Elsevier Editorial System(tm) for Geobios
Manuscript Draft

Manuscript Number: GEOBIO-D-10-00087R2

Title: A large new collection of Palaeostylops from the Paleocene of the Flaming Cliffs area (Ulan-Nur Basin, Gobi Desert, Mongolia), and an evaluation of the phylogenetic affinities of Arctostylopidae (Mammalia, Gliriformes)

Article Type: Original article / Article original

Keywords: Arctostylopidae; Gliriformes; Flaming Cliffs; Morphology; Biometry; Species discrimination; Phylogeny

Corresponding Author: Dr Pieter Missiaen, Ph.D.

Corresponding Author's Institution: Ghent University

First Author: Pieter Missiaen, Ph.D.

Order of Authors: Pieter Missiaen, Ph.D.; Gilles Escarguel, Ph.D.; Jean-Louis Hartenberger , Ph.D.; Thierry Smith, Ph.D.

Abstract: Arctostylopids are enigmatic mammals known from the Paleocene and early Eocene of Asia and North America. Based on molar similarities, they have most often been grouped with the extinct Notoungulata from South and Central America, but tarsal evidence links them to Asian basal gliriforms. Although Palaeostylops is the best known arctostylopid genus, some points of its content and species level taxonomy are uncertain. Here we report 255 upper and lower jaw fragments of Palaeostylops, five calcanea, three astragali, as well as the first known arctostylopid distal tibia. This new material was collected from the late Paleocene of the Flaming Cliffs area in Mongolia, in a single lens almost exclusively containing arctostylopid remains. Our study of the morphology and size of the new Palaeostylops dental material confirms the validity of two species, P. iturus and P. macrodon, and illustrates their morphological and biometrical variability and diagnostic differences. The distal tibia of Palaeostylops is relatively unspecialised and resembles the Asian gliriforms Pseudictops and Rhombomylus. We also review the relevance of the historically important genus Palaeostylops in view of other, more recently described but less abundant arctostylopid genera. Palaeostylops remains the reference taxon for the arctostylopid anterior dentition and postcranial morphology. For both anatomical regions, arctostylopids differ significantly from notoungulates, and present a mosaic of characters also seen in basal gliriforms. The notoungulate-like molars of Palaeostylops are highly specialized for arctostylopids and the arctostylopid molar morphotype is therefore better illustrated by the early middle Paleocene Asiostylops. This morphotype does not present any similarities to notoungulates, but shares a number of derived characters with basal gliriforms. Among gliriforms, the primitive arctostylopid morphotype is most similar to Astigale from the early Paleocene of South China, and we suggest that Arctostylopidae may therefore be more closely related to Astigalidae than to any other group.

A large new collection of *Palaeostylops* from the Paleocene of the Flaming Cliffs area (Ulan-Nur Basin, Gobi Desert, Mongolia), and an evaluation of the phylogenetic affinities of Arctostylopidae (Mammalia, Gliriformes)

Pieter Missiaen [a,b]*, Gilles Escarguel [c], Jean-Louis Hartenberger [d], Thierry Smith [b]

[a] Research Unit Palaeontology, Ghent University, Krijgslaan 281-S8, B-9000 Ghent, Belgium

[b] Royal Belgian Institute of Natural Sciences, Rue Vautier 29, B-1000 Brussels, Belgium

[c] Laboratoire de Géologie de Lyon : Terre, Planètes, Environnement, UMR CNRS 5276, Université Lyon 1, Boulevard du 11 Novembre 1918, F-69622 Villeurbanne cedex, France

[d] Institut des Science de l'Evolution, UMR CNRS 5554, Université Montpellier 2, Place E. Bataillon, CC 064, 34095 Montpellier cedex 5, France

* Corresponding author. E-mail address: Pieter.Missiaen@ugent.be

Abstract

Arctostylopids are enigmatic mammals known from the Paleocene and early Eocene of Asia and North America. Based on molar similarities, they have most often been grouped with the extinct Notoungulata from South and Central America, but tarsal evidence links them to Asian basal gliriforms. Although *Palaeostylops* is the best known arctostylopid genus, some points of its content and species level taxonomy are uncertain. Here we report 255 upper and lower jaw fragments of *Palaeostylops*, five calcanea, three astragali, as well as the first known arctostylopid distal tibia. This new material was collected from the late Paleocene of the Flaming Cliffs area in Mongolia, in a single lens almost exclusively containing arctostylopid remains. Our study of the morphology and size of the new *Palaeostylops* dental material confirms the validity of two species, *P. iturus* and *P. macrodon*, and illustrates their morphological and biometrical variability and diagnostic differences. The distal tibia of *Palaeostylops* is relatively unspecialised and resembles the Asian gliriforms *Pseudictops* and *Rhombomylus*. We also review the relevance of the historically important genus *Palaeostylops* in view of other, more recently described but less abundant arctostylopid genera. *Palaeostylops* remains the reference taxon for the arctostylopid anterior dentition and postcranial morphology. For both anatomical regions, arctostylopids differ significantly from notoungulates, and present a mosaic of characters also seen in basal gliriforms. The notoungulate-like molars of *Palaeostylops* are highly specialized for arctostylopids and the arctostylopid molar morphotype is therefore better illustrated by the early middle Paleocene *Asiostylops*. This morphotype does not present any similarities to notoungulates, but shares a number of derived characters with basal gliriforms. Among gliriforms, the primitive arctostylopid morphotype is most similar to *Astigale* from the early Paleocene of South China, and we suggest that Arctostylopidae may therefore be more closely related to Astigalidae than to any other group.

Keywords: Arctostylopidae; Gliriformes; Flaming Cliffs; Morphology; Biometry; Species discrimination; Phylogeny

In Memoriam Demberel Dashzeveg, 1936 - 2010.

1. Introduction

The late Paleocene and early Eocene Arctostylopidae are diverse and typical elements of Asian mammal faunas, as well as exceedingly rare elements in North American faunas where they were first discovered (Cifelli et al 1989; Wang et al., 2007). Based on striking molar resemblances, arctostylopids were initially grouped with the South and Central American Notoungulata (Matthew, 1915). This grouping implies early Tertiary mammal dispersal between North and South America. Arctostylopids have therefore figured prominently in various intercontinental dispersal hypotheses (Patterson and Pascual, 1972; Cifelli, 1983; Gingerich, 1985).

In 1989, interest in arctostylopids was revived by a phylogenetic revision of the group by Cifelli et al (1989). This paper featured the first arctostylopid tarsal bones, and based on dental and tarsal morphology, Cifelli et al. (1989) placed Arctostylopidae in a new order Arctostylopida, distinct from Notoungulata and all other mammals. Thereby they also dismissed the faunal exchange between North and South America during the late Paleocene or early Eocene. Another part of their study handled the classification of the two best known arctostylopids, *Palaeostylops iturus* and *P. macrodon*. Since their discovery, these two species had always been reported to co-occur, both in Mongolian and Chinese late Paleocene sites (Matthew and Granger, 1925; Matthew et al., 1929; Russell and Zhai, 1987; Meng et al., 1998; but see Missiaen and Smith, 2008). This co-occurrence in otherwise species-poor communities, of two taxa differing only by their size seemed to suggest the presence of a single, sexually dimorphic species. Cifelli and co-authors raised the possibility of sexual dimporphism, but quickly dismissed it. Presenting a number of novel morphological differences between both forms, they concluded that they represented two distinct species and genera: *Palaeostylops iturus* and "*Gashatostylops" macrodon* (Cifelli et al.,1989).

Since then, the validity of a separate genus "*Gashatostylops"* has been accepted by some studies (Ting, 1998; Meng et al., 1998), and rejected by others (Kondrashov and Lucas, 2004; Ni et al., 2007; Missiaen and Smith, 2008; this paper). Similarly, some researchers have accepted the placement of Arctostylopidae in a separate order Arctostylopida (Ting, 1998; Zack, 2004; Wang et al., 2008), while others have suggested to group them with Notoungulata based on unpublished new material (Bloch, 1999) or a rebuttal of the arguments of Cifelli and co-workers (Kondrashov and Lucas, 2004). Missiaen et al. (2006) published additional arctostylopid tarsals, assigned to *Palaeostylops iturus* from Inner Mongolia and *Arctostylops* from North America. Based on the tarsal evidence, they supported the exclusion of Arctostylopidae from Notoungulata, and moreover placed the family Arctostylopidae within the superorder Gliriformes.

Here we report on the discovery of 255 upper and lower jaw fragments of arctostylopids recovered from a small sandy lens in the late Paleocene of the Flaming Cliffs area in Mongolia (Fig. 1). This collection contains specimens referable to both *Palaeostylops iturus* and *P. macrodon* based on dental morphology and measurements, and represents a large, single sample from the type area of both forms. This collection is therefore perfectly suited to study the morphological and size variability of both forms, and to assess whether they represent two genera, two species or even one sexually dimorphic species.

In addition to the abundant dental remains, this lens also yielded a limited number of postcranial elements, including the previously unknown arctostylopid distal tibia, which provides additional data for reconstructing the higher-level phylogenetic position of arctostylopids.

In view of more recently described but less well known arctostylopids and of the new hypotheses on arctostylopid evolution, we critically review the relevance of the historically

important and abundant *Palaeostylops* fossils from Gashato for our understanding of arctostylopid evolution

2. Material and methods

Abbreviations: AMNH, American Museum of Natural History, New York, USA; IVPP, Institute of Vertebrate Paleontology and Paleoanthropology, Beijing, China; IMM, Inner Mongolian Museum, Hohhot, China; MCZ, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA; MLP, Facultad de Ciencias Naturales y Museo de La Plata, División Paleontología de Vertebrados, Buenos Aires, Argentina; MPC-M Mongolian Paleontological Center-Mammal Collection, Academy of Sciences of Mongolia, Ulaanbaatar, Mongolia.

2.1. Material

The famous Flaming Cliffs area in Mongolia has long been known to yield abundant late Cretaceous (Djadokhta Formation) and early Paleogene (Khashat Formation = Gashato Formation) vertebrates (Matthew and Granger 1925). During fieldwork at the Gashato locality in 1999, a small fossiliferous sandy lens (<1 m^3) in the late Paleocene Member 1 of the Khashat Formation was discovered and completely excavated and screenwashed by two of us (G.E. and J.-L.H.) and the late D. Dashzeveg (Fig. 1). Except for a few rare teeth of a large mixodont and one dentary fragment of a sarcodontid, the fossil mammal teeth in this lens belong exclusively to arctostylopids. A total of 730 teeth (canines, premolars and molars) in 255 upper and lower jaw fragments were recorded, representing a minimum number of individuals (MNI) of 48. In addition, this lens yielded a very limited number of identifiable postcranial remains. These include 5 calcanea and 3 astragali, which are attributed to *Palaeostylops* based on their abundance and their close similarity to the previously published

tarsals of *P. iturus* (Missiaen et al., 2006). The distal part of a left tibia is also attributed to *Palaeostylops*, based on its articulation with the *Palaeostylops* tarsals. The distal part of a humerus can be attributed to the typical late Paleocene multituberculate *Lambdopsalis bulla*, which is not represented by dental specimens in this collection. Finally, a large phalanx and the proximal part of a femur could not be identified with certainty.

Among the arctostylopid dental remains, two different morphotypes can be recognised, corresponding to *P. iturus* and *P. macrodon* as originally described from this area (Matthew and Granger, 1925; Matthew et al., 1929). When possible, dental remains were attributed to either of both morphs based on the enlarged M2/m2, the only criterion universally accepted as diagnostic between both forms (Matthew et al., 1929; Cifelli et al. 1989; Kondrashov and Lucas, 2004; Missiaen and Smith, 2008). Using this method, 111 of the 255 upper and lower jaw fragments, representing 376/730 teeth and a MNI of 32, were unambiguously identified as *P. iturus*, whereas 40/255 jaw fragments, representing 154/730 teeth and a MNI of 14, were identified as *P. macrodon*.

2.2 Biostatistical analyses

Parallel to the comparative analysis of cheek tooth morphologies, all arctostylopid teeth in this collection were measured using a binocular microscope with a graded eyepiece with a precision of 0.1 mm. Length and/or width was determined for 697/730 measurable teeth, of which 497 were unambiguously identified as either *P. iturus* or *P. macrodon*. In order to quantitatively describe and compare length and width measurements for each available cheek tooth position, we computed standard statistics using PAST v. 2.01 (Hammer et al., 2001), including: (i) usual univariate descriptive statistics, (ii) bivariate (Doornik and Hansen omnibus) tests for normality, (iii) Kolmogorov-Smirnov nonparametric test for two-sample univariate distribution comparison, and (iv) Wilks' λ test for multigroup multivariate

comparison (here, 2 groups [*P. iturus* and *P. macrodon*] and 2 variables [length and width cheek tooth dimensions]). Computation of two-group bivariate Wilks' λ (formally identical to an Hotelling's T^2-test) was preferred to the more usual combination of two univariate Student *t*-tests because several univariate distributions show significant departure from normality (results not shown), whereas all but one cheek tooth positions (*P. iturus*' P4) appear bi-normally distributed at the 95% confidence level (Table 1). Thus, based on the available sample distributions, the association of bivariate Wilks' λ with univariate Kolmogorov-Smirnov statistics offers the best possible compromise between power and robustness in order to test for significance both sample mean and individual distribution differences for each position.

We further investigated proportional differences in the lengths and widths of upper and lower cheek teeth between *P. iturus* and *P. macrodon* through:

- the computation of Simpson's (1941; see Simpson et al., 1960) Log-ratios, using *P. iturus* as the reference sample. A two-step Monte Carlo procedure (parametric bootstrap) allowed us to estimate: (i) the confidence intervals around the observed Log-Ratio values for *P. macrodon*, and (ii) the expected distributions of Log-Ratio values under the null hypothesis that *P. iturus* and *P. macrodon* share the same tooth dimensions (see Appendix A for computational details);
- the construction of bivariate scatterplots linking first and second upper or lower molars areas (estimated by a simple length × width product) coupled with one-way analysis of covariance (ANCOVA; Sokal & Rohlf, 1995) in order to test for equality of the second molar size when adjusted for covariance with the first molar size (regarded as a first-order proxy of overall dental size). Determination of the optimal M2/m2 surface cut-off value between *P. iturus* and *P. macrodon* follows Favre et al.'s (2008) method for determining the critical value ζ for

which the joint prediction error-risk of incorrectly attributing any specimen to one of the two groups is minimal.

Finally, in order to better characterize the taxonomic status of the studied fossil assemblage, we computed various complementary metrics focusing on distinct aspects of the sample distributions of three dental measurements: length (L), width (W) and ln(L × W) of the P3/p3 to M3/m3 of all measured *Palaeostylops* teeth, and measured teeth a priori assigned to *P. iturus* or to *P. macrodon* (see Appendix B for computational details). Two metrics, the unbiased coefficient of variation (*V**; Sokal & Rohlf, 1995) and the bimodality index (*b*; Der & Everitt, 2002), focus on the relative variability and shape of the sample distributions. Two other techniques, the dimorphism ratio of the "method-of-moments (MoM)" technique (*D*; Josephson et al., 1996) and maximum likelihood mixture analysis coupled with an evidence ratio-based model selection procedure (Titterington et al., 1985; Burnham & Anderson, 2002; Johnson & Omland, 2004), aim to estimate the dimorphism ratio involved by the available data. Ratios estimated for the "all-*Palaeostylops*" samples were compared to the expected ratios directly calculated from measured teeth a priori assigned to *P. iturus* or *P. macrodon*.

3. Dental morphology, size variation and species discrimination of the Flaming Cliffs arctostylopids

In the original description, *P. macrodon* was diagnosed as follows: "Cheek tooth series about 20 per cent longer than in *P. iturus*, molars relatively narrower, M2/m2 larger relative to other teeth" (Matthew et al., 1929: p.11). In their revision of arctostylopids, Cifelli and co-authors thought the difference between both forms was important enough to deserve a distinction at the genus level, and noted in their diagnosis of "*Gashatostylops*" *macrodon*: "… differing from *Palaeostylops*, the most closely similar genus, in having relatively enlarged upper and lower second molars; in having cuspules, variable in number and development, on the lingual

cingula of the upper molars; in the weakness or absence of a sulcus separating the lingual cusps of M1; in the presence of two rather than three upper incisors; and in having a laterally constricted snout, with the dental arcade multiply curved" (Cifelli et al., 1989: p. 15).

In the following sections we will evaluate how these differences apply to the arctostylopids reported here from the Flaming Cliffs area, and what the implications are for the taxonomic status of both forms.

3.1. Morphological variability

Although the upper molars assigned to *P. macrodon* generally have stronger lingual cuspules (Fig. 2(7, 9)) than those assigned to *P. iturus* (Fig. 2(5, 6)), these cuspules are sometimes rather well developed in *P. iturus* (Fig. 2(1, 3)), and moderately weak in *P. macrodon* (Fig. 2(8, 11)). Similarly, although many specimens of *P. macrodon* have only a weak lingual sulcus on M1 (Fig. 2(7, 9-11)) and many *P. iturus* M1s have a better developed sulcus (Fig. 1(2, 5)), there is also an important variation and overlap for this character with a marked sulcus in some specimens of *P. macrodon* (Fig. 2(8)) and only a weak one in some *P. iturus* specimens (Fig. 2(4, 6)).

Based on an uncatalogued IVPP specimen attributed to *P. macrodon*, Cifelli et al. (1989) concluded that this form had only two incisors and a multiply curved dental arcade. However, the specimen concerned is damaged in front of the root of I2, and therefore does not adequately establish the presence of only two incisors in *P. macrodon*. The apparent constriction of the snout and the curved dental arcade in this specimen may be the result of a break in the maxillary at the level of P3. In all of the *P. macrodon* specimens in our collection for which this region is present, the dental arcade is straight and the snout is not constricted (Fig. 2(8, 10)), exactly similar to the shape of the dental arcade in *P. iturus* (Fig. 2(2, 5, 6)).

3.2. Biometric variability

Tooth measurements confirm that on average *P. macrodon* is indeed larger than *P. iturus* (Table 1; Suppl. Fig. S1), as further evidenced by inter-sample comparisons of univariate (length or width) distributions and bivariate (length × width) means (Table 2). While the Wilks' λ tests unambiguously support differences between *P. iturus* and *P. macrodon* length × width means for each analysed cheek tooth position, a strong individual size overlap appears between the two forms, especially for premolars, leading to non-significant Kolmogorov-Smirnov test results in those cases. Only molars, and most particularly M2 and m2 allow unambiguous individual distinction between the two forms, due to the enlargement of this tooth locus in *P. macrodon*.

Simpson diagrams for upper and lower cheek teeth show that all but M2/m2 length an width dimensions vary isometrically between *P. iturus* (used as the reference sample) and *P. macrodon*, indicating that beyond size differences, the two species show significant proportional differences only in their upper and lower second molars (Fig. 3(A)). On average, *P. macrodon*'s upper and lower cheek tooth dimensions (excepted M2/m2) appear 11% and 17% larger than *P. iturus*' ones, respectively. While Matthew et al. (1929) indicated that *P. macrodon* shows molars that are relatively narrower than *P. iturus*, this fact is actually evidenced only for lower cheek teeth and the second upper molar.

