and *C. a. hesperius*

Table S6. Correlations of recent C. alpinus ssp. except C. a. alpinus and C. a. hespericus, tooth measurements and indices with coordinates of SZM and SHM descriptive models (E1, E4, K1, K2) and multiple *r*2 for multivariate linear regression model y = a1E1+a2E2+a3K1+a4K2 + a0 (backward stepwise method)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Measurements** | **E1** | **E2** | **K1** | **K2** | ***r*2** |
| **P4L** | 0.94 | 0.09 | 0.20 | -0.42 | 0.89 |
| **P4W** | 0.48 | 0.80 | -0.02 | 0.00 | 0.84 |
| **p2L** | 0.10 | 0.42 | -0.19 | 0.13 | 0.18 |
| **p4L** | 0.84 | -0.33 | 0.41 | -0.33 | 0.91 |
| **m1L** | 0.87 | 0.19 | -0.22 | -0.20 | 0.76 |
| **m1W** | 0.84 | 0.08 | -0.32 | -0.70 | 0.94 |
| **Index** | **E1** | **E2** | **K1** | **K2** | ***r*2** |
| **P4W/P4L** | -0.25 | 0.81 | -0.19 | 0.35 | 0.72 |
| **p2L/m1L** | -0.27 | 0.33 | -0.10 | 0.21 | 0.19 |
| **p4L/m1L** | 0.19 | -0.59 | 0.72 | -0.21 | 0.65 |
| **m1W/m1L** | -0.08 | -0.16 | -0.11 | -0.65 | 0.42 |