The bivariate scatterplots linking first and second upper or lower molars areas illustrate the proportional differences in the second/first molar size relationship between *P. iturus* and *P. macrodon* (Fig. 3(B)). Considering the first molar size as a first-order surrogate of the overall dentition size (in accordance with the isometric relations identified by the Log-ratio diagrams), intermediate sized individuals of the two species (i.e., large *P. iturus* and small *P. macrodon* specimens) show significantly distinct second molar areas, with optimal cut-off surface values between the two forms at 15.1 mm^2 for M2 and 8.1 mm^2 for m2. The one-way

ANCOVAs confirm that the larger size of the second molar in *P. macrodon* is not an isometric byproduct of larger individual size in this group:

- M2 differences between adjusted means: F = 157.1; d.f. = 1, 67; $p = 3.15\times10^{-19}$ (homogeneity of slopes: F = 1.477; $p = 0.229$, with isometric relations in both cases based on Reduced Major Axis slopes a: $a_{P.\ iturus} = 0.95 \pm 0.117$, $p_{a=1} = 0.67$; $a_{P.\ macrodon} = 1.15 \pm 0.196$, $p_{a=1} = 0.44$);

- m2 differences between adjusted means: F = 49.5; d.f. = 1, 39; $p = 1.91\times10^{-8}$ (homogeneity of slopes: F = 3.187; $p = 0.082$, with isometric relations in both cases based on Reduced Major Axis slopes a: $a_{P.\ iturus} = 1.45 \pm 0.227$, $p_{a=1} = 0.06$; $a_{P.\ macrodon} = 0.75 \pm 0.201$, $p_{a=1} = 0.24$).

Thus, while the individual size of the first and second upper and lower molars covaries isometrically within both groups, the size of the second upper and lower molars show a significant between-group proportional difference, making it a powerful size-free discriminant parameter between *P. iturus* and *P. macrodon*.

To further investigate the taxonomic status of both forms, we finally considered four complementary metrics focusing on distinct aspects of the sample distributions (Table S1; see Section 2.2. and Appendix B for details). At the all-*Palaeostylops* sample level (1st result column in Table S1), and whereas all the expected dimorphism ratios (based on the specimens assigned a priori to *P. iturus* and *P. macrodon*, respectively) are significantly larger than one at the 95% confidence level, the "method-of-moment (MoM)" and mixture analysis techniques only performed well in estimating the expected dimorphism ratio values for M1/m1 and M2/m2. In these cases, taking into account their associated 95% confidence intervals, the unbiased coefficient of variation (V^*) exceeds 10% and the bimodality index (b) exceeds 0.555 almost everytime, indicating high level of within-sample variability and distributions far from unimodality. Failure to satisfactorily recover a 2-group structure in

premolars and M3/m3 is most likely due to strong individual size overlap (Table 2) and marked abundance differences (Table 1) between the two forms. The latter point also probably explains why the mixture analysis outperformed the MoM technique in several M1/m1 and M2/m2 cases (Appendix B).

At the *P. iturus* or *P. macrodon* sample level (2nd and 3rd result columns in Table S1, respectively), V^* and b-values appears globally much lower, indicating relatively low levels of variability (V^* = ~6% on average, < 10% in almost all cases) and unimodal distributions ($b < 0.555$ in most cases). Nevertheless, dimorphism ratio values significantly larger than 1 are estimated by the MoM and/or the mixture analysis techniques in several cases. Most particularly, *P. iturus*' M2, M3 and m1, and *P. macrodon*'s M1-M3 and m1 areas show a size-dimorphism, with estimated dimorphism ratios ranging between 1.05 and 1.15. Remarkably, in all eight cases where dimorphism is detected by mixture analysis, the large-size group appears less abundant than the small-size one, returning an average abundance ratio of 1:2 between the two groups. Observation of a possible dimorphism in the upper and lower molars is especially noteworthy, as the size of these cheek teeth strongly covaries with overall body size in almost all extant mammal groups (e.g., Creighton, 1980; Gingerich et al., 1982; Legendre, 1989; Janis, 1990). This may indicate that a body-size dimorphism, possibly of sexual origin with an estimated sex-ratio of about 1 male (?) for 2 females (?), did exist within *P. iturus* as well as *P. macrodon*, which in turn cannot be considered as two sexual morphs of the same biological species.

3.3. Taxonomic implications

Because several of the dimorphism ratios at the *P. iturus* or *P. macrodon* sample level are significantly larger than 1 (Table S1), the results of our biometric analysis suggest the occurrence of a body-size dimorphism, possibly of sexual origin within the two forms. These

results are therefore not compatible with an interpretation of the two forms as males and females of a single dimorphic species. This interpretation of the two forms as two distinct taxa is corroborated by results from other late Paleocene sites, where both forms have been shown to occur with distinctly different ratios in different levels (Kondrashov, 2002) or where only one of the two forms was present (Missiaen and Smith, 2008).

Our biometric analysis shows that *P. macrodon* is clearly larger than *P. iturus*, but with an important inter-individual variability in both groups, leading to strong distribution overlaps in all but M2/m2 cheek tooth positions (Tables 1 and 2). In *P. macrodon*, the upper and lower second molars are disproportionately enlarged compared to *P. iturus* (Fig. 3(A)) and M2/m2 areas appear as powerful size-free discriminant parameters between *P. iturus* and *P. macrodon*, with optimal cut-off surface values between the two forms at 15.1 mm^2 for M2 and 8.1 mm^2 for m2 (Fig. 3(B)). These results reflect the original description by Matthew et al. (1929), who diagnosed *P. macrodon* from *P. iturus* based only on size differences. Some of the morphological characters that have been noted to distinguish both forms, such as the shape of the dental arcade and the number of incisors are shown here to be not diagnostic. For others, such as the presence of lingual cuspules on the upper molars and of a lingual sulcus on M1, there typically is a difference between specimens assignable to each form, but these characters also show considerable variability and overlap. The similarity between both species is also clearly illustrated by the fact that in this collection of 255 upper and lower jaw fragments, only 151 (<60%) can be unambiguously assigned to either species. We therefore conclude that differences between the two forms are minor and do not require a generic level distinction. We continue to consider *P. iturus* and *P. macrodon* as separate, closely related species within a single genus, and we confirm "*Gashatostylops*" as a junior synonym of *Palaeostylops*.

4. Postcranial remains from the Flaming Cliffs arctostylopids

4.1. Tarsal bones

The collection from the Flaming Cliffs contains five calcanea and three astragali of *Palaeostylops*, which are closely similar to the better preserved tarsals of *P. iturus* described from the late Paleocene Subeng site in Inner Mongolia (Missiaen et al., 2006). Although there is some size variation, the poor preservation and the limited number of these specimens does not allow us to distinguish two distinct morphotypes. Because of this, these tarsal bones are identified as *Palaeostylops*, but are not assigned to either *P. iturus* or *P. macrodon*.

4.2. Distal tibia

The collection from the Flaming Cliffs also yielded the distal part of a left tibia of *Palaeostylops*, the first ever identified in arctostylopids. This arctostylopid distal tibia has a generalised morphology. Preservation and breakage, especially of the laterodistal corner of the bone, limit the number of observable diagnostic features (Fig. 4). The medial malleolus of the tibia is small but distinct. The lateral astragalotibial facet is about as wide as it is long, and is saddle-shaped, with an anteroposteriorly oriented middle ridge separating the larger medial cavity from the smaller lateral cavity. The most remarkable feature is the presence of an incipient tibial posterior process, located posteromedial to this middle ridge.

The tibial morphology of Paleocene and early Eocene notoungulates is poorly known but a Casamayoran (middle? to late Eocene) notoungulate distal tibia was published by Shockey and Flynn (2007). This specimen differs from the *Palaeostylops* tibia by the larger and more oblique medial malleolus, by the less saddle-shaped lateral astragalotibial facet and by the absence of tibial posterior process. The tibial morphology of several basal gliriforms and glires has been documented, and at least some of these taxa share similarities with

Palaeostylops. As shown in Figure 5, the tibial morphology of *Palaeostylops* is similar to that of *Pseudictops* and *Rhombomylus*, two taxa that were also shown to have a similar tarsal morphology (Missiaen et al., 2006). *Palaeostylops* shares the pronounced saddle shape of the lateral astragalotibial facet with *Pseudictops*, and the presence of an incipient tibial posterior process with *Rhombomylus*, while the morphology of the medial malleolus in *Palaeostylops* is intermediate between that of *Pseudictops* and *Rhombomylus*. A distinct malleolus is a primitive feature of several gliriform taxa, including pseudictopids, eurymylids, alagomyids and ischyromyids, but is lost in true lagomorphs and advanced rodents (Meng et al. 2003). A tibial posterior process is a derived character of rodents, but an incipient process is also seen in *Rhombomylus* and *Mimolagus*, but also as in various other taxa including *Solenodon*, *Eomanis*, leptictids, palaeandonts and pantolestids (Szalay, 1985; Meng et al., 2003, Rose, 1999, Rose and Lucas, 2000, Horovitz et al., 2005, Rose and von Koenigswald, 2005). Therefore, although this tibial morphology does not offer any conclusive evidence, it does add support to the idea that arctostylopids are not notoungulates but instead are basal members of the Asian gliriform radiation.

5. Relevance of *Palaeostylops* for the phylogenetic affinities of arctostylopids

The first studies of *Arctostylops* from the North American Clark Fork Basin and of *Palaeostylops* from the Mongolian Ulan-Nur Basin (Matthew, 1915; Matthew and Granger, 1925) left very little room to doubt the surprising conclusion that these arctostylopids were related with South American notoungulates. Despite the discovery of numerous other arctostylopids, the sheer abundance of *Palaeostylops* specimens from Gashato and contemporary Asian mammal sites has meant that *Palaeostylops* remained an important reference taxon in phylogenetic comparisons (Cifelli et al., 1989; Kondrashov and Lucas,

2004; Missiaen et al., 2006). In the following paragraphs, we review the significance of *Palaeostylops* for the study of arctostylopid evolution based on its anterior dentition, molars, and postcrania. By integrating this information with that on other, less abundant arctostylopid taxa and on notoungulates, we support the hypothesis that arctostylopids are basal gliriforms and that the similarities with notoungulates are the results of convergent evolution. We present the novel suggestion that within the basal gliriforms, arctostylopids are most closely related to the poorly known Astigalidae.

5.1. Anterior dentition

The arctostylopid anterior dentition was first known from *Palaeostylops* specimens found at Gashato (Matthew and Granger, 1925; Matthew et al., 1929) (Fig. 6(3)) and more recent descriptions of the anterior dentition in other arctostylopids (Cifelli et al., 1989; Tong and Wang, 2006) have not significantly altered the characterisation of the anterior dentition in Arctostylopidae (Fig. 7). In all known arctostylopids, the anterior dentition forms an evenly graded, complete dental series, without conspicuous canines and without diastemata (Cifelli et al., 1989).

The anterior dentition of notoungulates varies considerably, but most groups have at least partly molarised premolars and often large diastemata are present (Simpson, 1948). A complete, evenly graded dentition is only seen to some extent in the primitive Henricosborniidae and more strongly in the advanced typothere family Interatheriidae (Simpson, 1948; Cifelli, 1993). In contrast, such a complete, evenly graded morphology is present in the basal gliriform families Anagalidae, Pseudictopidae and Astigalidae (Zhang and Tong, 1981; Meng et al., 2003).

A second characteristic of the arctostylopid anterior dentition are the serially multicuspid, blade-like lower antemolar teeth, which again are different from those of all Notoungulata but

similar to the lower incisors and anterior premolars of the basal gliriforms *Pseudictops* (Fig. 6(1)). In fact, this similarity of the anteriormost lower teeth led Matthew and Granger (1925) to describe specimen AMNH 20426 from Gashato as a p1 of an unknown, larger species of *Palaeostylops* (Fig. 6(2)), whereas this specimen was later re-identified as a p1 of *Pseudictops* (Sulimski, 1968).

5.2. Molars

The lower molars of arctostylopids were first known from *Arctostylops* (Fig. 7(5)) from the Clark Fork Basin, the upper molars from *Palaeostylops* from Gashato (Fig. 7(3)), and immediately the arctostylopid molar dentition was considered distinct and highly specialized, different from all northern taxa known at that time and only similar to notoungulates (Matthew, 1915). Since these first discoveries of arctostylopids, numerous other forms have been described (Zheng, 1979; Zheng and Huang, 1986; Huang and Zheng, 1997, 2003; Zack, 2004; Tong and Wang, 2006; Wang et al., 2008), both from strata that are significantly older and significantly younger than those yielding *Palaeostylops* or *Arctostylops* (Wang et al., 1998; Ting 1998; Missiaen, 2011).

Based on the quadrate upper molars with distinct parastyle, smooth ectoloph and strong protoloph and metaloph, and the biselenodont lower molars with extremely reduced trigonids and distinct entolophids, *Arctostylops* and *Palaeostylops* were judged to be an aberrant group of notoungulates (Matthew, 1915; Matthew and Granger, 1925).

However, all of these features are much less developed or completely absent in the arctostylopids from older strata. *Asiostylops* from the early Nongshanian (early middle Paleocene) Lannikeng Member of the Chijiang Formation is the oldest known arctostylopid (Missiaen, 2011), and its molar morphology (Fig. 7(2)) is much less specialized than that of *Palaeostylops*. The upper molars of *Asiostylops* differ from those of younger arctostylopids by

the more triangular shape with a distinct lingual protocone and posterolingual talon shelf, by only a very faint indication of a metaconule and a hypocone, and by the weaker ectolophs with a smaller parastyle and an unreduced paracone and metacone. The lower molars of *Asiostylops* differ from those of younger arctostylopids such as *Palaeostylops* by the unreduced trigonid with a distinct paracristid, protoconid and metaconid, by the lower and less smooth ectolophid and by only an incipient development of an entolophid. The oldest arctostylopid *Asiostylops* therefore also clearly has the most primitive molar morphology of all arctostylopids, and can itself be readily derived from a primitive mammal with tribosphenic teeth.

The question of whether the strong posterolingual cusp on M1 and M2 of notoungulates is a true hypocone or whether it is a pseudohypocone as it is in arctostylopids (Cifelli et al., 1989; Kondrashov and Lucas, 2004) is therefore not relevant for the evolutionary origin of Arctostylopidae. The quadrate shape with a strong pseudohypocone reminiscent of notoungulates is absent or weak in the oldest and most primitive arctostylopids, as well as in many of the youngest taxa (Fig. 7(4)). It is only in the evolution towards the *Arctostylops* and *Palaeostylops* type of dentition (Tong and Wang, 2006), that the upper molars become fully quadrate and develop a large pseudohypocone, but even within the genus *Palaeostylops* where the pseudohypocone is the most clearly developed, its development is variable, as shown above (Fig. 2).

The primitive arctostylopid molar morphotype as seen in the low-crowned and incipiently lophodont *Asiostylops* lacks all features originally used to link arctostylopids with notoungulates and neither does it present any of the diagnostic molar features of notoungulates (Simpson 1948, 1967). This means there is no longer any support for the hypothesis that arctostylopids were derived from notoungulates. Conversely, in some arctostylopids, most notably *Palaeostylops* and *Arctostylops*, molar characters such as the

high and smooth ectoloph and ectolophid, quadrate upper molars and lower molars with a short, reduced trigonid and a distinct entolophid may seem reminiscent of notoungulates such as *Leontinia* or *Notostylops* (Fig. 7(8)) but are not shared with all notoungulates including the Henricosborniidae, the most primitive notoungulate family (Fig. 7(6,7)). Notoungulates therefore cannot be derived from a *Palaeostylops*-like ancestor, and similarities between both groups evolved independently, representing a remarkable case of convergent evolution.

The primitive arctostylopid molar morphotype however does share a number of similarities with basal gliriforms (Anagalidae, Pseudictopidae and Astigalidae). Basal gliriform synapomorphies include reduction or loss of the upper molar stylar shelf, the partial fusion of the paracone and metacone, the development of a distinct precingulum and postcingulum without hypocone, and the partial reduction of the lower molar trigonid (Hu, 1993; Tong and Wang, 2006), all of which are also present in arctostylopids. Basal gliriform upper molars further resemble those of primitive arctostylopids by the weak upper molar conules, the development of distinct crests running from the protocone to the paracone and metacone and by undergoing heavy wear. Among these basal gliriforms, the early Paleocene *Astigale* (Fig.8(1)) and *Zhujegale* from the poorly known Asian family Astigalidae (Zhang and Tong, 1981) seem to have the strongest morphological similarities with Arctostylopidae, and thus possibly the closest phylogenetic affinities. While pseudictopids and anagalids have transversely elongated upper molars, early Paleocene astigalids have more nearly square upper molars with stronger lingual cingula. Similarly, these primitive astigalids have a less reduced and less anteroposteriorly compressed lower molar trigonid, and are less unilaterally hypsodont than other basal gliriforms. Astigalidae however differ from Arctostylopidae by the larger canines, the larger upper molar hypocones and the absence of an entolophid or a strong ectolophid on the lower molars.

5.3. Postcrania

Detailed studies of arctostylopid postcrania have been limited to the astragalus and calcaneum of *Palaeostylops* and *Arctostylops* (Cifelli et al., 1989; Missiaen et al., 2006), although the *Arctostylops* tarsals are in fact part of a partial skeleton (Bloch, 1999). Our new *Palaeostylops* collection from Gashato yielded the first *Palaeostylops* tibia. None of the arctostylopid postcrania known so far present any diagnostic characters or typical synapomorphies of notoungulates. Instead, they share a mosaic of derived characters with primitive gliriforms, most notably the genera *Pseudictops* and *Rhombomylus*. This has led to the hypothesis that Arctostylopidae are not related to Notoungulata, but instead are basal members of the Asian gliriform radiation (Cifelli et al., 1989, Missiaen et al., 2006) which is further supported by the *Palaeostylops* tibia reported here from Gashato.

6. Conclusions

The abundant arctostylopid dental remains reported here come from a large single sample from the type area of *Palaeostylops*. They can be divided into two closely similar morphotypes that correspond well with *Palaeostylops iturus* and *P. macrodon* as originally published (Matthew and Granger, 1925; Matthew et al., 1929). Both morphotypes however display an important and overlapping morphological variation. The biometrical analysis of the new material confirms the existence of significant absolute and relative cheek teeth size differences between the two forms. Because our study of morphological and biometrical variability is most consistent with the interpretation of both forms as separate species in a single genus, we identify these two morphotypes as *P. iturus* and *P. macrodon*, and confirm "*Gashatostylops*" as a junior synonym of *Palaeostylops*.

The material reported here also includes new postcranial remains, including the first *Palaeostylops* tarsals known from Mongolia and the distal part of the previously unknown

Palaeostylops tibia. The morphology of this partial tibia is not highly specialised, but does present similarities with the primitive gliriforms *Pseudictops* and *Rhombomylus*, such as a small but distinct medial malleolus, a saddle-shaped lateral astragalotibial facet and an incipient tibial posterior process.

Historically, the Gashato area was the second region where arctostylopid fossils were found, and because of the abundance of these fossils, *Palaeostylops* has remained an important taxon for phylogenetic comparisons. In the light of more recent discoveries of other arctostylopid fossils, we critically review the importance of *Palaeostylops* for the understanding of the phylogeny of the family. *Palaeostylops* arguably remains the best available source of information on the arctostylopid anterior dentition. The evenly graded, complete dentition of arctostylopids with their characteristic serially multicuspid, blade-like lower premolars is unlike that of notoungulates, but does resemble that of basal gliriforms. Although the molar morphology of *Palaeostylops* presents some similarities with South American Notoungulata, more recently discovered stratigraphically older and morphologically more primitive arctostylopids show that these similarities arose independently in both groups. The primitive arctostylopid molar morphotype is therefore better exemplified by the early middle Paleocene *Asiostylops.* We show that this morphotype exhibits a number of gliriform synapomorphies and that, among basal gliriforms, Arctostylopidae seem most similar to the poorly known early Paleocene family Astigalidae.

Acknowledgements

The authors thank Lilian Bergqvist (UFRJ, Rio de Janeiro), Judy Galkin and Carl Mehling (AMNH, New York), and Christian de Muizon and Claire Sagne (MNHN, France) for access to reference material. Gregg Gunnell (University of Michigan, Ann Arbor), Nathalie Van Hoey and Wilfried Miseur (RBINS, Brussels) assisted during the photography of some of the

specimens. Khishigjav Tsogtbaatar (MPC, Ulaanbaatar) helped with cataloguing the specimens. Chris Beard (Carnegie Museum of Natural History, Pittsburgh), Rodolphe Tabuce (Université Montpellier II, Montpellier) and one anonymous reviewer commented and improved the manuscript of this paper. Pieter Missiaen is a Postdoctoral Fellow of the Research Foundation Flanders – FWO Vlaanderen. Monte Carlo computations of Fig. 3 achieved with the IDL-program SIST (Escarguel, 2008; available on request to G.E. [gilles.escarguel@univ-lyon1.fr]). We dedicate this paper to the late Demberel Dashzeveg (1936-2010), who was a pioneer in Mongolian and international paleontology for almost 50 years, and part of the field team who found the *Palaeostylops* fossils published here.

Appendices A, B. Supplementary material

Supplementary material (appendices A, B, Table S1 and Fig. S1) associated with this article can be found, in the online version, at doi:

References

Bloch, J.I., 1999. Partial skeleton of *Arctostylops* from the Paleocene of Wyoming: arctostylopid-notoungulate relationship revisited. Journal of Vertebrate Paleontology 19, 32A.

Burnham, K. P., Anderson, D. R., 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer-Verlag, New York, USA.

Cifelli, R.L., 1983. Eutherian tarsals from the late Paleocene of Brazil. American Museum Novitates 2761, 1-31.

Cifelli, R.L., Schaff, C.R., McKenna, M.C., 1989. The relationships of the Arctostylopidae (Mammalia): new data and interpretation. Bulletin of the Museum of Comparative Zoology 152, 1-44.

Cifelli, R.L., 1993. The Phylogeny of the Native South American Ungulates, in: Szalay, F.S., Novacek, M.J., and McKenna, M.C. (Eds.), Mammal Phylogeny: Placentals. Springer Verlag, New York, pp. 195-216 .

Creighton, G.K., 1980. Static allometry of mammalian teeth and the correlation of the tooth size and the body size in contemporary mammals. Journal of Zoology, London 191, 435-443.

Der, G., Everitt, B.S., 2002. A handbook of statistical analyses using SAS, 2nd ed. Chapman & Hall/CRC, Boca Ratón, CA.

Escarguel, G., 2008. Macroécologie en temps profond: motifs, rythmes et modalités des changements de biodiversité à l'échelle des temps géologiques. Mémoire pour le Diplôme d'Habilitation à Diriger des Recherches, Université Claude Bernard Lyon 1 (inédit), 428 p.

Gingerich, P.D., 1981. Variation, sexual dimorphism, and social structure in the early Eocene horse *Hyracotherium* (Mammalia, Perissodactyla). Paleobiology 7, 443-455.

Gingerich, P.D. 1985. South American Mammals in the Paleocene of North America, in: Stehli F.G., Webb, S.D. (Eds.), The Great American Biotic Interchange. Plenum Press, New York, pp. 123-137.

Gingerich, P.D., Smith, B.H., Rosenberg, K., 1982. Allometric scaling on the dentition of primates and prediction of body weight from tooth size in fossils. American Journal of Physical Anthropology 58, 81-100.

Hammer Ø., Harper D.A.T., Ryan P.D., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4, 9 p.

Holm S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65-70.

Horovitz, I., Storch, G., Martin, T., 2005. Ankle structure in Eocene pholidotan mammal *Eomanis krebsi* and its taxonomic implications. Acta Palaeontologica Polonica 50, 545-548.

Hu, Y., 1993. Two new genera of Anagalidae (Anagalida, Mammalia) from the Paleocene of Qianshan,Anhui and the phylogeny of anagalids. Vertebrata Palasiatica 31, 153-181.

Huang, X.-S. 2003 Mammalian remains from the late Paleocene of Jiashan, Anhui. Vertebrata Palasiatica 41, 42-54.

Huang X S, Zheng J J, 1997. A new Tillodont from the Paleocene of Nanxiong Basin, Guangdong. Vertebrata Palasiatica, 35, 290-306.

Huang, X.-S., Zheng J.-J, 2003. Note on two new mammalian species from the late Paleocene of Nanxiong, Guangdong. Vertebrata Palasiatica 41, 271-277.

Janis, C.M., 1990. Correlation of cranial and dental variables with body size in ungulates and macropodoids, in: Damuth J., MacFadden B.J. (Eds.), Body size in mammalian paleobiology: Estimation and biological implications. Cambridge University Press, Cambridge, pp. 255-299.

Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19, 101-108.

Josephson, S.C., Juell, K.E., Rogers, A.R., 1996. Estimating sexual dimorphism by method-of-moments. American Journal of Physical Anthropology 100, 191-206.

Kondrashov, P.E., 2002. Late Paleocene Arctostylopidae (Mammalia, Notoungulata) from Mongolia. Journal of Vertebrate Paleontology 22, 75A.

Kondrashov, P.E., Lucas, S.G., 2004. Palaeostylops iturus from the Upper Paleocene of Mongolia and the status of Arctostylopida (Mammalia, Eutheria). New Mexico Museum of Natural History and Science Bulletin 26, 195-203.

Legendre, S., 1989. Les communautés de mammifères du Paléogène (Eocène supérieur et Oligocène) d'Europe occidentale: structures, milieux et evolution. Münchner Geowissenschaftlichen Abhandlungen, A 16, 1-110.

Marshall, L. G., Muizon, C. de, Sigé, B., 1983, *Perutherium altiplanense*, un notongulé du Crétacé Supérieur du Pérou. Palaeovertebrata 13, 145-155.

Matthew, W.D. 1915. A revision of the lower Eocene Wasatch and Wind River faunas. Part IV. Entelonychia, Primates, and Insectivora (part). Bulletin of the American Museum of Natural History 34, 429-483.

Matthew, W.D., Granger, W., 1925. Fauna and correlation of the Gashato Formation of Mongolia. American Museum Novitates 189, 1-12.

Matthew, W.D., Granger, W., Simpson, G.G., 1929. Additions to the fauna of the Gashato Formation of Mongolia. American Museum Novitates 376, 1-12.

Meng, J., Zhai, R., Wyss, A.R., 1998. The late Paleocene Bayan Ulan fauna of Inner Mongolia, China. Bulletin of Carnegie Museum of Natural History 34, 148-185.

Meng, J., Hu, Y., Li, C., 2003. The osteology of *Rhombomylus* (Mammalia, Glires): Implications for phylogeny and evolution of Glires. Bulletin of the American Museum of Natural History 275, 1-247.

Missiaen, P., 2011. An updated mammalian biochronology and biogeography for the early Paleogene of Asia. Vertebrata Palasiatica 49, 29-52 .

Missiaen, P., Smith, T., 2008. The Gashatan (late Paleocene) mammal fauna from Subeng, Inner Mongolia, China. Acta Palaeontologica Polonica 53, 357-378.

Missiaen, P., Smith, T., Guo, D.-Y., Bloch J.I., Gingerich, P.D., 2006. Asian Gliriform origin for arctostylopid mammals. Naturwissenschaften 93, 407-411.

Ni, X., Beard, K.C., Meng, J., Wang, Y., Gebo, D.L., 2007. Discovery of the First Early Cenozoic Euprimate (Mammalia) from Inner Mongolia. American Museum Novitates 3571, 1-11.

Patterson, B., Pascual, R., 1972. The fossil mammal fauna of South America, in: Keast, A., Erk, F.C., Glass, B.P. (Eds.), Evolution, mammals and southern continents. State University of New York Press, New York, pp. 247-309.

Rose, K.D., 1999. Postcranial skeleton of eocene Leptictidae (Mammalia), and its implications for behavior and relationships. Journal of Vertebrate Paleontology 19, 355-372.

Rose, K.D., Luvas, S.G., 2000. An early Paleocene palaeanodont (Mammalia, ?Pholidota) from New Mexico, and the origin of Palaeanodonta. Journal of Vertebrate Paleontology 20139-156.

Rose, K.D., von Koenigswald, W., 2005. An Exceptionally Complete Skeleton of *Palaeosinopa* (Mammalia,Cimolesta,Pantolestidae) from the Green River Formation,and Other Postcranial Elements of the Pantolestidae from the Eocene of Wyoming (USA). Palaeontographica 273, 55-96.

Russell, D.E., Zhai, R.J., 1987. The Paleogene of Asia: mammals and stratigraphy. Mémoires du Muséum national d'Histoire Naturelle, C 52, 1-488.

Shockey, B., Flynn, J.J., 2007. Morphological Diversity in the Postcranial Skeleton of Casamayoran (?Middle to Late Eocene) Notoungulata and Foot Posture in Notoungulates. American Museum Novitates 3601,1-26.

Simpson G.G., 1941. Explanation of Ratio Diagrams. In: G.G. Simpson (Ed.), Large Pleistocene Felines of North America. American Museum Novitates 1136, 1-27.

Simpson, G.G. 1948. The beginning of the age of mammals in South America. Part 1. Bulletin of the American Museum of Natural History 91, 1-232.

Simpson, G.G. 1967. The beginning of the age of mammals in South America. Part 2. Bulletin of the American Museum of Natural History 137, 1-260.

Simpson G.G., Roe A., Lewontin R.C., 1960. Quantitative Zoology, revised edition. Harcourt, Brace, and World, New York.

Sokal, R. R., Rohlf, F. J. (1995). Biometry (3rd Edition). W. H. Freeman and Co., New York.

Sulimski, A., 1968. Paleocene genus Pseudictops Matthew, Granger and Simpson 1929 (Mammalia) and its revision. Palaeontologia Polonica 19, 101-132.

Szalay, F.S., 1985. Rodent and lagomorph morphotype adaptations, origins and relationships: Some postcranial attributes analyzed. In: Luckett, W.P., Hartenberger, J.L. (Eds.), Evolutionary relationships among rodents: A multidisciplinary analysis. Plenum Press, New York, pp. 83-132.

Ting, S., 1998. Paleocene and early Eocene Land Mammal Ages of Asia. Bulletin of Carnegie Museum of Natural History 34, 124-147.

Titterington, D., Smith, A., Makov, U., 1985. Statistical analysis of finite mixture distributions. John Wiley & Sons, Chichester, U.K.

Tong, Y., Wang, J., 2006. Fossil Mammals from the Early Eocene Wutu Formation of Shandong Province. Palaeontologia Sinica, new series C 192, 1-195.

Wang, Y., Hu, Y., Chow, M., Li, C., 1998. Chinese Paleocene Mammal faunas and their correlation. In: Beard, K.C., Dawson, M.R. (Eds.), Dawn of the Age of Mammals in Asia. Bulletin of Carnegie Museum of Natural History 34, 89-123.

Wang, Y., Meng, J., Ni, X., Beard, K.C., 2008. A new Early Eocene arctostylopid (Arctostylopida, Mammalia) from the Erlian Basin, Nei Mongol (Inner Mongolia), China. Journal of Vertebrate Paleontology 28, 553-558.

Wang, Y., Meng, J., Ni, X., Li, C., 2007. Major events of Paleogene mammal radiation in China. Geological Journal 42, 415-430.

Wright S. P., 1992. Adjusted p-values for simultaneous inference. Biometrics 48, 1005-1013.

Zack, S.P., 2004. An Early Eocene Arctostylopid (Mammalia: Arctostylopida) from the Green River Basin, Wyoming. Journal of Vertebrate Paleontology 24, 498-501.

Zhang, Y., Tong, Y., 1981. New anagaloid mammals from Paleocene of South China. Vertebrata Palasiatica 19, 133-144.

Zheng, J. J., 1979. The Paleocene notoungulates of Jiangxi, in: The Mesozoic and Cenozoic Red Beds of South China. Science Press, Beijing. pp. 387–394 [in Chinese].

Zheng, J. J., Huang, X.S.,1986. New arctostylopids (Notoungulata, Mammalia) from the late Paleocene of Jiangxi. Vertebrata PalAsiatica 24, 121-128.

Captions

Table 1. Measurements for *Palaeostylops iturus* and *P. macrodon* cheek teeth from the Gashatan of the Khashat Formation in the Flaming Cliffs area (Mongolia). All measurements in mm. N: number of measured specimens; min.: minimum; max.: maximum; Std.-Dev.: standard-deviation; Bivariate D&H *p*-val.: *p*-value of the Doornik and Hansen omnibus significance test for bivariate (Length × Width) normality, indicating that all positions but *P. iturus*' P4 show normal length × width distributions. Non-normality of *P. iturus* P4's distribution is due to a single specimen (MPC-M 30/100) showing an unusually large length; removal of this outlier returns a non-significant D&H *p*-value (0.632).

Table 2. Wilks' λ tests of the null hypothesis that the *P. iturus* and *P. macrodon* samples come from populations with equal bivariate (Length × Width) means, and Kolmogorov-Smirnov univariate nonparametric tests of the null hypothesis that the *P. iturus* and *P. macrodon* samples come from populations with equal distributions. Holm's *p*-val.: significance level based on a sequential Bonferroni correction (Holm's [1979] procedure; see Wright, 1992) for multiple testing; bold values indicate significant differences at a 95% experimentwise confidence level.

Fig. 1. **A.** Geographic location of the studied fossil locality (F) in the Flaming Cliffs area (Ulan-Nur Basin, Gobi Desert, Mongolia); **B.** Panoramic view of the Khashat Formation in the Flaming Cliffs area; **C.** Close-up view of the fossiliferous small sandy lens (white star; <1 m^3) from which the arctostylopid collection studied here comes from (featuring D. Dashzeveg on the left and J.-L. Hartenberger on the right); **D.** Simplified stratigraphic log of the Late Cretaceous-Early Paleogene section in the studied fossil

locality area; the black star indicates the stratigraphic position of the fossiliferous lens within the Member I (= Khashat Mb = Khashat Svita of Russell and Zhai, 1987: fig. 22) of the Kashat (= Gashato) Formation.

Fig. 2. *Palaeostylops* upper dentitions from the Gashatan of the Khashat Formation in the Flaming Cliffs area (Mongolia). **1-6.** *P. iturus*. 1. MPC-M 30/143, left M1-2; 2. MPC-M 30/234a, right P2-M3; 3. MPC-M 30/146, right M2-3; 4. MPC-M 30/283, right M1-2; 5. MPC-M 30/236b, left P3-M3; 6. MPC-M 30/234b, left P2-M3. **7-11**. *P. macrodon*. 7. MPC-M 30/287, left M1-2; 8. MPC-M 30/233a, right P2-M3; 9. MPC-M 30/133, left P3-M2; 10. MPC-M 30/233b, left P3-M3; 11. MPC-M 30/109, left M1-2. All in occlusal view. Scale bar = 5 mm.

Fig. 3. **A**. Simpson diagrams of Log-ratios based on the mean lengths and widths of upper and lower cheek teeth (Table 1). Reference sample: *P. iturus* from the Gashatan of the Khashat Formation in the Flaming Cliffs area (MNI = 32); thin black lines: 95% bootstrapped confidence intervals around the observed Log-Ratio values (circles) for *P. macrodon* from the same fossil locality (MNI = 14); bold gray lines: expected distribution (95% C.I.) under the null hypothesis that *P. iturus* and *P. macrodon* share the same tooth dimensions. Sample and null hypothesis confidence intervals estimated from 10,000 parametric bootstrap iterations. **B**. Scatterplots of the Ln-areas of the first vs. second upper and lower molars. Ellipses show 95% sample concentration under a bivariate normal distribution working hypothesis; dotted lines within the ellipses: reduced major axes; horizontal dashed lines: optimal cut-off values between *P. iturus* and *P. macrodon* M2/m2 surfaces (M2: Proba(indv. $\in$ *P. iturus* $\mid$ L×W < 15.1 mm^2 = Proba(indv. $\in$ *P. macrodon* $\mid$ L×W > 15.1 mm^2 = 99.7%; m2:

Proba(indv. ∈ *P. iturus* | L×W < 8.1 mm^2 = 99.9%, Proba(indv. ∈ *P. macrodon* | L×W > 8.1 mm^2 = 99.8%; see Favre et al. [2008] for methodological details).

Fig. 4. Distal tibia of *Palaeostylops* from the Gashatan of the Khashat Formation in the Flaming Cliffs area (Mongolia). MPC-M 30/328 in posterior (**1**), distal (**2**), anterior (**3**) and medial (**4**) views. MM: Medial malleolus; TPP: Tibial posterior process; LAT: Lateral astragalotibial facet. Scale bar = 5 mm.

Fig. 5. Comparison of the distal tibia of *Pseudictops* (**1**, **2**), *Palaeostylops* (**3**, **4**), and *Rhombomylus* (**5**, **6**) in posterior (1, 3, 5) and distal (2, 4, 6) views. MM: Medial malleolus; TPP: Tibial posterior process. Illustration of *Pseudictops* based on Sulimski (1968); illustration of *Rhombomylus* based on Meng et al. (2003). Scale bars = 5 mm.

Fig. 6. Comparison of the lower anterior dentition of *Pseudictops* (**1**, **2**) and *Palaeostylops* (**3**) in lingual view. 1**.** i1-3 and c based on MgM-II/13 (reversed) and p1-m3 based on MgM-II/15, modified from Sulimski (1968); 2**.** AMNH 20422, originally described as *Palaeostylops* sp. by Matthew and Granger 1925; 3**.** AMNH 20414, holotype of *Palaeostylops iturus*. Scale bars = 1 cm (1, 2), 5 mm (3).

Fig. 7. Comparison of upper and lower dentition of Astigalidae (**1**), Arctostylopidae (**2-5**) and Notoungulata (**6-8**) in occlusal view. 1. *Astigale nanxiongensis* from the Shanghuan (early Paleocene) Shanghu Formation in Jintang, Guandong Province, China. IVPP V5215: left P3-M3 and p3-m3; 2. *Asiostylops spanios* from the Nongshanian (middle Paleocene) Chijiang Formation in Laolingbei, Jiangxi Province, China. IVPP V5042: P3-M3 (reversed from right side) and left c-m3; 3. *Palaeostylops iturus* from the Gashatan (late Paleocene) Khashat

Formation in the Flaming Cliffs area, Mongolia. MPC-M 30/234b: left P2-M3. MPC-M 30/288: c-m3 (reversed from right side); 4. *Migrostylops* from the Bumbanian (early Eocene) Wutu Formation in Wutu, Shandong Province, China. IVPP V10734: left I1-M3 of *M. rosella*. IVPP V10733-4: left p4-m3 from *M. roboreus*; 5. *Arctostylops steini* from the Tiffanian-5a (late Paleocene) Fort Union Formation of Polecat Bench, Bighorn Basin, Wyoming. MCZ 20004: left C-M3 and left dentary with i3-m1; 6. *Colbertia magellanica* (Typotheria: Oldfieldthomasiidae) from the Itaboraian (late Paleocene) of Itaborai, Brazil. AMNH 49873: left P3-M3. AMNH 49879: p3-m3 (reversed from right side); 7. *Simpsonotus praecursor* (Notioprogonia: Henricosbornidae) from the Riochican (late Paleocene) Mealla Formation in Jujuy, Argentina. MLP 73-VII-3-II: left I2-M3 and left i3-m3; 8. *Notostylops murinus* (Notioprogonia: Notostylopidae) from the Casamayoran (early Eocene) Casamayor Formation in Colhue Huapi, Chubut, Argentina. AMNH 28956: left I3-M3. AMNH 28727: left p2-m3. Scale bars = 5 mm.

Supplementary material

Appendix A. Statistical analysis of Simpson's (1941) Log-ratios.

Appendix B. Biostatistical analysis of dental measurements: Computational details regarding *V**, *b* and *D* metrics, and mixture analysis results

Table S1. Unbiased coefficient of variation (***V****), bimodality index (***b***), Dimorphism ratio of the "method-of-moments(MoM)" technique (***D***) and mixture analysis (Mixt.) results for the length (L), width (W) and Ln(L × W) of the upper and lower third premolars to third molars of all measured *Palaeostylops* teeth, and measured teeth a priori assigned to *P. iturus* or to *P. macrodon*. Ex.R.: expected ratio between the largest and smallest group means based on the a priori assignment of specimens to one of the two groups.