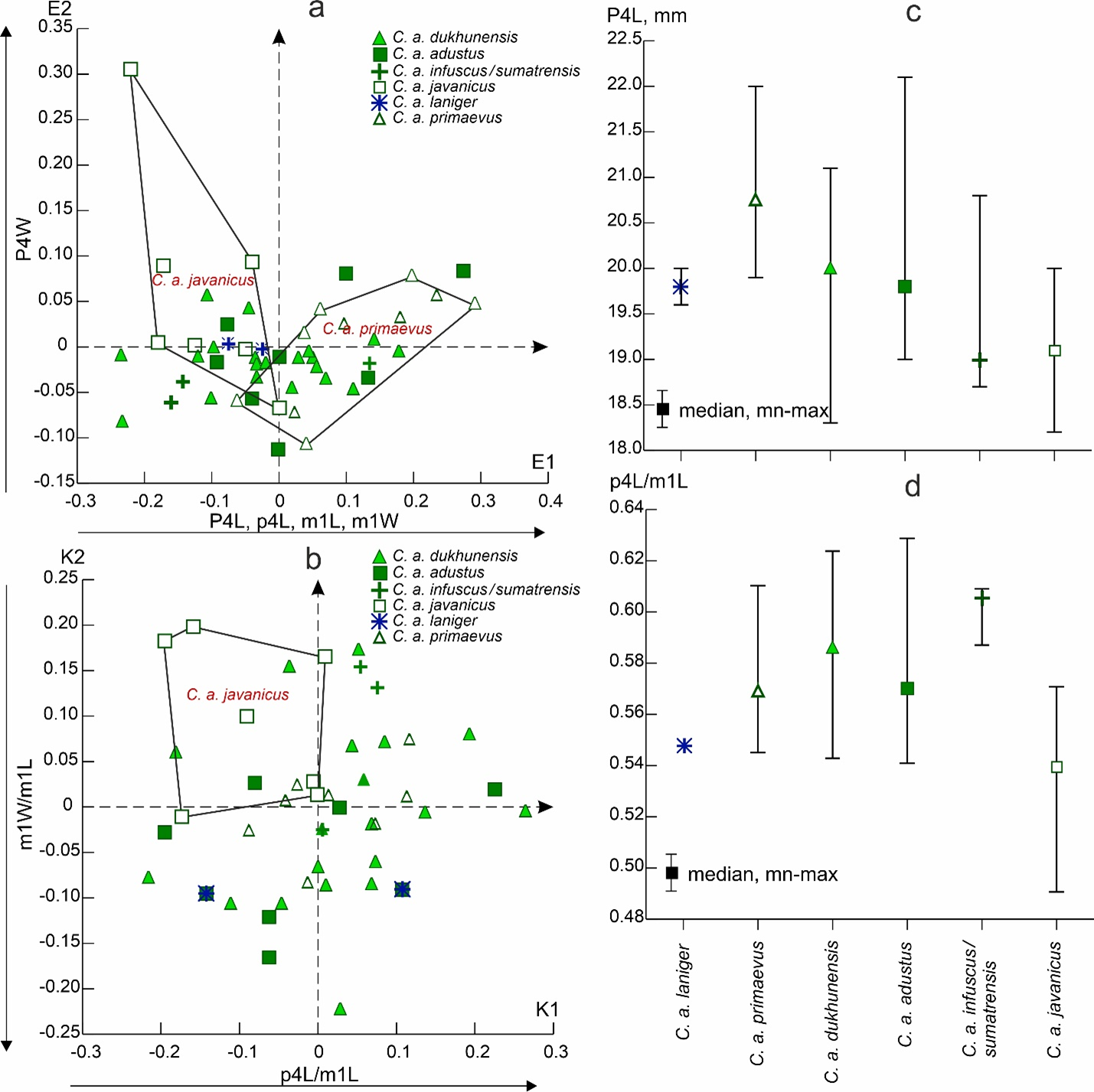


Fig. S2. Descriptive models of variability in tooth size (a) and shape (b) in the extant “southern” group of subspecies of *C. alpinus*; c, d - variability in P4L and p4L/m1L index.

**3. Supplementary Material for “Assessing the "quality" of the a priori taxonomic partitioning of the sample”**

Five different partitioning methods were used. Three of them were non-hierarchical clustering methods: K-means (KM) (Rencher 2002), medoid partition (PAM) (Kaufman, Rousseeuw 1990), and Gaussian finite mixture model (GFM) (Quiles et al. 2005). The next ones, UPGMA and McQuitty (weighted mean) methods (MQ), belonged to the hierarchical clustering methods (Charrad et al. 2014).

The "optimal number" of clusters for KM, UPGMA and MQ methods has been estimated in the R package **NbClust** (Charrad et al. 2014). The authors of the package use up to 30 tests (indices) to estimate this number. The decision is made according to the "majority rule" (MR), i.e. the maximum number of positive "tests" (for details see (Charrad et al. 2014)).

R package **fpc** (Henning 2010) was used to find the optimal partitioning of the sample using the PAM method. In this case, the estimation of the optimal partitioning was based on the "average silhouette width" criterion.

We applied GFM, implemented in the R **mclust** package (Scrucca et al., 2016),

**References**

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. (2014): Nbclust: An R package for determining the relevant number of clusters in a data set. – Journal of Statistical Software, 61(5): 1–36. <https://doi.org/10.18637/jss.v061.i06>

Henning Ch. (2010): Fpc: Flexible Procedures for Clustering. R Package.

<https://doi.org/10.1093/oxfordjournals.aob.a083391>

Kaufman, L., Rousseeuw, P. J. (1990): Partitioning Around Medoids (Program PAM). In:Finding groups in data: An introduction to cluster analysis. John Wiley & Sons, pp. 68–125. <https://doi.org/10.1002/9780470316801.ch2>

Quiles, J., Monchot, H., Pacher, M. (2005): Mixture analysis : application to cave bear sex-ratio determination. In: Proceedings of the 10th International Cave Bear Symposium, Mas d’Azil (Ariège, France), 29/09 – 03/10/2004, pp. 29–37.

Rencher, A. C. (2002): Methods of Multivariate Analysis (Second Ed.). New York, USA, Wiley-Intersience, John Wiley & Sons, Inc., 1–708 pp.

Scrucca, L., Fop, M., Murphy, T. B., Raftery, A. E. (2016): Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. – R Journal, 8(1): 289–317. <https://doi.org/10.32614/rj-2016-021>

**List of libraries of R packages used:**

> library(cluster)

> library(NbClust)

> library(mclust)

> library(fpc)

> library(factoextra)

> library(ggplot2)

**Abbreviation:**

**X** – data frame (raw variable(-s)

**R code:**

**##Package for partitioning around medoids with estimation of number of clusters (library(fpc))**

> set.seed(123)

> clust<-pamk(**X**, krange=2:10, criterion="asw",critout=TRUE)

> pamk(**X**, krange=2:5, criterion="asw",critout=TRUE)