Figure S1. Bivariate Length × Width scatterplots for *Palaeostylops iturus* and *P. macrodon* cheek teeth from the Gashatan of the Khashat Formation in the Flaming Cliffs area (Mongolia). All measurements in mm.

Table 1

Table 1.

		Length					Width					Bivariate
		N	min	max	mean	Std.-dev.	N	min	max	mean	Std.-dev.	D&H *p* -val.
P. iturus	P2	11	1.6	2.0	1.79	0.122	11	1.1	1.8	1.41	0.192	0.203
	P3	15	1.6	2.1	1.95	0.119	15	1.3	1.9	1.67	0.198	0.055
	P4	34	1.8	2.4	1.97	0.115	35	1.7	2.6	2.22	0.192	0.0008
	M1	50	2.2	3.0	2.61	0.185	50	2.4	3.3	2.77	0.201	0.844
	M2	58	2.7	3.8	3.25	0.245	58	2.8	3.6	3.31	0.203	0.246
	M3	39	1.9	2.8	2.34	0.179	37	2.3	3.3	2.90	0.215	0.061
	p2	4	1.7	1.9	1.83	0.096	4	0.9	1.1	1.00	0.082	---
	p3	9	1.6	2.1	1.92	0.164	9	0.9	1.2	1.04	0.088	0.355
	p4	16	2.0	2.5	2.30	0.137	16	1.1	1.4	1.19	0.089	0.27
	m1	31	2.5	3.0	2.76	0.136	29	1.3	1.6	1.42	0.094	0.441
	m2	56	2.8	3.8	3.33	0.190	55	1.4	2.0	1.73	0.135	0.681
	m3	26	2.2	2.7	2.52	0.196	32	1.1	1.7	1.35	0.139	0.387
P. macrodon	P2	7	1.9	2.2	2.07	0.095	7	1.2	1.5	1.31	0.135	---
	P3	10	2.0	2.4	2.16	0.143	10	1.6	2.1	1.86	0.165	0.844
	P4	15	2.0	2.5	2.14	0.159	14	2.1	2.7	2.42	0.167	0.414
	M1	20	2.8	3.3	3.04	0.150	20	2.9	3.6	3.20	0.209	0.917
	M2	21	4.4	5.5	5.01	0.290	21	3.7	4.7	4.17	0.285	0.981
	M3	16	2.2	2.8	2.46	0.186	15	3.0	3.6	3.23	0.168	0.834
	p2	2	1.9	2.0	---	---	2	1.0	1.1	---	---	---
	p3	6	2.0	2.3	2.23	0.121	6	1.0	1.3	1.18	0.117	---
	p4	7	2.4	2.8	2.66	0.127	8	1.2	1.4	1.29	0.083	---
	m1	14	2.9	3.7	3.24	0.221	14	1.5	1.8	1.61	0.092	0.383
	m2	19	4.3	5.1	4.71	0.216	19	2.0	2.4	2.18	0.108	0.792
	m3	7	2.8	3.4	3.01	0.212	9	1.4	1.6	1.50	0.087	---

Table 2

Table 2.

	Kolmogorov-Smirnov test						Wilks' λ test		
	Length			Width			Length $\times$ Width		
	D	*p* -value	Holm's *p* -val.	*D*	*p* -value	Holm's *p* -val.	λ	*p* -value	Holm's *p* -val.
P2	0.766	5.5E-03	5.6E-02	0.364	5.2E-01	7.8E-01	2.8E-01	6.29E-05	**2.51E-04**
P3	0.633	8.3E-03	7.5E-02	0.400	2.2E-01	7.8E-01	5.8E-01	2.6E-03	**7.73E-03**
P4	0.441	2.9E-02	2.3E-01	0.437	3.1E-02	2.3E-01	6.3E-01	4.07E-05	**2.44E-04**
M1	0.780	1.4E-08	**2.5E-07**	0.740	9.1E-08	**1.5E-06**	4.1E-01	1.12E-13	**1.01E-12**
M2	1.000	7.6E-15	**1.7E-13**	1.000	7.6E-15	**1.7E-13**	9.6E-02	2.35E-39	**2.58E-38**
M3	0.317	1.9E-01	7.8E-01	0.683	3.9E-05	**5.5E-04**	0.6376	2.04E-05	**1.43E-04**
p3	0.833	5.1E-03	5.6E-02	0.556	1.4E-01	7.0E-01	4.5E-01	8.1E-03	**8.06E-03**
p4	0.857	5.0E-04	**6.0E-03**	0.384	3.8E-01	7.8E-01	3.8E-01	5.92E-05	**2.96E-04**
m1	0.821	1.8E-06	**2.9E-05**	0.679	1.5E-04	**2.0E-03**	0.3167	1.84E-10	**1.47E-09**
m2	1.000	1.1E-13	**2.2E-12**	0.964	9.7E-13	**1.8E-11**	0.09352	2.94E-37	**2.94E-36**
m3	1.000	7.0E-06	**1.1E-04**	0.520	6.7E-02	4.0E-01	0.4118	0.004873	**9.75E-03**

Figure 1
Click here to download high resolution image

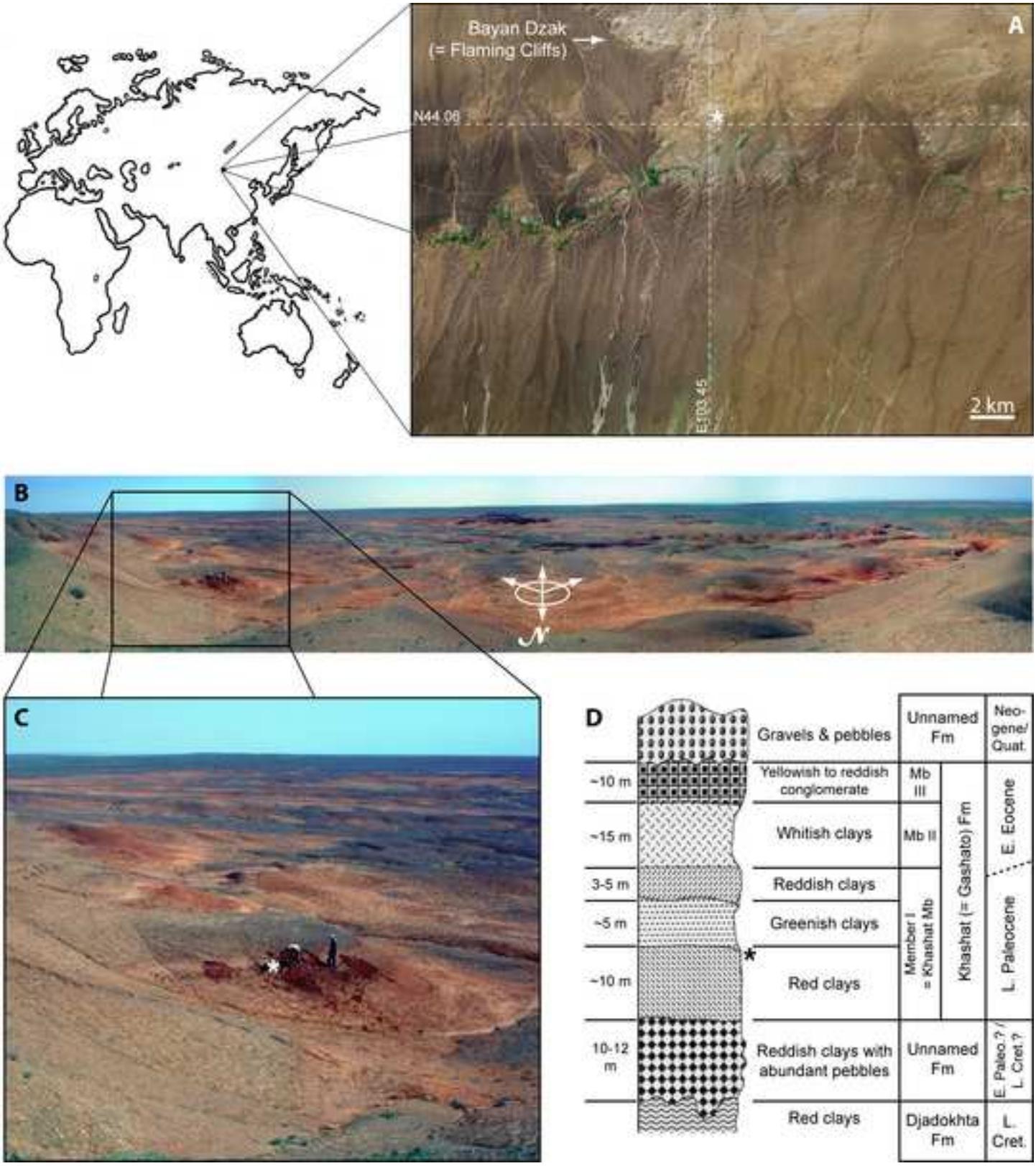


Figure 2
Click here to download high resolution image

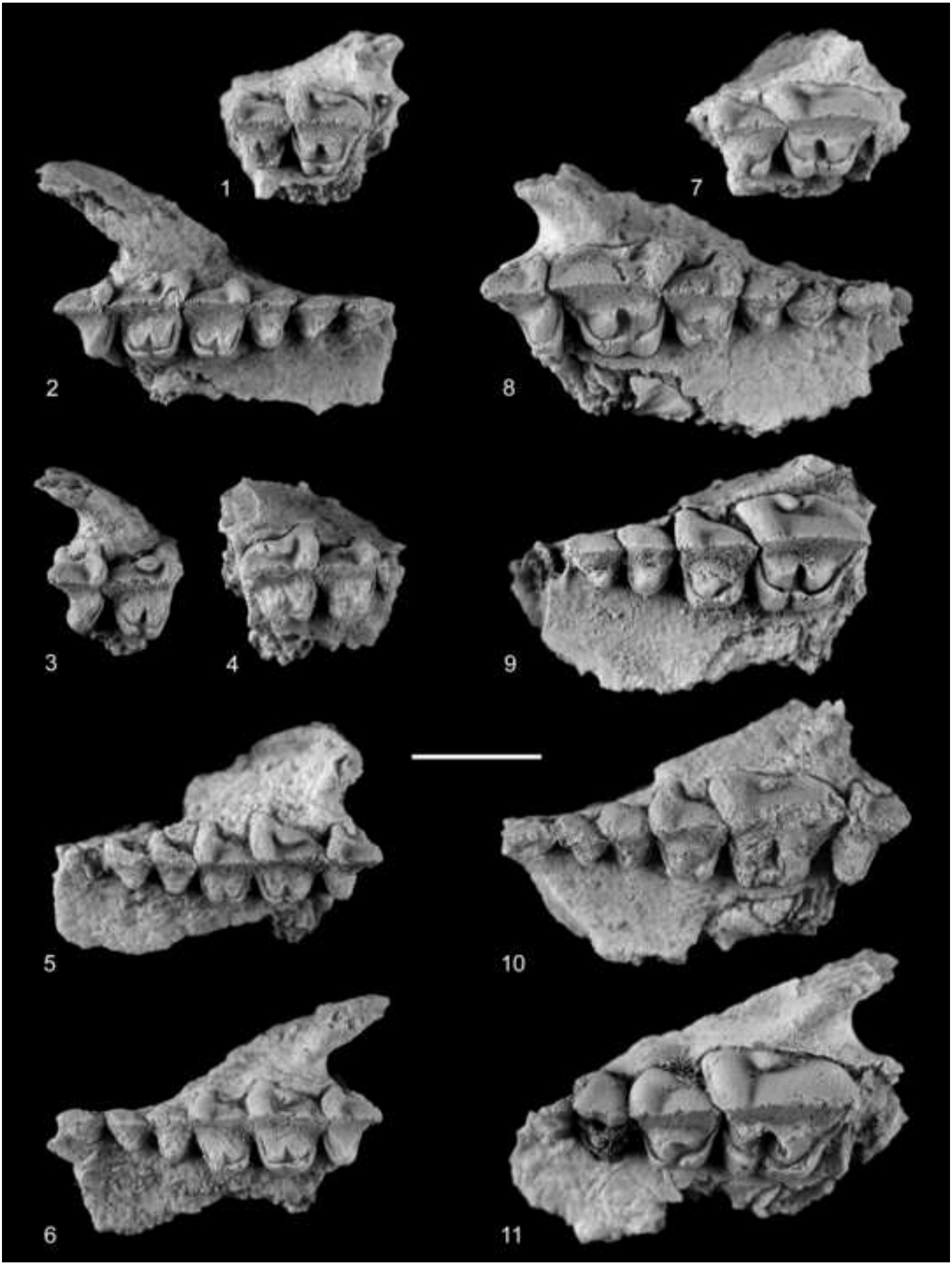


Figure 3
Click here to download high resolution image

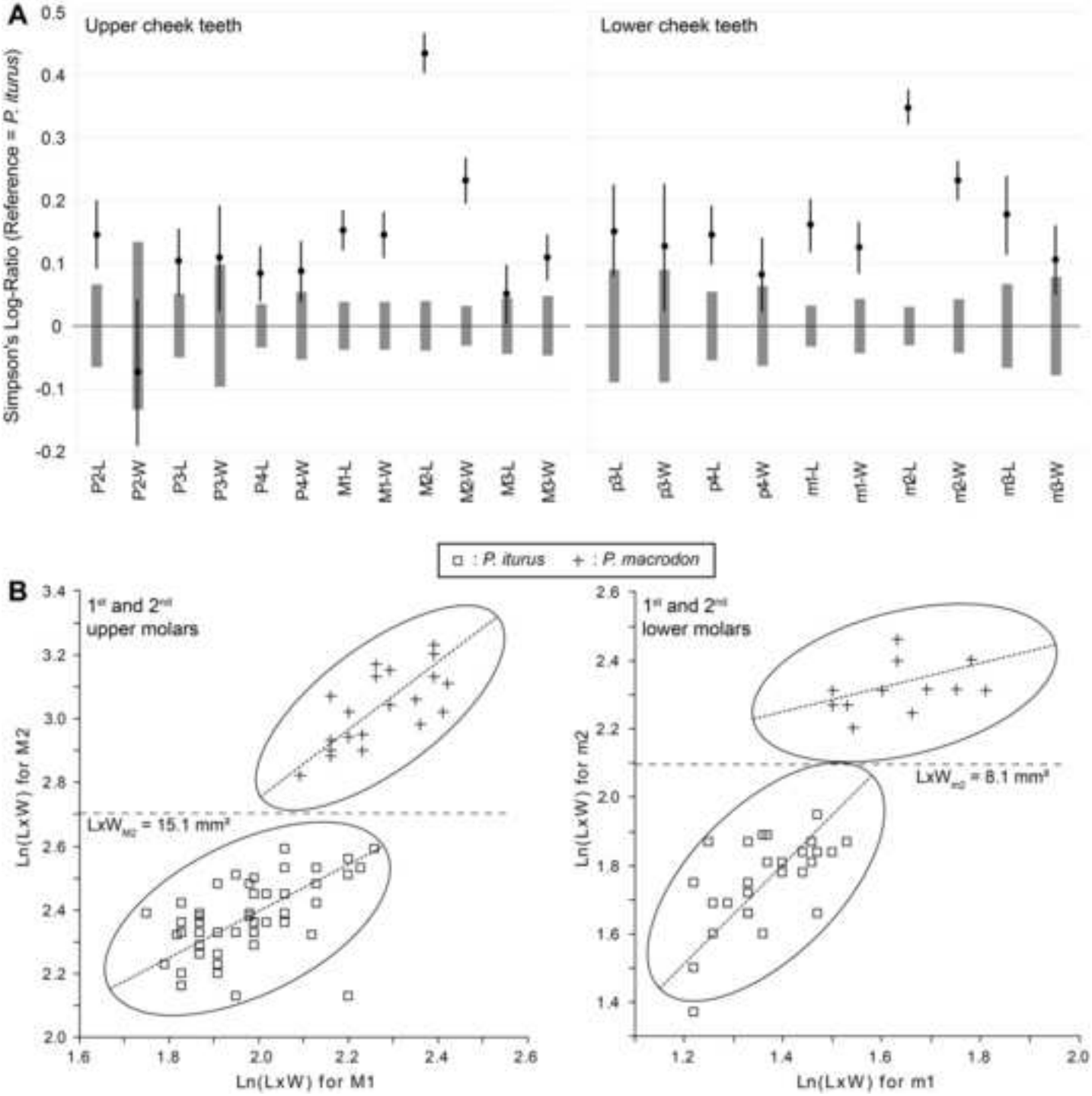


Figure 4
Click here to download high resolution image

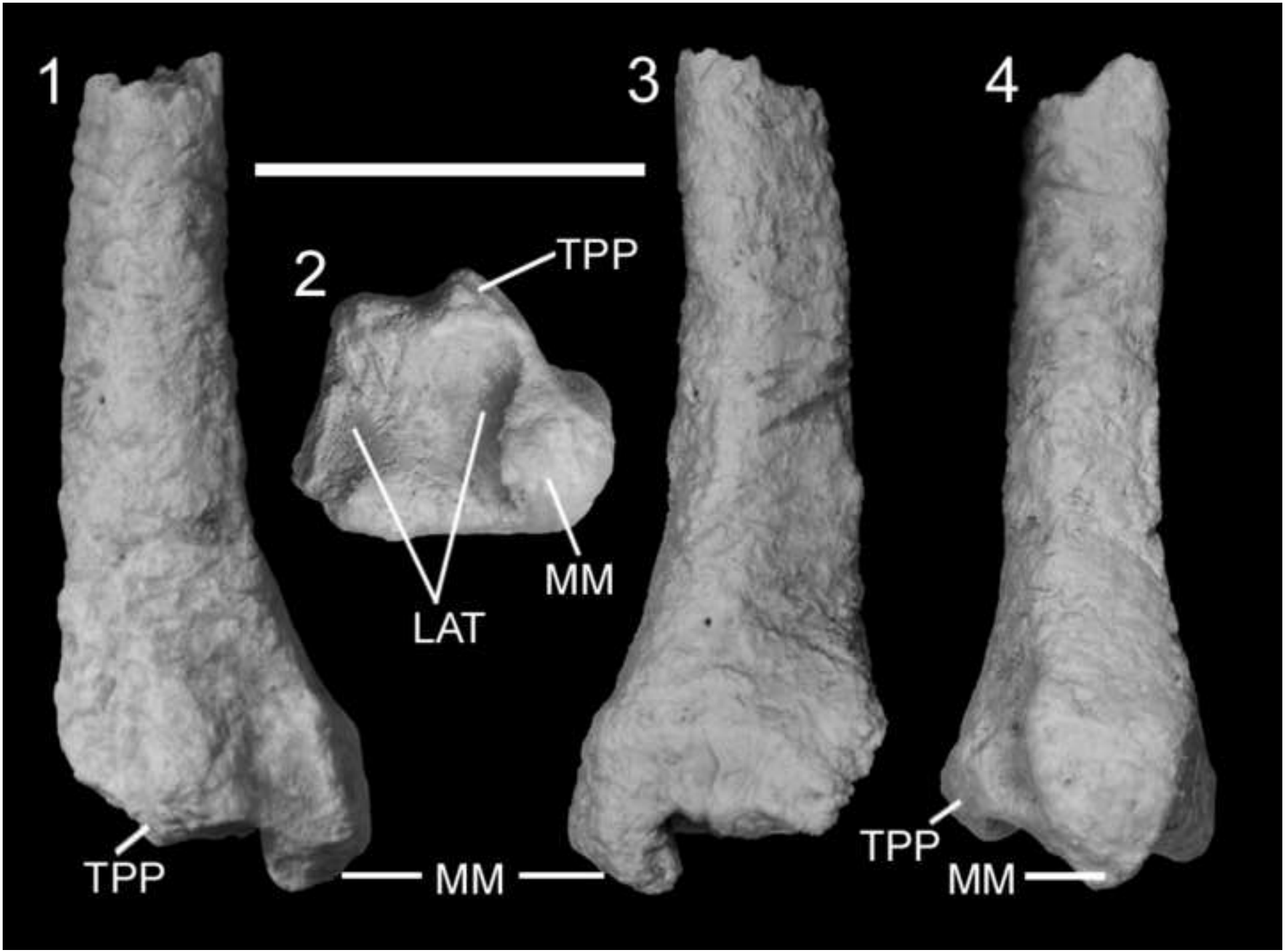


Figure 5
Click here to download high resolution image

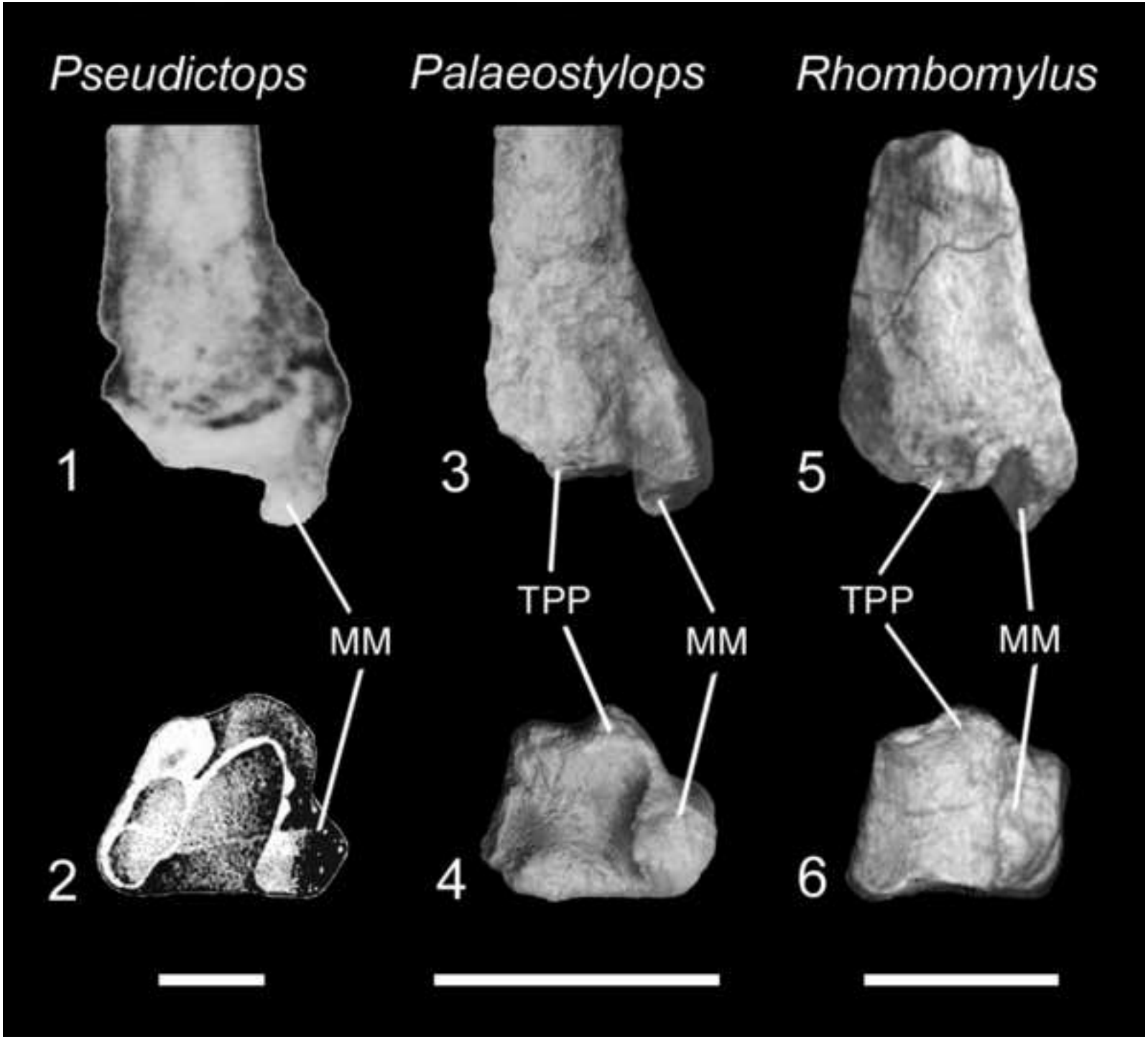


Figure 6
Click here to download high resolution image

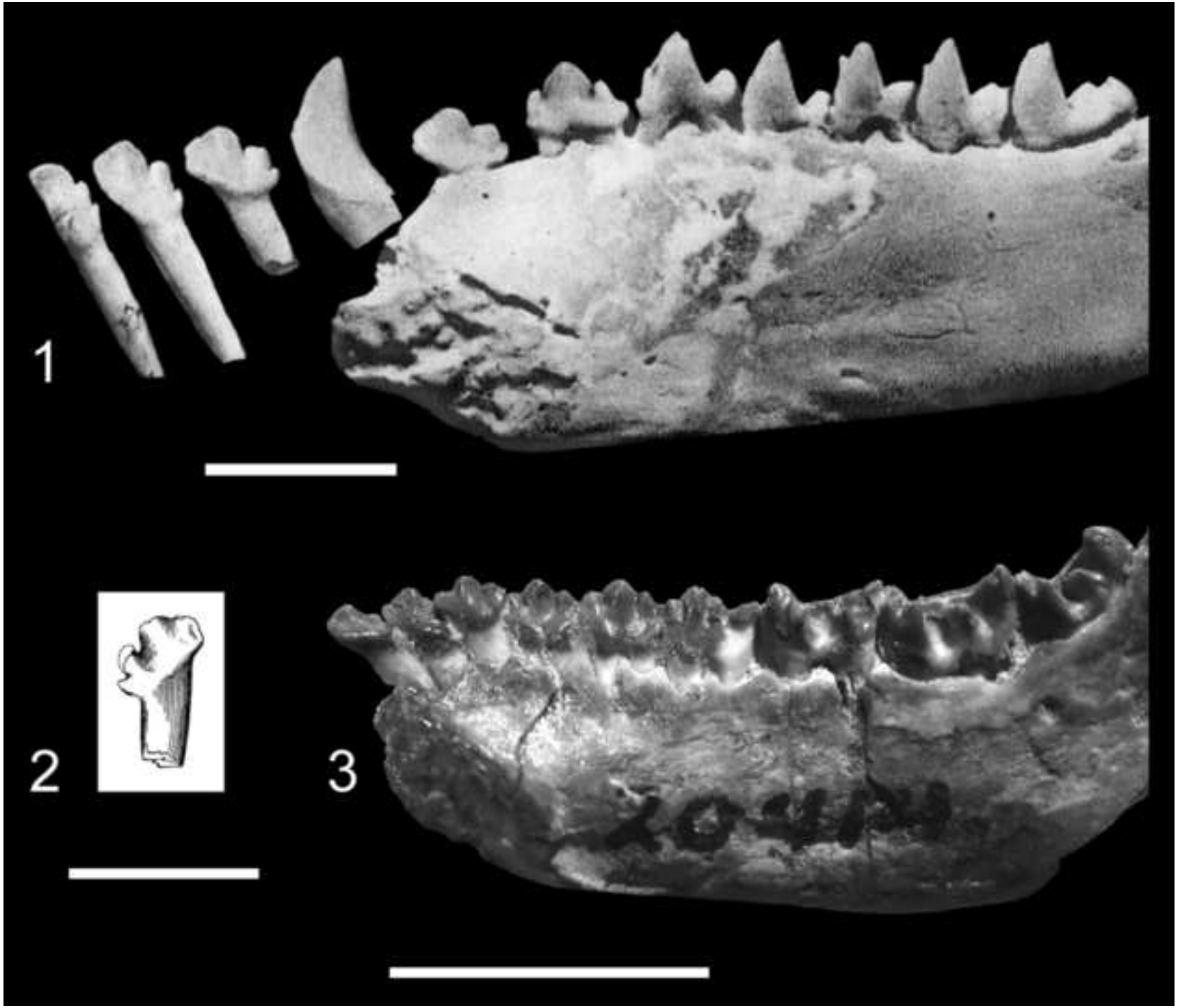
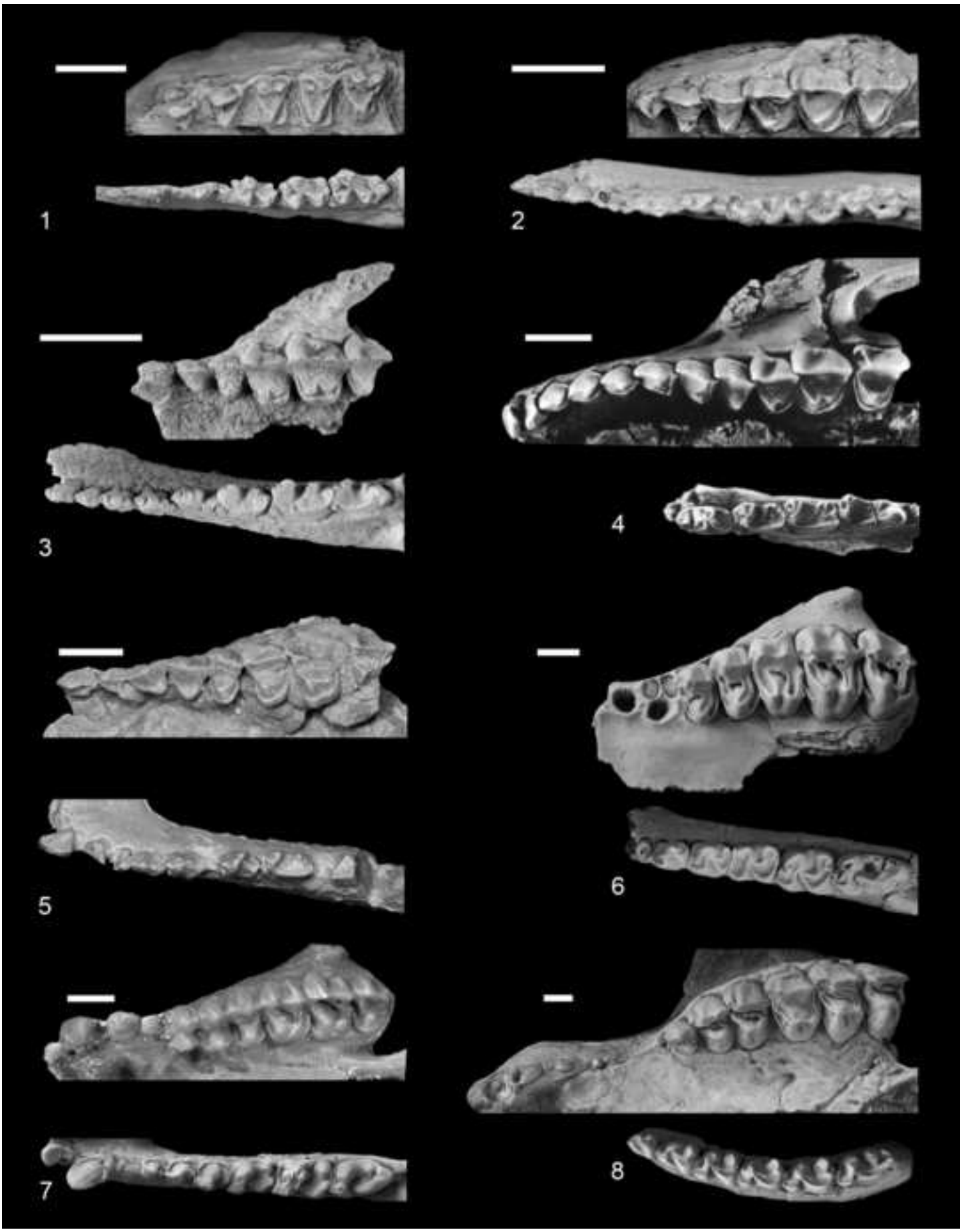


Figure 7
Click here to download high resolution image



Appendix A. Statistical analysis of Simpson's (1941) Log-ratios (G. Escarguel).

Biometric differences between groups of individuals always combine isometric and allometric variations, only the later involving proportional differences corresponding to shape differences. A classical way to untangle isometric and allometric components within a biometric dataset consists in standardizing each studied sample by a given, homogeneous and usually large reference sample, e.g., by calculating Simpson's Log-ratios (Simpson, 1941; Simpson et al. 1960). By allowing between-group comparisons in terms of proportions, i.e., in terms of relative and *not* absolute differences, the use of Log-ratios makes possible the separate analysis of isometric and allometric *relative* differences between groups (Meadow, 1999). From a strictly biological point of view, such a size-scaling approach is justified by the fact that the living world does not evolve in an *arithmetic* (i.e., additive) but in a *geometric* (i.e., multiplicative) space: any comparison between organs' or organisms' size and shape must be done in terms of proportions – nature has nothing to do with meters, liters or grams: only proportional relations between structures are of interest, whatever the quality, and thus measurement units of these structures (Gingerich, 2000).

For any analyzed sample *i* and biometric descriptor *j*, Simpson's Log-ratio (= "Log-Size Index" *sensu* Meadow, 1999) is

$$SR_{i,j} = \ln\left(\frac{\overline{X_{i,j}}}{\overline{R_j}}\right) = \ln\left(\overline{X_{i,j}}\right) - \ln\left(\overline{R_j}\right),$$

where $\overline{X_{i,j}}$ is the empirical mean value for sample *i* and descriptor *j*, and $\overline{R_j}$ is the empirical mean value of the reference sample for descriptor *j*. By definition, *SR* is a non-dimensional index, but it still depends on the unit of the measured descriptors[1] (the logarithm of a ratio, i.e., a difference of logarithms keeps the same dimensional proportionality as a standard deviation; Lande, 1977; Gingerich, 2001).

Once calculated for commensurate (i.e., same-unit) biometric descriptors of different samples (always using the same reference sample), Simpson's Log-ratio allows multivariate comparisons of proportional differences between samples. Classically, such comparisons are done graphically, through the drawing of a "Simpson diagram" where each proportional difference between empirical means is plotted for each analyzed sample and descriptor. Such diagrams are widespread in the paleontological literature, but as far as we know, little or no

---

[1] The following example illustrates this simple, while rather counter-intuitive fact. Let's *A* and *B* be two cubes of line lengths *L*(*A*) and *L*(*B*) m, surfaces *S*(*A*) and *S*(*B*) m^2, and volumes *V*(*A*) and *V*(*B*) m^3. As $V(\cdot) = L(\cdot)^3$ and $S(\cdot) = 6.L(\cdot)^2$, their volume and surface logarithmic differences are:

$$SR = \ln\left(\frac{V(A)}{V(B)}\right) = \ln\left(V(A)\right) - \ln\left(V(B)\right) = \ln\left(L(A)^3\right) - \ln\left(L(B)^3\right) = 3\cdot\left(\ln\left(L(A)\right) - \ln\left(L(B)\right)\right)$$

and

$$SR = \ln\left(\frac{S(A)}{S(B)}\right) = \ln\left(S(A)\right) - \ln\left(S(B)\right) = \ln\left(6\cdot L(A)^2\right) - \ln\left(6\cdot L(B)^2\right)$$
$$= \left[\ln(6) + 2\cdot\ln\left(L(A)\right)\right] - \left[\ln(6) + 2\cdot\ln\left(L(B)\right)\right] = 2\cdot\left(\ln\left(L(A)\right) - \ln\left(L(B)\right)\right).$$

Thus, the Log-ratio of the volume of two cubes is 1.5 times larger than the Log-ratio of their associated surfaces, and 3 times larger than the Log-ratio of their associated line lengths: even if non-dimensional, *SR* still depends on the units of the measured descriptors (here, m, m^2 or m^3). As a consequence, when measured with Simpson's Log-ratios, only proportional differences of descriptors with the same unit of measurement are commensurate and can be directly compared.

attention has yet been paid to the statistical (descriptive and inferential) issues underlying such comparisons. On the one hand, when the number of studied samples becomes large, multi-sample multivariate comparisons can be made easier (and less subjective than direct graphical comparisons of Simpson diagrams) using usual hierarchical or nonhierarchical clustering and/or metric or nonmetric ordination techniques as available in many statistical books and computational packages (e.g., UPGMA, Neighbor-Joining, *k*-mean, PCA/PCoA, NM-MDS, etc.). On the other hand, as for any sample value, confidence intervals are obviously associated with each empirical mean proportional difference. These intervals, which directly depend on the confidence intervals associated with the sample means $\overline{X_{i,j}}$ and $\overline{R_j}$, must be estimated and taken into account in order to test the observed differences between samples for statistical significance.

Estimates of Simpson's Log-ratio confidence intervals can readily be achieved using parametric bootstrap (Efron and Tibshirani, 1993; see Appendix B of the present paper for a short summary of the core operational concept underlying the Bootstrap theory). For each descriptor, *B* pseudo-values (indicated hereafter by the "¤" symbol) of the mean of the studied and reference samples are randomly and independently generated, based on their sample Gaussian distributions, following:

$$X_{i,j}^{¤} \sim N\left(\overline{X_{i,j}}, \frac{s_{X_{i,j}}}{\sqrt{n_{X_{i,j}}}}\right) \text{ and } R_j^{¤} \sim N\left(\overline{R_j}, \frac{s_{R_j}}{\sqrt{n_{R_j}}}\right),$$

where *s* is the sample standard deviation, *n* is the number of observed specimens (sample size), and $N\left(m, \frac{s}{\sqrt{n}}\right)$ is the normal distribution with mean *m* and standard deviation $s/\sqrt{n}$, the standard error of the mean *m*. The *B* $\left(X_{i,j}^{¤}, R_j^{¤}\right)$ couples then allows the computation of *B* pseudo-values $SR_{i,j}^{¤} = \ln\left(X_{i,j}^{¤}\right) - \ln\left(R_j^{¤}\right)$, forming together a bootstrap distribution of pseudo-values (*B* = 10,000 in this work, including the observed *SR*-value), from which bootstrap mean and standard deviation, as well as nonparametric bilateral (1 – α)% confidence interval limits (i.e., the $\alpha/2^{th}$ and $(1 - \alpha/2)^{th}$ percentiles of the cumulated distribution functions) can be extracted.

Following the same principle, the confidence interval associated with the null hypothesis (H_0) that *the studied and reference samples share the same descriptor's reference distribution (mean and standard deviation)*, can be estimated, for each descriptor *j*, by randomly and independently generating the pseudo-values

$$X_{i,j}^{¤} \sim N\left(\overline{R_j}, \frac{s_{R_j}}{\sqrt{n_{X_{i,j}}}}\right) \text{ and } R_j^{¤} \sim N\left(\overline{R_j}, \frac{s_{R_j}}{\sqrt{n_{R_j}}}\right).$$

Note that the standard deviation of the normal distribution underlying the *B* $X_{i,j}^{¤}$-pseudo-values is calculated using the actual size of sample *i* ($n_{X_{i,j}}$) and *not* the size of the reference sample (n_{R_j}). The *B* $\left(X_{i,j}^{¤}, R_j^{¤}\right)$ couples then allows the computation of *B* pseudo-values $SR_{i,j}^{H_0}$, which form an unbiased estimate ($\overline{SR}_{i;j}^{H_0} \cong 0$) of the confidence interval associated with H_0.