> pamx<-pam(**X**, clust$nc, metric = "euclidean")

> pamx[["clusinfo"]]

> plot(pamx)

> class<-clust[["pamobject"]][["clustering"]]

> write.csv(class, file = "class.txt")

**##NbClust - Package for determining the best number of clusters** (**library(NbClust))**

**List of indices and criteria** see in (Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. (2014): NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. – Journal of Statistical Software, 61(6): 1–36. <https://doi.org/10.18637/jss.v061.i06>)

> set.seed(123)

**##K-means method**

> NbClust<- NbClust(**X**, diss=NULL, distance = "euclidean", min.nc=2, max.nc=10, method = "kmeans", index = "alllong", alphaBeale = 0.1)

> NbClust

> hist(NbClust$Best.nc[1,], breaks = max(na.omit(NbClust $Best.nc[1,])))

> class<-NbClust[["Best.partition"]]

> write.csv(class, file = "class.txt")

**##UPGMA method**

> NbClust<- NbClust(**X**, diss=NULL, distance = "euclidean", min.nc=2, max.nc=10, method = "average", index = "alllong", alphaBeale = 0.1

> NbClust

> hist(NbClust$Best.nc[1,], breaks = max(na.omit(NbClust $Best.nc[1,])))

> class<-NbClust[["Best.partition"]]

> write.csv(class, file = "class.txt")

**##McQuitty method**

> NbClust<- NbClust(**X**, diss=NULL, distance = "euclidean", min.nc=2, max.nc=10, method = "mcquitty", index = "all", alphaBeale = 0.1)

> NbClust

> hist(NbClust$Best.nc[1,], breaks = max(na.omit(NbClust $Best.nc[1,])))

> class<-NbClust[["Best.partition"]]

> write.csv(class, file = "class.txt")

**##Gaussian finite mixture model fitted by EM algorithm (library(Mclust))**

> set.seed(123)

> BIC<-mclustBIC(**X**)

> summary(BIC)

> plot(BIC)

> mod<-Mclust(**X**, modelName="**abbreviation of the model**", G= **number of groups**)

> summary(mod, parameters = TRUE)

> plot(mod)

> class<-mod[["classification"]]

> write.csv(class, file = "class.txt")

**Protocol**

**1) Upper canine**

Variables CL, CW

**Partitioning around medoids with estimation of number of clusters**

$pamobject

Medoids:

ID CL CW

[1,] 10 10.0 5.70

[2,] 24 10.9 6.85

Clustering vector:

Objective function:

build swap

0.6298548 0.5719021

$nc – **number of clusters**

[1] 2

$crit

[1] 0.0000000 0.5157607 0.4209089 0.4224442 0.4313289 0.4011278 0.4229869

[8] 0.4018960 0.4436365 0.4281104





**Determining the best number of clusters**

**K-means method**

\* Among all indices:

\* 10 proposed 2 as the best number of clusters

\* 3 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

\* 5 proposed 5 as the best number of clusters

\* 2 proposed 8 as the best number of clusters

\* 7 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**UPGMA method**

\* Among all indices:

\* 10 proposed 2 as the best number of clusters

\* 4 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\* 4 proposed 6 as the best number of clusters

\* 2 proposed 7 as the best number of clusters

\* 1 proposed 8 as the best number of clusters

\* 4 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**McQuitty method**

\* Among all indices:

\* 8 proposed 2 as the best number of clusters

\* 4 proposed 3 as the best number of clusters

\* 2 proposed 4 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\* 2 proposed 6 as the best number of clusters

\* 1 proposed 7 as the best number of clusters

\* 3 proposed 9 as the best number of clusters

\* 2 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**Gaussian finite mixture model fitted by EM algorithm**

Best BIC values:

EEE,1 EEV,1 EVE,1

BIC -138.2029 -138.2029 -138.2029

BIC diff 0.0000 0.0000 0.0000



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

----------------------------------------------------

Mclust EII (spherical, equal volume) model with 2 components:

log-likelihood n df BIC ICL

-61.17989 32 6 -143.1542 -146.2262

Clustering table:

1 2

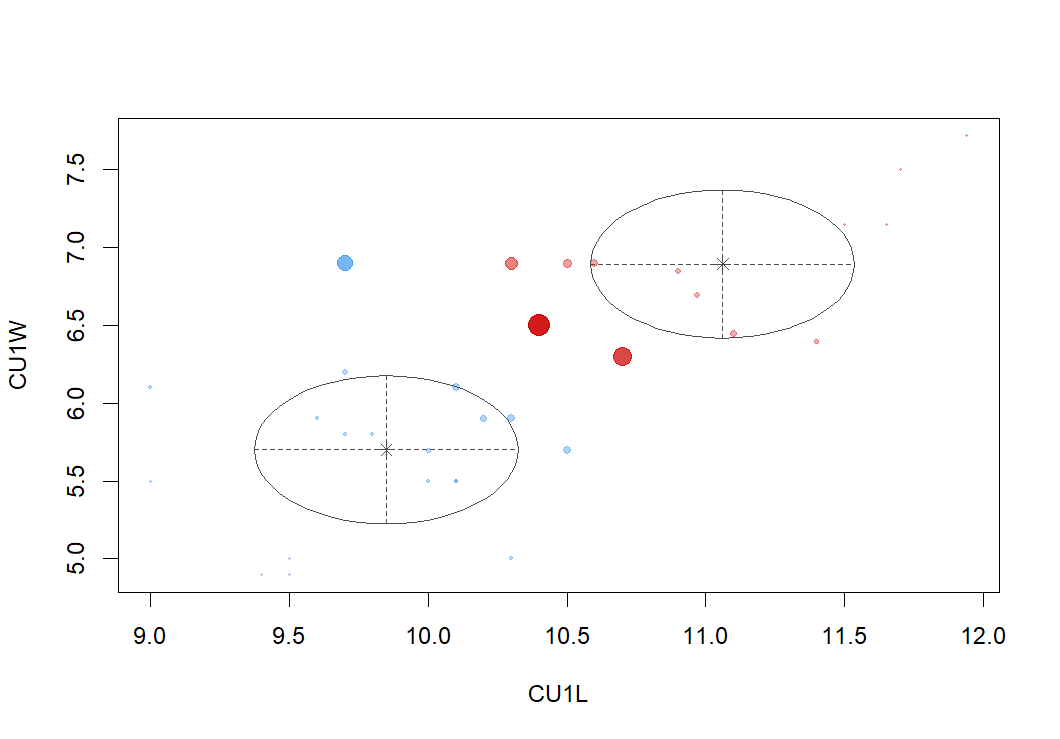
19 13

Mixing probabilities:

1 2

0.6133183 0.3866817





**2) Upper premolar P4**

Variables P4L, P4W

**Partitioning around medoids with estimation of number of clusters**

$pamobject

Medoids:

ID PU4Lw PU4Ww

[1,] 75 21.1 10.5

[2,] 24 20.0 7.3

Objective function:

build swap

1.0758296 0.9775286

$nc

[1] 2

$crit

[1] 0.0000000 0.5671092 0.4642028 0.4434530 0.4777443 0.4481494 0.4548525

[8] 0.4114896 0.3988272 0.4160850





**Determining the best number of clusters**

**K-means method**

\* Among all indices:

\* 11 proposed 2 as the best number of clusters

\* 4 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

\* 6 proposed 5 as the best number of clusters

\* 2 proposed 7 as the best number of clusters

\* 3 proposed 8 as the best number of clusters

\* 1 proposed 9 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**UPGMA method**

\* Among all indices:

\* 10 proposed 2 as the best number of clusters

\* 3 proposed 3 as the best number of clusters

\* 3 proposed 4 as the best number of clusters

\* 2 proposed 5 as the best number of clusters

\* 1 proposed 6 as the best number of clusters

\* 3 proposed 7 as the best number of clusters

\* 5 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**McQuitty method**

\* Among all indices:

\* 10 proposed 2 as the best number of clusters

\* 1 proposed 3 as the best number of clusters

\* 7 proposed 4 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\* 1 proposed 6 as the best number of clusters

\* 1 proposed 7 as the best number of clusters

\* 3 proposed 8 as the best number of clusters

\* 1 proposed 9 as the best number of clusters

\* 3 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**Gaussian finite mixture model fitted by EM algorithm**

Best BIC values:

VEV,2 VEV,3 VEE,2

BIC -561.5747 -564.796384 -564.80128

BIC diff 0.0000 -3.221733 -3.22663



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

---------------------------------------------------

Mclust VEV (ellipsoidal, equal shape) model with 2 components:

log-likelihood n df BIC ICL

-257.8625 98 10 -561.5747 -562.8471

Clustering table:

1 2

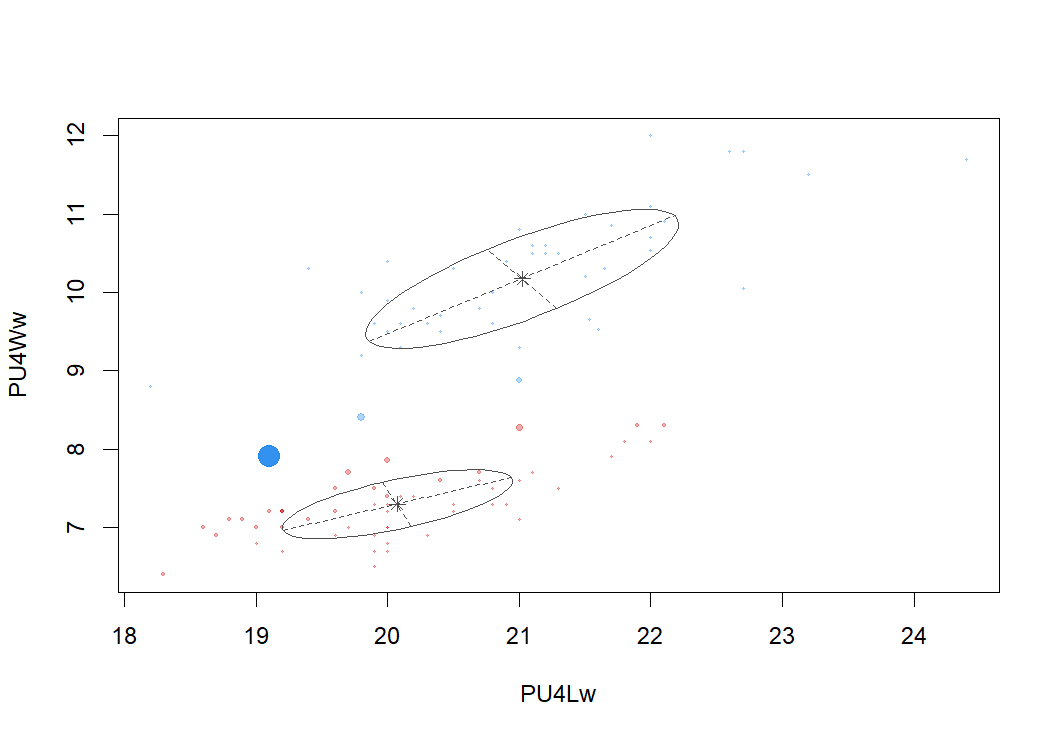
45 53

Mixing probabilities:

1 2

0.4550426 0.5449574





**3) Lower canine c**

Variables cL, cW

**Partitioning around medoids with estimation of number of clusters**

$pamobject

Medoids:

ID CL1L CL1W

[1,] 17 9.70 6.70

[2,] 29 10.70 7.00

[3,] 36 11.94 8.18

Objective function:

build swap

0.5213742 0.4942002

$nc

[1] 3

$crit

[1] 0.0000000 0.4514590 0.4572436 0.4149121 0.3950609 0.4121066 0.4379410

[8] 0.4055146 0.4196943 0.4404196





**Determining the best number of clusters**

**K-means method**

\* Among all indices:

\* 6 proposed 2 as the best number of clusters

\* 10 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

\* 2 proposed 7 as the best number of clusters

\* 1 proposed 8 as the best number of clusters

\* 5 proposed 9 as the best number of clusters

\* 2 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 3



**UPGMA method**

\* Among all indices:

\* Among all indices:

\* 11 proposed 2 as the best number of clusters

\* 9 proposed 3 as the best number of clusters

\* 1 proposed 7 as the best number of clusters

\* 3 proposed 8 as the best number of clusters

\* 4 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**McQuitty method**

\* Among all indices:

\* 11 proposed 2 as the best number of clusters

\* 8 proposed 3 as the best number of clusters

\* 4 proposed 7 as the best number of clusters

\* 1 proposed 8 as the best number of clusters

\* 4 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**Gaussian finite mixture model fitted by EM algorithm**