Finally, various univariate and multivariate parametrical tests [see Sokal & Rohlf (1995), Legendre & Legendre (1998) and Zar (1998) for comprehensive descriptions and full computational details] can be achieved based on these mean and associated variance bootstrap estimates (provided the bootstrapped *SR*-distributions are actually normally distributed):

- a **1-sample Student's *t*-test** can be done in order to test the studied sample *SR*-value against any expected *SR*-value, including 0 (corresponding to the null hypothesis that the study sample comes from a population with *SR* = 0 for the descriptor of interest). Another closely related, but less stringent null hypothesis of interest is that, based on the studied sample, the observed (empirical) *SR*-value does not differ from the bootstrapped null distribution (corresponding to the null hypothesis that the studied and reference samples share the same reference distribution [mean and standard deviation] for the descriptor of interest);

- a **2-sample Student's or Welch's *t*-test** can be done in order to compare two samples standardized with the same reference sample (tested null hypothesis: for the descriptor of interest, the two samples come from populations with similar *SR*-values). Welch's test (Welch, 1947; Sawilowsky, 2002) is favored over Student's one when the two compared samples significantly depart from homoscedasticity (= homogeneity of variances). In order to select Student's or Welch's tests, the null hypothesis of homoscedasticity is first tested with Fisher's and Bartlett's (1937) tests;

- a **one-way Analysis of Variance (ANOVA)** can be done in order to simultaneously compare several samples (= groups) standardized with the same reference sample (tested null hypothesis: the studied samples share a single *SR*-value for the descriptor of interest). Overall homogeneity of variances is first tested with Bartlett's (1937) test; Welch's (1951) unequal-variance ANOVA must be favored over "classical" (equal-variance) ANOVA when the compared samples significantly depart from homoscedasticity (leading to an over-liberal *F*-based result). A significant ANOVA result then legitimates the computation of "post-hoc" pairwise comparisons (contrast analysis), e.g., using Tukey's HSD test.
  In addition to the usual *F*-statistic, it can be useful here to compute the effect-size ω^2-statistic, a measure of the proportion of the total variability explained by between-group differences (with large samples, a highly significant ANOVA result can be reached even if the studied samples largely overlap…), such as:

  $$\omega^2 = \frac{SS_{Between} - \left(df_{Between} \times MS_{Within}\right)}{SS_{Total} + MS_{Within}}$$

  (ω^2 is an unbiased version of the simpler and more classic, but biased eta²-statistic $\eta^2 = \frac{SS_{Between}}{SS_{Total}}$). ω^2 ranges between < 0 and 1; the closer to 1, the more different the sample means relative to the overall variability, implying that the between-group variability is large with respect to the overall within-group variability;

- Provided that for each studied sample, each available specimen is known for every analyzed descriptors, a **multiple Analysis of Variance (MANOVA)** – eventually coupled with a **Canonical Variate Analysis (CVA)**, i.e., a **multi-group Discriminant Analysis** – can be achieved in order to simultaneously compare several multivariate samples standardized with the same reference sample (tested null hypothesis: the studied samples share the same *SR*-values for the descriptors under analysis). As for a one-way ANOVA, a significant MANOVA result (based on, e.g., Wilk's λ or the [more robust] Pillai trace) then legitimates the computation of "post-hoc" pairwise comparisons (contrast analysis), e.g., using Hotelling's T^2-test, preferably corrected for multiple testing in order to control for the increase in (type I) experimentwise error rate.

<<<<<>>>>>

In order to help readers to go through these computations, we have designed a **user-friendly 5-sheet Excel file** (done with Excel 2003) called "**Simpson Ratio Confidence Interval.xls**" (available with this Appendix or on simple request to G. Escarguel [gilles.escarguel@univ-lyon1.fr]) allowing the following "one-descriptor" computations do be automatically done (see below, the screen-copies of the five sheets):

- **"Individual SR-values":** this sheet calculates the individual *SR*-values for each specimen of a studied sample (up to 50 specimens per run), based on the **detailed sampled measures and on the empirical mean of the reference sample** (the number of specimens and standard deviation of the reference sample are not useful here). When all individual values are known for one or more analyzed descriptor and for each studied sample, such individual *SR*-values allow for the direct statistical comparison between samples through univariate (Student, Mann-Whitney, Kolmogorov-Smirnov, ANOVA, Kruskal-Wallis) or multivariate (MANOVA) parametric and nonparametric procedures as available in most standard statistical softwares (e.g., PAST).

- **"Confidence intervals":** this sheet calculates parametric bootstrap estimates (10,000 iterations) of a studied sample confidence interval and null-hypothesis confidence interval based on the **size (number of measured specimens) and empirical mean and standard deviation of the studied and reference samples**. Computations include the bootstrapped mean *SR*-value and its associated standard error, standard deviation, variance, skewness and kurtosis, as well as the bootstrapped median *SR*-value and associated 90%, 95% and 99% "nonparametric" bilateral confidence intervals. Normality of the two bootstrapped distributions can be evaluated (even if not formally tested here) by direct comparison of the skewness (g_1; coefficient of asymmetry) and kurtosis (g_2; coefficient of "peakedness") metrics against their respective 95% confidence intervals under the null hypothesis that $g_1 = 0$ and $g_2 = 0$ (expected values for a normal distribution, given within brackets on the right). In addition, percent absolute deviation between "parametric" $\pm 1.645s^{¤}$ (90%), $\pm 1.96s^{¤}$ (95%) and $\pm 2.58s^{¤}$ (99%), and "nonparametric" 90%, 95% and 99% bilateral C.I. are computed (the lower these percentages, the closer the bootstrapped distributions to normality). Statistically null g_1 and g_2 values, as well as very-low $\%_{A.D.}$ are required for further parametric testing based on bootstrapped mean and variance values.

- **"1-Sample t-test":** this sheet automatically calculates two Student's *t*-tests linked to two closely related, but distinct null hypothesis. Based on the studied sample, the left test contrasts the bootstrapped mean *SR*-value to any given expected (parametric) *SR*-value, including 0 (➔ test of nullity of the empirical *SR*-value), while the right test contrasts the bootstrapped null distribution to the observed (empirical) *SR*-value. All data values in this sheet are directly imported from the "Confidence intervals" sheet, excepted the **expected *SR*-value** against which the studied sample is to be compared (left test; default setting: 0 ➔ test of nullity of the empirical *SR*-value).

- **"2-Sample t-test":** this sheet allows the comparison of two samples for a given biometric descriptor, using a Student's (equal variance) or Welch's (unequal variances) *t*-test (Fisher's and Bartlett's tests for homogeneity of variances are also provided), based on the **size (number of measured specimens) and bootstrapped *SR*-mean and variance (*not* standard deviation) of the two samples**.

- **"multi-Sample ANOVA":** this sheet allows the computation of a one-way ANOVA, including a preliminary Bartlett's test for homogeneity of variances. Equal-variance ANOVA (*F*-statistic), unequal-variance ANOVA (Welch's *F**-statistic), and effect-size ω^2-statistics are computed, based on the **size (number of measured specimens) and bootstrapped *SR*-mean and variance (*not* standard deviation) of $k < 25$ samples**.

Technical caveats:

- depending on your operating system settings, the **decimal mark** can be either the period (“.”) or the comma (“,”); no space or symbol is expected for the thousands separator;
- in all 5 sheets, **modifiable cells** to be informed are indicated by a gray (optional) or black (obligatory) star (“*”). ***All other cells are protected and cannot be modified***;
- **data** must be entered first, then press the <**F9**> function-key to **run the computations** (please wait: bootstrap computations can take a few seconds, depending on the speed of your computer);
- **results** can be freely copied/pasted on another sheet, and/or printed following a pre-formatted one-page model for each sheet;
- in the last 3 sheets (t-tests and ANOVA), ***p*-values in scientific notation** are also provided in italics within grey cells (useful for very small *p*-values);

References

Bartlett, M. S., 1937. Properties of sufficiency and statistical tests. Proceedings of the Royal Statistical Society Series A 160, 268-282.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman & Hall, London.

Gingerich, P.D., 2000. Arithmetic or geometric normality of biological variation: an empirical test of theory. Journal of Theoretical Biology 204, 201-221.

Gingerich, P.D., 2001. Rates of evolution on the time scale of the evolutionary process. Genetica 112-113, 127-144.

Lande, R., 1977. On comparing Coefficients of Variation. Systematic Zoology 26, 214-217.

Legendre P. and Legendre L., 1998. Numerical Ecology (2nd English Edition). Developments in Environmental Modelling, 20, Elsevier, Amsterdam.

Meadow, R., 1999. The use of size index scaling techniques for research on archaeozoological collections from the Middle East. In: Becker G., Manhart H., Peters J. & Schibler J. (edits.), Historia Animalium ex Ossibus, Festschrift für Angela von den Driesch. Marie Leidorf, Rahden/Westfahlen, 285-300.

Simpson, G.G., 1941. Explanation of Ratio Diagrams. In: Simpson G.G. (edit.), Large Pleistocene Felines of North America. American Museum Novitates 1136, 1-27.

Simpson, G.G., Roe, A., Lewontin, R.C., 1960. Quantitative Zoology. Harcourt & Brace, New-York.

Sokal, R. R., and Rohlf, F. J. (1995). Biometry (3rd Edition). W. H. Freeman and Co., New York.

Sawilowsky, S.S., 2002. Fermat, Schubert, Einstein, and Behrens–Fisher: The Probable Difference Between Two Means When $\sigma 1 \neq \sigma 2$. Journal of Modern Applied Statistical Methods 1(2), 461-472.

Welch, B.L., 1947. The generalization of "Student's" problem when several different population variances are involved. Biometrika 34(1-2), 28-35.

Welch, B.L., 1951. On comparison of several mean values: an alternative approach. Biometrika 38(3/4), 330-336.

Zar, J.H., 1998. Biostatistical analysis, 4th edition. Prentice Hall, Upper Saddle River, New Jersey.

<<<<<>>>>>

Screen-copy for the sheet "Individual SR-values"

SIMPSON'S LOG-RATIO (*SR*)
INDIVIDUAL *SR*-VALUES

1. SAMPLE DATA (measured descriptor)

Data information:	M2-Length P. macrodon (ref.: P. iturus)

Studied sample (calculated)	
Nb. of specimens	21
Mean	5,01
Standard deviation	0,290

Reference Sample	
Nb. of specimens	58
Mean	3,25
Standard deviation	0,245

2. Studied sample (detailed values)

#	Measured specimens	Value
1	1	5,1
2	4	4,7
3	15	4,8
4	16	5,1
5	20	5,2
6	24	4,6
7	27	4,9
8	32	5,1
9	36	4,4
10	43	4,7
11	48	4,9
12	50	5,2
13	51	5,4
14	150-1	5,3
15	150-2	5,4
16	160-1	5,5
17	160-2	5,1
18	174	4,7
19	194	5,0
20	204	4,9
21	222	5,2
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		

3. CALCULATE: PRESS <F9>

Measured specimens	*SR*-Value
1	0,4506
4	0,3689
15	0,3900
16	0,4506
20	0,4700
24	0,3474
27	0,4106
32	0,4506
36	0,3029
43	0,3689
48	0,4106
50	0,4700
51	0,5077
150-1	0,4891
150-2	0,5077
160-1	0,5261
160-2	0,4506
174	0,3689
194	0,4308
204	0,4106
222	0,4700

G.E.-10/2011

Screen-copy for the sheet "Confidence intervals"

SIMPSON'S LOG-RATIO (*SR*)
BOOTSTRAPPED CONFIDENCE INTERVALS

1. SAMPLE DATA (measured descriptor)

Data information:	M2-Length P. macrodon (ref.: P. iturus)	*

Studied sample			Reference Sample		
Nb. of specimens	21	*	Nb. of specimens	58	*
Mean	5,01	*	Mean	3,25	*
Standard deviation	0,290	*	Standard deviation	0,245	*

2. CALCULATE: PRESS <F9>

3. RESULTS

Observed *SR* = 0,433

Bootstrap Confidence interval for the studied sample		Bootstrap Confidence interval for the null distribution§		
Mean *SR*-value	0,433	Mean *SR*-value	0,000	(expected: 0.0)
Standard error of the mean	0,0159	Standard error of the mean	0,0191	
Standard deviation	0,0731	Standard deviation	0,0876	
Variance	0,00534	Variance	0,00767	
Skewness	0,0195	Skewness	-0,0144	[-0.047, +0.047]
Kurtosis	0,0347	Kurtosis	-0,0120	[-0.096, +0.096]
99% C.I. - lower limit	0,392	99% C.I. - lower limit	-0,050	
95% C.I. - lower limit	0,402	95% C.I. - lower limit	-0,038	
90% C.I. - lower limit	0,407	90% C.I. - lower limit	-0,031	
Median *SR*-value	0,433	Median *SR*-value	0,000	(expected: 0.0)
90% C.I. - upper limit	0,459	90% C.I. - upper limit	0,032	
95% C.I. - upper limit	0,465	95% C.I. - upper limit	0,037	
99% C.I. - upper limit	0,474	99% C.I. - upper limit	0,048	

Percent absolute deviation between bootstrapped "parametric" and "nonparametric" C.I.

Studied sample		Null distribution	
99% C.I. - lower limit	0,076%	99% C.I. - lower limit	0,674%
99% C.I. - upper limit	0,026%	99% C.I. - upper limit	2,645%
95% C.I. - lower limit	0,000%	95% C.I. - lower limit	1,829%
95% C.I. - upper limit	0,172%	95% C.I. - upper limit	0,101%
90% C.I. - lower limit	0,092%	90% C.I. - lower limit	0,306%
90% C.I. - upper limit	0,002%	90% C.I. - upper limit	1,371%
Mean % absolute deviation	**0,062%**	Mean % absolute deviation	**1,154%**

§ *Null hypothesis: no difference between the studied and reference samples*

Parametric bootstrap estimates; 10,000 iterations

Boot. 99% C.I.: 0.005 (lower) & 0.995 (upper) percentiles
Boot. 95% C.I.: 0.025 (lower) & 0.975 (upper) percentiles
Boot. 90% C.I.: 0.050 (lower) & 0.950 (upper) percentiles

G.E.-10/2011

Screen-copy for the sheet "1-Sample t-test"

SIMPSON'S LOG-RATIO (*SR*)
1-SAMPLE STUDENT'S *t*-TEST

1. SAMPLE DATA

Data information:	M2-Length P. macrodon (ref.: P. iturus)

Studied sample		Reference Sample		
Nb. of specimens	21	Nb. of specimens	58	*Values from the CI sheet*
Desc. mean	5,01	Desc. mean	3,25	*Values from the CI sheet*
Desc. standard deviation	0,290	Desc. standard deviation	0,245	*Values from the CI sheet*

Bootstrapped C.I. for the studied sample		Bootstrapped C.I. for the null distribution		
Bootstrapped mean *SR*	0,433	Bootstrapped mean *SR*	0,000	*Values from the CI sheet*
Boot. variance *SR*	0,00534	Boot. variance *SR*	0,00767	*Values from the CI sheet*

SR-value against which the studied sample is tested			SR-value against which the null distribution is tested		
Expected *SR*-value §	0,000	*	Observed *SR*-value	0,433	*Value from the CI sheet*

§ Default setting: 0 ==> test of nullity

2. CALCULATE: PRESS <F9>

3. STUDENT'S *t*-TESTS

Studied sample *vs*. expected *SR*	Observed *SR* *vs*. null distribution
t = 5,926	*t* = 4,942
d.f. = 20	*d.f.* = 20
***p* = 0,00001**	***p* = 0,00008**
8,5E-06	*7,9E-05*

Tested null hypothesis (bilateral):

H_0: the studied sample comes from a population with the given *SR*-value	H_0: the studied and reference samples share the same reference distribution (mean & std. deviation)

G.E.-10/2011

Screen-copy for the sheet "2-Sample t-test"

SIMPSON'S LOG-RATIO (*SR*)
2-SAMPLE STUDENT'S AND WELCH'S *t*-TESTS

1. SAMPLE DATA

Data information:	Two virtual samples	*

Studied sample #1			Studied sample #2		
Nb. of specimens	13	*	Nb. of specimens	35	*
Bootstrapped mean *SR*	0,652	*	Bootstrapped mean *SR*	0,372	*
Boot. variance *SR*	0,2456	*	Boot. variance *SR*	0,0754	*

2. CALCULATE: PRESS <F9>

3. FISHER'S- and BARTLETT'S TESTS for homogeneity of variances

Fisher's *F* = 3,257	Bartlett's *B* = 6,918
***p* = 0,00666**	***p* = 0,00853**

CAUTION:

If Fisher's & Bartlett's *p*-values small (< ~0.05; reject homoscedasticity), then Student's test is over-liberal ==> favor Welch's test

4. STUDENT'S and WELCH'S *t*-TESTS

STUDENT:	*t* =	2,491	
	d.f. =	46	
	***p* =**	**0,01642**	*1,6E-02*
WELCH:	*t* =	1,930	
	d.f. =	14,8	
	***p* =**	**0,07412**	*7,4E-02*

Tested null hypothesis (bilateral):

H_0: the two samples come from populations with similar *SR*-values

G.E.-10/2011

Comment: in this virtual example, the two samples significantly depart from homoscedasticity (i.e., their two bootstrapped *SR*-variances significantly differ). In that case, Welch's test result (H_0 not rejected at the 95% confidence level) must be favored over Student's test result (which suggests rejecting H_0 at the 95% confidence level).

Screen-copy for the sheet "multi-Sample ANOVA"

SIMPSON'S LOG-RATIO (*SR*)
MULTI-SAMPLE ANALYSIS OF VARIANCE

1. SAMPLE DATA (max: 25 samples)

Data information: Five virtual samples

Sample #	Sample name	Nb. of specimens	Bootstrapped mean *SR*	Bootstrapped variance *SR*	*Data checking*
1	Aaaaaa	8	0,213	0,5120	*Data OK*
2	Bbbbbb	13	0,456	0,1076	*Data OK*
3	Cccccc	26	0,438	0,0759	*Data OK*
4	Dddddd	15	0,189	0,2841	*Data OK*
5	Eeeeee	18	0,631	0,1055	*Data OK*
6					*No Data*
7					*No Data*
8					*No Data*
9					*No Data*
10					*No Data*
11					*No Data*
12					*No Data*
13					*No Data*
14					*No Data*
15					*No Data*
16					*No Data*
17					*No Data*
18					*No Data*
19					*No Data*
20					*No Data*
21					*No Data*
22					*No Data*
23					*No Data*
24					*No Data*
25					*No Data*

2. CALCULATE: PRESS <F9>

3. BARTLETT'S TEST for homogeneity of variances

B **= 17,082**	d.f. = 4	***p*** **=** 0,00186

CAUTION:
If Bartlett's *p*-value small (< ~0.05; reject homoscedasticity), then "classical" (equal-variance) ANOVA is over-liberal ==> favor Welch's (unequal-variance) ANOVA

4. ANALYSIS OF VARIANCE

Number of studied samples: *k* = 5
Number of studied specimens: *N* = 80
Grand mean: M = 0,415
Proportion of variance explained: ω² = 0,088

ANOVA table

Source	S.S.	d.f.	M.S.	***F***
Between samples	1,968	4	0,4920	**2,942**
Within samples	12,544	75	0,1672	
Total	14,512	79	0,1837	

Significance (*F*): *p* = 0,02575 *2,6E-02*

Unequal-variance ANOVA using Welch's correction

Source	"M.S."	d.f.	***F****
Between samples	2,4937	4	**2,320**
Within samples	1,0748	26,8	

Significance (*F): *p* = 0,08356** *8,4E-02*

Tested null hypothesis: H_0: the studied samples share a single SR-value

G.E.-10/2011

Comment: in this virtual example, the five samples significantly depart from homoscedasticity (i.e., at least one sample shows a bootstrapped *SR*-variance that significantly differs from the others). In that case, Welch's unequal-variance result (H_0 not rejected at the 95% confidence level) must be favored over the "classical" equal-variance ANOVA result (which suggests rejecting H_0 at the 95% confidence level).

Appendix B. Biostatistical analysis of dental measurements: Computational details regarding *V**, *b* and *D* metrics, and mixture analysis results (G. Escarguel)

In order to better characterize the biometrical homogeneity, and thus possible taxonomic status of the studied fossil assemblage, we computed various complementary metrics focusing on distinct aspects of the sample distributions of three dental measurements: length (L), width (W) and ln(L × W) of the P3/p3 to M3/m3 of all measured *Palaeostylops* teeth, and measured teeth a priori assigned to *P. iturus* or to *P. macrodon* (Table S1). The two first metrics focus on the relative variability and shape of the sample distributions, whereas the last two techniques aim at estimating the dimorphism ratio involved by the available data. Ratios estimated for the "all-*Palaeostylops*" samples (first column of results in Table S1) can be directly compared to the expected ratios directly calculated from measured teeth a priori (and independently) assigned to *P. iturus* or *P. macrodon*.

Unbiased coefficient of variation (*V**)

The unbiased coefficient of variation,

$$V^* = \left(1 + \frac{1}{4N}\right)\frac{100s}{\bar{x}}$$

where N is the number of sampled specimens (sample size), and $\bar{x}$ and s are the sample mean and standard deviation, i.e., the first and second moments of the studied distribution, respectively. V^* is basically a standard deviation expressed as a percentage of the mean; it accounts for the relative amount of variation in a population, thus allowing for direct comparisons of variability between samples with different empirical means (Sokal & Rohlf, 1995).

Over the last decades, several studies (e.g., Simpson et al., 1960; Gingerich, 1974, 1981; Plavcan & Cope, 2001) have shown that empirical values for cranio-dental and post-cranial linear dimensions in large homogenous samples of well-defined extant mammal species usually range between 4 and 10, even when sexual dimorphism is observed. Nevertheless, noteworthy counter-examples do exist, with values significantly larger than 10 even in the absence of strong sexual dimorphism (e.g., Polly, 1997, Plavcan & Cope, 2001). Thus, if a *large V*-value can usually be considered as a rough indicator of sample heterogeneity (due to, e.g., intra-specific sexual dimorphism or to multi-species mixing), this simple metric cannot be used alone in order to safely characterize the taxonomic homogeneity and/or dimorphism status of a given biometric sample (Kelley & Plavcan, 1998).

Bimodality index (*b*)

Complementary to the coefficient of variation, which focuses on the relative amount of inter-individual variability regardless of the shape of the underlying sample distribution, the bimodality index *b* focuses on the shape of the sample distribution – regardless on the relative amount of inter-individual variability. This coefficient relies on the computation of the third and fourth standardized moments about the mean of the studied distribution, i.e., its skewness (M_3) and excess kurtosis (M_4), following:

$$b = \frac{M_3^2 + 1}{M_4 + \frac{3(N-1)^2}{(N-2)(N-3)}},$$

where

$$M_3 = \frac{N\left(\sum_{i=1}^{i=N}\left(x_i - \bar{x}\right)^3\right)}{(N-1)(N-2)\left(\sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\left(x_i - \bar{x}\right)^2}\right)^3}$$

and

$$M_4 = \frac{N(N+1)\left(\sum_{i=1}^{i=N}\left(x_i - \bar{x}\right)^4\right)}{(N-1)(N-2)(N-3)\left(\sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\left(x_i - \bar{x}\right)^2}\right)^4} - \frac{3(N-1)^2}{(N-2)(N-3)}$$

are the estimators of the population skewness and excess kurtosis, respectively, and $\bar{x}$ is the sample mean (Der & Everitt, 2002). The skewness metric quantifies the lack of symmetry of the distribution ($M_3 > 0$ indicates a right-tailed distribution; $M_3 < 0$ a left-tailed distribution), whereas the excess kurtosis metric mainly measures the "peakedness" of the distribution ($M_4 > 0$ indicates excess of very small and large deviations around the mean, leading to a distribution with an acute mode and fat tails [leptokurtic distribution]; $M_4 < 0$ indicates excess of medium deviations around the mean, leading to a distribution with a wide mode and thin tails [platykurtic distribution]).

The value of *b* ranges between less than zero and one. Expected values of *b* for some remarkable distributions are:

* **1/3** = 0.33$\underline{3}$ for a single normal distribution (where, by definition, $M_3 = M_4 = 0$);
* **5/9** = 0.55$\underline{5}$ for a continuous uniform distribution (where, by definition, $M_3 = 0$ and $M_4 = -6/5$);
* **1** for a Bernoulli distribution with $p = 1 - q = 0.5$ (where, by definition, $M_3 = 0$ and $M_4 = -2$, the lowest possible excess kurtosis, and thus most platykurtic distribution of all).

Thus, a *b*-value significantly lower than 5/9 may indicate unimodality, whereas $b > 5/9$ is likely to reflect a two (or more) group structure, and thus bimodality (or multi-modality) of the underlying distribution.

Dimorphism ratio of the "method-of-moments" technique (*D*)

In any given sample assemblage, large *V** and/or *b*-values can be the consequence of the mixing of two distinct groups with different mean (and standard deviation). Nevertheless, small *V** and/or *b*-values can be obtained even in actually dimorphic species when groups strongly overlap, i.e., when group-means only slightly differ relative to their associated standard deviation. In any cases, different methods are available in order to identify the mixed occurrence of two distinct groups, and then to estimate the ratio between the largest and smallest group means. Among these methods, the dimorphism ratio *D* of the "method-of-moments (MoM)" technique (Josephson et al., 1996) appears as one of the most reliable method, especially in conditions of low within-group coefficient of variation (~5%) and

relatively large sample size (~50 or more), which is frequently the case here (Kościński & Pietraszewski, 2004).

For any given biometric descriptor, based on an available pooled sample made of N measures with overall mean $\bar{x}$, variance s^2 and kurtosis m_4, the MoM technique makes use of the fact that the greater the difference between means of the two groups is, the lower the pooled-sample kurtosis (i.e., the more platykurtic the pooled distribution). From this premise, it estimates the proportional difference, noted δ, between the pooled-mean $\bar{x}$ and the largest group-mean, as:

$$\delta = \left(-\bar{x}^4 + \frac{3}{2} s^2 - \frac{1}{2} m_4 \right)^{1/4}$$

(negative values of δ are set to 0, making $\delta \geq 0$ by definition).

For any variable X, computation of δ is made easier by using the following 4-step protocol (Josephson et al., 1996):

1. Ln-transform the sample data: $Y = \mathrm{Ln}(X)$;
2. Standardize (i.e., center-reduce) Y: $U = \dfrac{Y - \bar{Y}}{s_Y}$;
3. Calculate the kurtosis of U: $m_4 = \dfrac{1}{N} \sum_{i=1}^{N} u_i^4$;
4. Calculate $\delta' = \left(\dfrac{3 - m_4}{2} \right)^{1/4}$, then $\delta = \delta' s_Y$.

This protocol makes obvious that δ is computed in Ln(X) (*not* X)-unit, and thus represents a proportional difference, i.e., a Ln-ratio between two quantities. Under the working hypothesis that the two groups within the pooled sample are equi-abundant (i.e., $\bar{x}$ is half-way the two group means), the dimorphism ratio D between them is finally estimated as $D = e^{2\delta}$. Nevertheless, D remains reasonably unbiased and accurate even with strongly unbalanced abundances (e.g., 80%/20%), provided the true underlying dimorphism ratio remains moderate (< ~1.2) (Josephson et al., 1996; Rehg & Leigh, 1999; Kościński & Pietraszewski, 2004).

Mixture analysis and associated Akaike information criterion-based evidence ratio

In unbalanced abundance conditions, most of the index-based methods available to estimate the dimorphism ratio between to groups – including truly nonparametric techniques such as Lee's (2001) Assigned Resampling Method – performs quite poorly (increasing bias and/or decreasing precision), all the more when dealing with large dimorphism ratio-values (> ~1.3) and/or large within-group variability (Kościński & Pietraszewski, 2004). In those cases, use of maximum-likelihood Mixture Analysis can be a useful complementary procedure in a model selection-based approach contrasting solutions with one or two normal distributions (Titterington et al., 1985; Harper, 1999). When the best model given the available data turns out to be an admixture of two groups, the ratio of their estimated means (corresponding to Meiri *et al.*'s [2005] Sexual Size Dimorphism [SSD] index) provides a direct estimate of the degree of dimorphism between the two identified groups.

In this work we used the Mixture Analysis routine implemented in PAST, v. 2.07 (Hammer et al., 2001). This routine allows the selection of the "best", i.e., the most likely

mixture of normal distributions given the available data, using the Akaike information criterion (Akaike, 1973, 1974) corrected for small-sample bias (AIC_c; Sugiura, 1978; Hurvich & Tsai, 1989) as the selection criterion, such as:

$$AIC_c = -2\ln(L) + 2k + \frac{2k(k+1)}{N-k-1} = -2\ln(L) + \frac{2kN}{N-k-1},$$

where N is the sample size, k is the number of parameters to be estimated in the mixture model, and L is the likelihood of the mixture model given the available data. In the 3-part definition (1st formula; the two first elements representing the [biased] *AIC* as originally defined by Akaike), the first term measures the lack of model fit to the observed data, the second term accounts for the complexity of the model, and the third term is a correction factor accounting for a negative sample-size bias of the two first ones. Such an approach combining model's power and parsimony has its foundation in Kullback-Leibler Information Theory: for large sample size N and relatively small K values, *AIC* is an approximately unbiased estimator of the expected Kullback discrepancy between the (unknown) *true* generating model (or, at least, a *true probabilistic characterization* of it) and the (fitted) approximating model M derived from the available sample (Akaike, 1973; Burnham & Anderson, 2002).

The mixture solution with the lowest AIC_c-value is preferred as the best fit without overfitting of the model to the empirical data (Johnson & Omland, 2004). In some cases, the mixture analysis failed due to the lack of convergence of the EM algorithm used by PAST toward a 2-group stable solution, suggesting that given the available data, the 1-group solution is more likely. For each mixture solution, the PAST routine also provides the mean, standard deviation and proportion of each normal distribution constituting the mixture model; maximum likelihood-based assignment of each data points to one of the groups is also possible.

Based on the AIC_c-values obtained for 1-group and 2-group solutions, selection of the simplest model that adequately accommodates the observed data can be achieved based on the observed difference (noted Δ^{AIC}) between the highest and the lowest AIC_c-values, corresponding to the poorest and best mixture model based on the available data, respectively. The Δ^{AIC}-metric, even if not a distance, is a convenient measure of how close two models are one each other given the available data.

In the simple case of the comparison between two models, M_i and M_j, the latter being the lowest-AIC_c model ($\Rightarrow \Delta_j^{AIC} = 0$ and $\exp\left(-\frac{\Delta_j^{AIC}}{2}\right) = 1$, where $\exp\left(-\frac{\Delta^{AIC}}{2}\right)$ is proportional to the relative likelihood of the compared models),

$$w_i = \frac{\exp\left(-\frac{\Delta_i^{AIC}}{2}\right)}{1+\exp\left(-\frac{\Delta_i^{AIC}}{2}\right)} \text{ and } w_j = \frac{1}{1+\exp\left(-\frac{\Delta_i^{AIC}}{2}\right)}$$

are the Akaike weights (i.e., normalized relative likelihoods) associated with models M_i and M_j (Burnham & Anderson 2002). Most particularly, w_i (< 1) measures the probability (the *risk*) that model M_i is more likely than the minimum *AIC* (or AIC^c) model M_j, whereas

$$w_j / w_i = \frac{1}{\exp\left(-\frac{\Delta_i^{AIC}}{2}\right)} = \exp\left(\frac{\Delta_i^{AIC}}{2}\right)$$

is the evidence ratio between models M_j and M_i, i.e., the relative support for the lowest-AIC_c model M_j over model M_i: the higher the w_j/w_i-value, the stronger the evidence for model M_j over model M_i.

Nonparametric Bootstrap estimates of the confidence intervals associated with sample *V*, b, D* and Ex.R.-values

As shown by Kościński & Pietraszewski (2004) in their simulation-based comparative analysis of various techniques to estimate dimorphism ratio, in most situations the error of estimation resulted mainly from sampling error, not from the error of method. We thus estimated statistical confidence intervals around the empirical (sample) values obtained for the *V**, *b* and *D*-metrics, and the expected dimorphism ration (Ex.R.), using nonparametric bootstrap, a computer-intensive technique based on random resampling with replacement of the available data (Efron & Tibshirani, 1993; Manly, 1997). When the mixture analysis favored a 2-group solution, the statistical confidence interval around the corresponding SSD-index was estimated using parametric bootstrap (Efron & Tibshirani, 1993) based on the mixture solution (estimated number of individuals, mean and standard error of the mean for each group).

The core operational concept of the Bootstrap theory is that, for any random variable with empirical value x estimating without bias the parametric value ζ, the distribution of x around ζ (which is unknown as far as its sampling error cannot be analytically calculated) can be estimated by the simulated distribution of "bootstrap" pseudo-values $x^{¤}$ around x. In the very same way x is calculated from a given sample S, a pseudo-values $x^{¤}$ is calculated:

- from a pseudo-sample $S^{¤}$ whose individual values are randomly drawn with replacement from S (nonparametric bootstrap);
- by random sorting within the parametric (e.g., Normal, Log-normal, Exponential…) distributions of the underlying computational elements of x, with distribution parameters directly estimated from S (parametric bootstrap).

In other words, the probability that $x^{¤} = x$ is *about* the same that the conditional probability, given x, that $x = \zeta$, which is to say, using the bootstrap standard deviation $s_X^{¤}$ ($z_{\alpha/2}$ being the centered-reduced normal deviation with probability $100[1 - {}^{\alpha}/_{2}]\%$):

$$x - z_{\alpha/2} s_X^{¤} < \zeta < x + z_{\alpha/2} s_X^{¤} \Leftrightarrow \zeta - z_{\alpha/2} s_x^{¤} < x < \zeta + z_{\alpha/2} s_x^{¤}.$$

Rather than calculating a bootstrap standard deviation from the resulting bootstrapped distribution (requiring that $x^{¤}$ is normally distributed around x, which is generally not the case when dealing with ratios or bounded quantities such as *V**, *b*, *D*, SSD and Ex.R.), we extracted the nonparametric confidence interval limits directly from the distribution of pseudo-values. In all cases, 100,000 bootstrap iterations were done, leading to a bootstrap distribution of 100,000 pseudo-values (including the original sample one), from which bilateral 95% confidence interval limits (i.e., the 2.5^{th} and 97.5^{th} percentiles of the cumulated distribution functions) were extracted for *V**, *b* and Ex.R., and the lower limit of the unilateral 95% confidence interval (i.e., the 5th percentile of the cumulated distribution function) was extracted for *D* and SSD. This later values (lower limits of the unilateral 95% confidence intervals for *D* and SSD) allows to test for significance (at the 5% significance level) the null hypothesis that $D = 1$ or SSD = 1 (absence of dimorphism) against the unilateral alternate hypothesis that $D > 1$ or SSD > 1 (involving the presence of a dimorphism).

References

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In: 2nd International Symposium on Information Theory, B. N. Petrov & F. Csaki (eds.), Akademia Kiado, Budapest, 267-281.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Akaike, H., 1983. Information measures and model selection. *In*: Proceedings of the 44th Session of the International Statistical Institute, 277-291.

Burnham, K. P., Anderson, D. R., 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer-Verlag, New York, USA.

Der, G., Everitt, B.S., 2002. A handbook of statistical analyses using SAS, 2nd ed. Chapman & Hall/CRC, Boca Ratón, CA.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman & Hall, London.

Gingerich, P.D., 1974. Size variability of the teeth in living mammals and the diagnosis of closely related sympatric fossil species. Journal of Paleontology 48, 895-903.

Gingerich, P.D., 1981. Variation, sexual dimorphism, and social structure in the early Eocene horse *Hyracotherium* (Mammalia, Perissodactyla). Paleobiology 7, 443-455.

Hammer, Ø., Harper, D.A.T., Ryan, P. D., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1), 9 p.

Harper, D.A.T., 1999. Numerical Palaeobiology. John Wiley & Sons, Chichester.

Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. Biometrika 76(2), 297-307.

Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19(2), 101-108.

Josephson, S.C., Juell, K.E., Rogers, A.R., 1996. Estimating sexual dimorphism by method-of-moments. American Journal of Physical Anthropology 100, 191-206.

Kelley, J., Plavcan, J.M., 1998. A simulation test of hominoid species number at Lufeng, China: implications for the use of the coefficient of variation in paleotaxonomy. Journal of Human Evolution 35, 577-596.

Kościński, K., Pietraszewski, S., 2004. Methods to estimate sexual dimorphism from unsexed samples: A test with computer-generated samples. Anthropological Review 67, 33-55.

Lee, S.-H., 2001. Assigned Resampling Method: A new method to estimate size sexual dimorphism in samples of unknown sex. Anthropological Review 64, 21-39.

Manly, B.F.J., 1997. Randomization, Bootstrap and Monte Carlo methods in biology (2nd Edition). Chapman & Hall/CRC, Boca Raton.

Meiri, S., Dayan, T., Simberloff, D., 2005. Variability and sexual size dimorphism in carnivores: testing the niche variation hypothesis. Ecology 86, 1432-1440.

Plavcan, J.M., Cope, D.A., 2001. Metric variation and species recognition in the fossil record. Evolutionary Anthropology 10, 204-222.

Polly, P.D., 1997. Ancestry and species definition in paleontology: a stratocladistic analysis of Paleocene-Eocene Viverravidae (Mammalia, Carnivora) from Wyoming. Contributions from the Museum of Paleontology, The University of Michigan 30(1), 1-53.

Rehg, J.A., Leigh, S.R., 1999. Estimating sexual dimorphism and size differences in the fossil record: a test of methods. American Journal of Physical Anthropology 110, 95-104.

Simpson, G.G., Roe, A., Lewontin, R.C., 1960. Quantitative Zoology. Harcourt & Brace, New-York.

Sokal, R. R., Rohlf, F. J. (1995). Biometry (3rd Edition). W. H. Freeman and Co., New York.

Sugiura, N., 1978. Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections. Communications in Statistics: Theory and Methods A7, 13-26.

Titterington, D., Smith, A., Makov, U., 1985. Statistical analysis of finite mixture distributions. John Wiley & Sons, Chichester, U.K.

Supplementary Table S1. Unbiased coefficient of variation (***V****), bimodality index (***b***), Dimorphism ratio of the "method-of-moments (MoM)" technique (***D***) and mixture analysis (**Mixt.**) results for the length (L), width (W) and Ln(L × W) of the upper and lower third premolars to third molars of all measured *Palaeostylops* teeth, and measured teeth a priori assigned to *P. iturus* or to *P. macrodon* (***N***: number of measured specimens). **Ex.R.**: expected ratio between the largest and smallest group means based on the a priori assignment of specimens to one of the two groups. Mom- and mixture analysis-based results indicating a 2-group structure of the analyzed sample are highlighted in red. See **Appendix B** for full computational details.

*V**, *b*, Ex.R.: sample value and 95% bilateral bootstrapped confidence intervals (within brackets; nonparametric bootstrap, 100,000 iterations). *D*: sample value and lower limit of the 95% unilateral bootstrapped confidence interval (within brackets; nonparametric bootstrap, 100,000 iterations). Mixt.: evidence ratio for the best supported (1 or 2 groups) mixture model, based on their respective Akaike criterion values; within parenthesis (when the 2-group solution is favored): ratio between the inferred largest and smallest group means and lower limit of the 95% unilateral bootstrapped confidence interval (within brackets; parametric bootstrap, 100,000 iterations). Some 2-group mixture analyses failed due to the lack of convergence toward a stable solution.