Best BIC values:

EEE,2 EVE,2 EEE,1

BIC -158.8845 -160.833499 -160.918790

BIC diff 0.0000 -1.948976 -2.034268



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

----------------------------------------------------

Mclust EII (spherical, equal volume) model with 3 components:

log-likelihood n df BIC ICL

-65.65658 37 9 -163.8114 -170.5912

Clustering table:

1 2 3

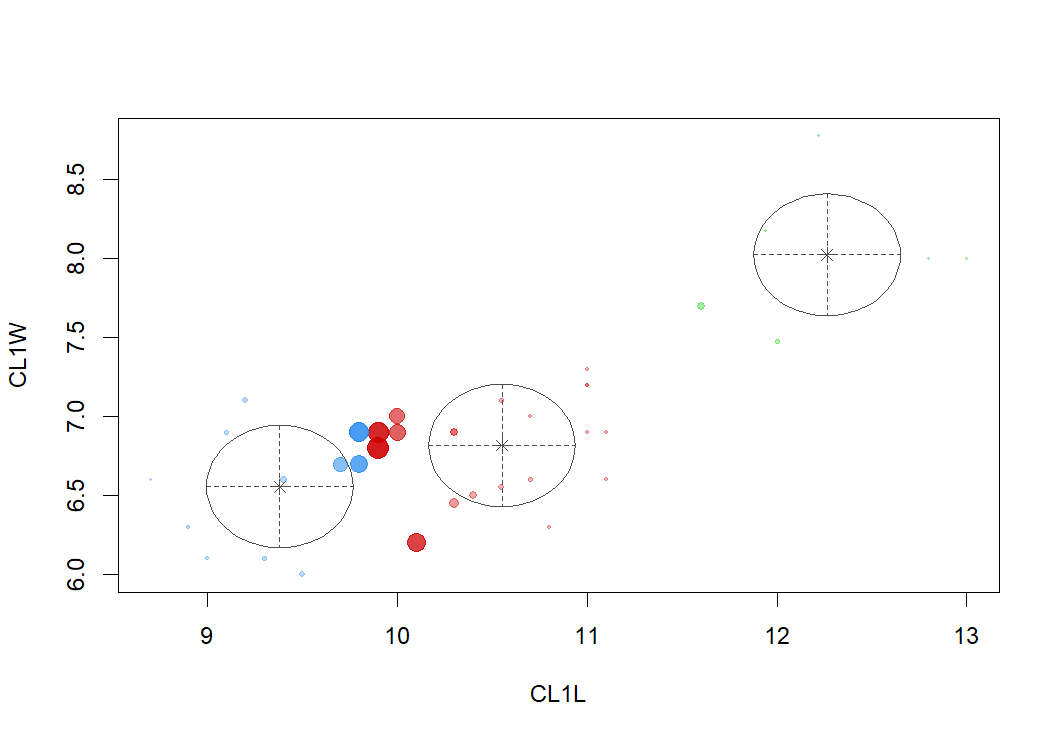
11 20 6

Mixing probabilities:

1 2 3

0.3218778 0.5169125 0.1612097





**4) Lower premolar p2**

Variable p2L

**Partitioning around medoids with estimation of number of clusters**

$pamobject

Medoids:

ID p2L

[1,] 14 9.3

[2,] 19 4.9

Objective function:

build swap

0.5437079 0.3751685

$nc

[1] 2

$crit

[1] 0.0000000 0.8643397 0.7097927 0.7171625 0.5289211





**Determining the best number of clusters**

**K-means method**

\* Among all indices:

\* 3 proposed 2 as the best number of clusters

\* 2 proposed 3 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2

**UPGMA method**

\* Among all indices:

\* 3 proposed 2 as the best number of clusters

\* 2 proposed 3 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2

**McQuitty method**

\* Among all indices:

\* 3 proposed 2 as the best number of clusters

\* 1 proposed 3 as the best number of clusters

\* 2 proposed 5 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2

**Gaussian finite mixture model fitted by EM algorithm**

Best BIC values:

V,2 E,2 V,3

BIC -271.256 -276.242599 -281.80141

BIC diff 0.000 -4.986551 -10.5453



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

----------------------------------------------------

Mclust V (univariate, unequal variance) model with 2 components:

log-likelihood n df BIC ICL

-124.4064 89 5 -271.256 -271.2561

Clustering table:

1 2

47 42

Mixing probabilities:

1 2

0.5280895 0.4719105





**5) Lower premolar p4**

Variable p4L

**Partitioning around medoids with estimation of number of clusters**

$pamobject

Medoids:

ID p4L

[1,] 60 12.4

[2,] 88 13.7

[3,] 78 14.8

[4,] 107 13.0

[5,] 12 10.7

[6,] 55 11.9

[7,] 38 11.5

Objective function:

build swap

0.1538318 0.1471028

$nc

[1] 7

$crit

[1] 0.0000000 0.5728952 0.5630908 0.5480242 0.5740187 0.5725034 0.6192097

[8] 0.5869608 0.5863207 0.6001285





**Determining the best number of clusters**

**K-means method**

\* Among all indices:

\* 1 proposed 2 as the best number of clusters

\* 1 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

\* 2 proposed 9 as the best number of clusters

\* 1 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 9

**UPGMA method**

\* Among all indices:

\* 1 proposed 2 as the best number of clusters

\* 1 proposed 3 as the best number of clusters

\* 3 proposed 5 as the best number of clusters

\* 1 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 5

**McQuitty method**

\* Among all indices:

\* 1 proposed 2 as the best number of clusters

\* 1 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\* 1 proposed 8 as the best number of clusters

\* 1 proposed 9 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2

**Gaussian finite mixture model fitted by EM algorithm**

Best BIC values:

E,1 V,1 V,2

BIC -348.9989 -348.9989 -355.219256

BIC diff 0.0000 0.0000 -6.220348



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

----------------------------------------------------

Mclust V (univariate, unequal variance) model with 2 components:

log-likelihood n df BIC ICL

-165.9276 107 5 -355.2193 -368.092

Clustering table:

1 2

92 15

Mixing probabilities:

1 2

0.8788096 0.1211904



**5) Lower molar m1**

Variables m1L, m1W

**Partitioning around medoids with estimation of number of clusters**

$pamobject

Medoids:

ID m1L m1W

[1,] 21 21.0 8.3

[2,] 53 23.2 9.1

Objective function:

build swap

0.9469766 0.8012431

$nc

[1] 2

$crit

[1] 0.0000000 0.5086987 0.4728095 0.4151682 0.4242009 0.4398191 0.4559570

[8] 0.4250771 0.3865862 0.3666331





**Determining the best number of clusters**

**K-means method**

\* Among all indices:

\* 9 proposed 2 as the best number of clusters

\* 10 proposed 3 as the best number of clusters

\* 2 proposed 5 as the best number of clusters

\* 1 proposed 8 as the best number of clusters

\* 3 proposed 9 as the best number of clusters

\* 2 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 3



**UPGMA method**

\* Among all indices:

\* 9 proposed 2 as the best number of clusters

\* 11 proposed 3 as the best number of clusters

\* 2 proposed 5 as the best number of clusters

\* 4 proposed 9 as the best number of clusters

\* 2 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 3



**McQuitty method**

\* Among all indices:

\* 9 proposed 2 as the best number of clusters

\* 5 proposed 3 as the best number of clusters

\* 8 proposed 4 as the best number of clusters

\* 1 proposed 5 as the best number of clusters

\* 1 proposed 8 as the best number of clusters

\* 3 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2



**Gaussian finite mixture model fitted by EM algorithm**

Best BIC values:

EEV,2 EVE,2 EVV,2

BIC -644.738 -645.0485407 -646.398433

BIC diff 0.000 -0.3105217 -1.660414



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

----------------------------------------------------

Mclust EEV (ellipsoidal, equal volume and shape) model with 2 components:

log-likelihood n df BIC ICL

-300.3287 134 9 -644.738 -661.8349

Clustering table:

1 2

117 17

Mixing probabilities:

1 2

0.849077 0.150923



----------------------------------------------------

Gaussian finite mixture model fitted by EM algorithm

----------------------------------------------------

Mclust VEI (diagonal, equal shape) model with 3 components:

log-likelihood n df BIC ICL

-315.4412 134 12 -689.6566 -725.0598

Clustering table:

1 2 3

54 39 41

Mixing probabilities:

1 2 3

0.4013872 0.2883765 0.3102362