Upper teeth		*Palaeostylops*	*P. iturus*	*P. macrodon*
P3 - Length	*N*	49		
	*V**	8.46 [6.58 - 10.07]		
	b	0.309 [0.237 - 0.452]		
	D	1.000 [1.000]		
	Mixt.	1 gr.: 2.2 (---)		
	Ex.R.	1.110 [1.059 - 1.167]		
P3 - Width	*N*	48		
	*V**	9.31 [7.00 - 11.37]		
	b	0.368 [0.270 - 0.555]		
	D	1.000 [1.000]		
	Mixt.	*Failed*		
	Ex.R.	1.112 [1.029 - 1.202]		
P3 - Ln(LxW)	*N*	48		
	*V**	13.20 [9.73 - 16.50]		
	b	0.424 [0.333 - 0.579]		
	D	1.000 [1.000]		
	Mixt.	2 gr.: 1.82 (1.184 [1.112])		
	Ex.R.	1.182 [1.083 - 1.301]		
P4 - Length	*N*	63	34	15
	*V**	7.75 [6.09 - 9.17]	5.87 [3.66 - 8.01]	7.58 [4.26 - 9.44]
	b	0.377 [0.268 - 0.535]	0.376 [0.306 - 0.633]	0.514 [0.344 - 0.695]
	D	1.000 [1.000]	1.000 [1.000]	1.094 [1.000]
	Mixt.	2 gr.: 3.7 (1.198 [1.165])	*Failed*	*Failed*
	Ex.R.	1.088 [1.045 - 1.135]		

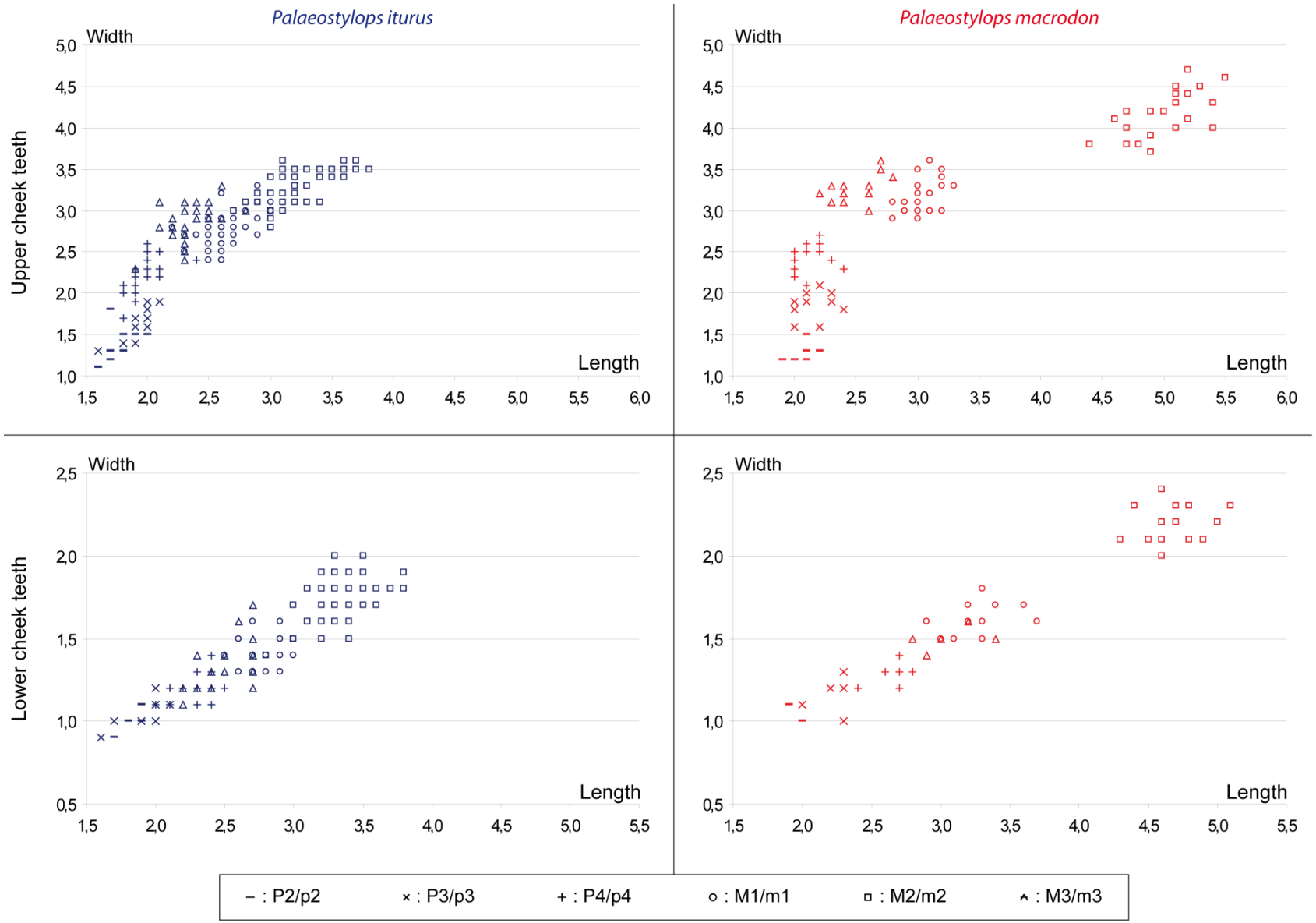
P4 - Width	*N*	63	35	14
	*V**	9.18 [7.53 - 10.69]	8.71 [6.41 - 10.77]	7.03 [4.40 - 8.88]
	b	0.337 [0.278 - 0.477]	0.291 [0.235 - 0.487]	0.314 [0.246 - 0.547]
	D	1.000 [1.000]	1.000 [1.000]	1.104 [1.000]
	Mixt.	1 gr.: 4.5 (---)	*Failed*	1 gr.: 25.3 (---)
	Ex.R.	1.092 [1.044 - 1.142]		
P4 - Ln(LxW)	*N*	61	34	14
	*V**	9.83 [7.94 - 11.48]	9.02 [6.65 - 11.16]	6.21 [4.33 - 7.36]
	b	0.412 [0.331 - 0.530]	0.312 [0.255 - 0.542]	0.410 [0.318 - 0.607]
	D	1.000 [1.000]	1.000 [1.000]	1.112 [1.046]
	Mixt.	1 gr.: 1.0 (---)	*Failed*	1 gr.: 5.0 (---)
	Ex.R.	1.109 [1.063 - 1.157]		
M1 - Length	*N*	74	50	20
	*V**	9.41 [8.14 - 10.51]	7.12 [5.74 - 8.28]	5.00 [3.61 - 5.96]
	b	0.435 [0.368 - 0.545]	0.349 [0.276 - 0.501]	0.386 [0.284 - 0.570]
	D	1.156 [1.116]	1.089 [1.000]	1.084 [1.032]
	Mixt.	2 gr.: 3.0 (1.162 [1.139])	1 gr.: 1.49 (---)	*Failed*
	Ex.R.	1.163 [1.130 - 1.197]		
M1 - Width	*N*	75	50	20
	*V**	9.81 [8.39 - 11.01]	7.29 [5.79 - 8.55]	6.62 [4.85 - 7.81]
	b	0.458 [0.377 - 0.563]	0.372 [0.292 - 0.510]	0.436 [0.336 - 0.603]
	D	1.154 [1.090]	1.083 [1.000]	1.113 [1.052]
	Mixt.	2 gr.: 4.3 (1.145 [1.116])	1 gr.: 3.0 (---)	1 gr.: 2.7 (---)
	Ex.R.	1.155 [1.116 - 1.195]		
M1 - Ln(LxW)	*N*	74	50	20
	*V**	8.67 [7.61 - 9.56]	6.31 [5.14 - 7.23]	4.46 [3.43 - 5.17]
	b	0.505 [0.434 - 0.596]	0.450 [0.346 - 0.581]	0.446 [0.344 - 0.631]
	D	1.156 [1.127]	1.096 [1.000]	1.079 [1.051]
	Mixt.	2 gr.: 42.5 (1.151 [1.130])	2 gr.: 1.49 (1.136)	2 gr.: 3.6 (1.079 [1.064])
	Ex.R.	1.150 [1.121 - 1.180]		
M2 - Length	*N*	79	58	21
	*V**	22.21 [19.80 - 23.79]	7.57 [6.21 - 8.72]	5.85 [4.16 - 7.14]
	b	0.779 [0.722 - 0.832]	0.383 [0.292 - 0.562]	0.356 [0.283 - 0.541]
	D	1.407 [1.103]	1.093 [1.000]	1.087 [1.000]
	Mixt.	2 gr.: 2.5×10^{19} (1.537 [1.499])	2 gr.: 2.3 (1.145 [1.122])	1 gr.: 9.9 (---)
	Ex.R.	1.541 [1.493 - 1.588]		

M2 - Width	*N*	80	58	21
	*V**	12.42 [10.48 - 13.97]	6.17 [5.03 - 7.16]	6.93 [5.10 - 8.21]
	b	0.560 [0.462 - 0.657]	0.503 [0.410 - 0.628]	0.388 [0.298 - 0.552]
	D	1.161 [1.000]	1.061 [1.000]	1.119 [1.064]
	Mixt.	2 gr.: 97734 (1.260 [1.226])	2 gr.: 944 (1.096 [1.078])	1 gr.: 3.9 (---)
	Ex.R.	1.259 [1.218 - 1.300]		
M2 - Ln(LxW)	*N*	79	58	21
	*V**	12.48 [10.96 - 13.59]	5.23 [4.31 - 6.01]	3.81 [2.82 - 4.54]
	b	0.685 [0.617 - 0.758]	0.377 [0.302 - 0.483]	0.391 [0.312 - 0.568]
	D	1.216 [1.097]	1.071 [1.000]	1.063 [1.033]
	Mixt.	2 gr.: $1.2x10^{13}$ (1.276 [1.255])	1 gr.: 2.1 (---)	1 gr.: 5.0 (---)
	Ex.R.	1.280 [1.254 - 1.306]		
M3 - Length	*N*	56	39	16
	*V**	8.50 [6.59 - 10.20]	7.68 [5.64 - 9.41]	7.69 [5.41 - 9.14]
	b	0.273 [0.217 - 0.436]	0.263 [0.208 - 0.494]	0.422 [0.340 - 0.595]
	D	1.000 [1.000]	1.000 [1.000]	1.133 [1.000]
	Mixt.	1 gr.: 1.42 (---)	1 gr.: 5.6 (---)	1 gr.: 1.30 (---)
	Ex.R.	1.048 [1.004 - 1.094]		
M3 - Width	*N*	53	37	15
	*V**	8.43 [6.36 - 10.23]	7.45 [5.14 - 9.32]	5.27 [3.07 - 6.61]
	b	0.277 [0.208 - 0.494]	0.474 [0.299 - 0.674]	0.438 [0.289 - 0.645]
	D	1.000 [1.000]	1.000 [1.000]	1.066 [1.000]
	Mixt.	2 gr.: 2.7 (1.031 [1.004])	2 gr.: 8.8 (1.078 [1.045])	1 gr.: 6.3 (---)
	Ex.R.	1.115 [1.078 - 1.155]		
M3 - Ln(LxW)	*N*	52	36	15
	*V**	7.77 [5.78 - 9.59]	6.91 [4.77 - 9.06]	5.33 [3.16 - 6.33]
	b	0.282 [0.223 - 0.491]	0.367 [0.286 - 0.611]	0.537 [0.368 - 0.682]
	D	1.000 [1.000]	1.000 [1.000]	1.083 [1.000]
	Mixt.	2 gr.: 20.1 (1.028 [1.001])	2 gr.: 2.5 (1.055 [1.027])	2 gr.: 13.7 (1.112 [1.097])
	Ex.R.	1.085 [1.050 - 1.123]		

Lower teeth		*Palaeostylops*	*P. iturus*	*P. macrodon*
p3 - Length	*N*	38		
	*V**	9.62 [7.25 - 11.52]		
	b	0.37 [0.265 - 0.510]		
	D	1.040 [1.000]		
	Mixt.	*Failed*		
	Ex.R.	1.162 [1.089 - 1.244]		
p3 - Width	*N*	38		
	*V**	10.44 [8.19 - 12.21]		
	b	0.366 [0.285 - 0.494]		
	D	1.163 [1.000]		
	Mixt.	1 gr.: 7.4 (---)		
	Ex.R.	1.133 [1.031 - 1.233]		
p3 - Ln(LxW)	*N*	37		
	*V**	24.55 [18.55 - 29.67]		
	b	0.354 [0.273 - 0.486]		
	D	1.000 [1.000]		
	Mixt.	1 gr.: 3.8 (---)		
	Ex.R.	1.400 [1.186 - 1.694]		
p4 - Length	*N*	50	16	7
	*V**	9.31 [7.15 - 11.27]	6.03 [3.40 - 7.73]	4.96 [1.44 - 7.07]
	b	0.262 [0.211 - 0.469]	0.464 [0.280 - 0.672]	0.408 [0.119 - 0.645]
	D	1.000 [1.000]	1.054 [1.000]	1.000 [1.000]
	Mixt.	*Failed*	1 gr.: 1.51 (---)	*Failed*
	Ex.R.	1.155 [1.104 - 1.204]		
p4 - Width	*N*	52	16	8
	*V**	10.22 [7.76 - 12.37]	7.57 [4.50 - 9.73]	6.68 [3.74 - 7.94]
	b	0.413 [0.309 - 0.584]	0.422 [0.267 - 0.723]	0.307 [0.178 - 0.627]
	D	1.000 [1.000]	1.080 [1.000]	1.116 [1.053]
	Mixt.	2 gr.: 2.2 (1.154 [1.089])	1 gr.: 7.3 (---)	1 gr.: 1746 (---)
	Ex.R.	1.084 [1.026 - 1.144]		
p4 - Ln(LxW)	*N*	50	16	7
	*V**	15.95 [12.87 - 18.70]	11.35 [7.36 - 14.31]	7.67 [3.20 - 10.33]
	b	0.338 [0.279 - 0.509]	0.290 [0.238 - 0.555]	0.237 [0.154 - 0.526]
	D	1.000 [1.000]	1.161 [1.000]	1.098 [1.000]
	Mixt.	1 gr.: 5.0 (---)	1 gr.: 24.8 (---)	1 gr.: 24343 (---)
	Ex.R.	1.215 [1.128 - 1.306]		

m1 - Length	*N*	69	31	14
	*V**	10.60 [8.45 - 12.44]	4.94 [3.91 - 5.74]	6.93 [3.88 - 8.84]
	b	0.408 [0.304 - 0.567]	0.421 [0.336 - 0.580]	0.327 [0.209 - 0.567]
	D	1.000 [1.000]	1.082 [1.049]	1.082 [1.000]
	Mixt.	2 gr.: 14.2 (1.172 [1.149])	2 gr.: 1.77 (1.089 [1.074])	1 gr.: 5.1 (---)
	Ex.R.	1.173 [1.129 - 1.220]		
m1 - Width	*N*	68	29	14
	*V**	9.01 [7.43 - 10.36]	6.68 [5.09 - 7.82]	5.81 [3.53 - 7.23]
	b	0.387 [0.300 - 0.515]	0.417 [0.323 - 0.553]	0.350 [0.241 - 0.595]
	D	1.106 [1.000]	1.108 [1.000]	1.085 [1.000]
	Mixt.	*failed*	1 gr.: 3.7 (---)	1 gr.: 20.7 (---)
	Ex.R.	1.131 [1.090 - 1.174]		
m1 - Ln(LxW)	*N*	66	28	14
	*V**	12.03 [10.14 - 13.63]	6.78 [5.29 - 7.88]	6.53 [4.49 - 7.71]
	b	0.415 [0.320 - 0.557]	0.414 [0.323 - 0.562]	0.390 [0.266 - 0.583]
	D	1.177 [1.000]	1.117 [1.072]	1.116 [1.055]
	Mixt.	2 gr.: 10.5 (1.259 [1.230])	1 gr.: 1.65 (---)	1 gr.: 2.8 (---)
	Ex.R.	1.208 [1.160 - 1.257]		
m2 - Length	*N*	75	56	19
	*V**	17.26 [14.92 - 18.80]	5.72 [4.43 - 6.86]	4.64 [3.07 - 5.75]
	b	0.771 [0.708 - 0.830]	0.252 [0.211 - 0.454]	0.308 [0.225 - 0.577]
	D	1.274 [1.000]	1.000 [1.000]	1.062 [1.000]
	Mixt.	2 gr.: $1x10^{19}$ (1.413 [1.384])	1 gr.: 9.0 (---)	1 gr.: 21.1 (---)
	Ex.R.	1.413 [1.378 - 1.449]		
m2 - Width	*N*	78	55	19
	*V**	12.79 [11.02 - 14.24]	7.82 [6.45 - 9.01]	5.04 [3.64 - 6.02]
	b	0.494 [0.404 - 0.597]	0.389 [0.328 - 0.527]	0.448 [0.358 - 0.736]
	D	1.208 [1.124]	1.110 [1.000]	1.084 [1.000]
	Mixt.	2 gr.: 110 (1.255 [1.221])	1 gr.: 1.82 (---)	2 gr.: 3.2 (1.089 [1.070])
	Ex.R.	1.256 [1.220 - 1.295]		
m2 - Ln(LxW)	*N*	74	55	19
	*V**	14.52 [12.61 - 15.97]	6.74 [5.26 - 8.22]	3.27 [2.27 - 3.96]
	b	0.613 [0.494 - 0.741]	0.355 [0.285 - 0.531]	0.359 [0.265 - 0.545]
	D	1.237 [1.000]	1.000 [1.000]	1.050 [1.000]
	Mixt.	2 gr.: $8.8x10^{11}$ (1.328 [1.303])	1 gr.: 1.73 (---)	1 gr.: 7.8 (---)
	Ex.R.	1.328 [1.299 - 1.359]		

m3 - Length	*N*	46	26	7
	*V**	9.96 [7.49 - 12.14]	7.85 [6.08 - 8.99]	7.27 [1.76 - 9.18]
	b	0.313 [0.234 - 0.512]	0.605 [0.468 - 0.756]	0.411 [0.180 - 0.645]
	D	1.000 [1.000]	1.148 [1.068]	1.097 [1.000]
	Mixt.	1 gr.: 1.07 (---)	*Failed*	1 gr.: 2164 (---)
	Ex.R.	1.197 [1.135 - 1.269]		
m3 - Width	*N*	54	32	9
	*V**	9.96 [8.32 - 11.36]	10.41 [7.97 - 12.47]	5.93 [3.48 - 6.85]
	b	0.399 [0.323 - 0.536]	0.379 [0.291 - 0.574]	0.350 [0.176 - 0.607]
	D	1.169 [1.112]	1.159 [1.000]	1.107 [1.000]
	Mixt.	*Failed*	*Failed*	1 gr.: 403 (---)
	Ex.R.	1.114 [1.059 - 1.171]		
m3 - Ln(LxW)	*N*	45	25	7
	*V**	14.16 [11.45 - 16.46]	14.06 [10.58 - 16.76]	7.34 [2.66 - 8.42]
	b	0.381 [0.319 - 0.553]	0.404 [0.319 - 0.583]	0.420 [0.259 - 0.645]
	D	1.233 [1.000]	1.247 [1.000]	1.125 [1.000]
	Mixt.	1 gr.: 1.65 (---)	1 gr.: 3.0 (---)	1 gr.: 11.7 (---)
	Ex.R.	1.222 [1.139 - 1.316]		



Supplementary Figure S1. Bivariate Length × Width scatterplots for *Palaeostylops iturus* and *P. macrodon* cheek teeth from the Gashatan of the Khashat Formation in the Flaming Cliffs area (Mongolia). All measurements in mm.