

1 ***A 'long-fuse domestication' of the horse? Tooth shape suggests explosive change in***
2 ***modern breeds compared to extinct populations and living Przewalski horses***

3

4 Krish Seetah ^{1†}, Andrea Cardini ^{2,3} & Graeme Barker ⁴

5

6 ^{1†} Corresponding author:

7 Department of Anthropology, Stanford University, Stanford, CA 94305, USA.

8 Email: kseetah@stanford.edu

9 Tel: 1 650 721 5939

10

11 ²Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia,
12 Via Campi 103, 41125 Modena, Italy

13

14 ³Centre for Forensic Science, The University of Western Australia, 35 Stirling Highway,
15 Crawley, WA 6009, Australia.

16

17 ⁴McDonald Institute for Archaeological Research, Cambridge, CB2 3ER, United Kingdom.

18

19

20 **Abstract**

21 Archaeological and molecular data suggests that horses were domesticated comparatively
22 recently, the genetic evidence indicating that this was from several maternal haplotypes but
23 only a single paternal one. However, although central to our understanding of how humans
24 and environmental conditions shaped animals during domestication, the phenotypic
25 changes associated with this idiosyncratic domestication process remain unclear. Using
26 geometric morphometrics on a sample of horse teeth including Pleistocene wild horses,
27 modern Icelandic and Thoroughbred domestic horses, Przewalski wild horses of recent age,
28 and domestic horses of different ages through the Holocene, we show that, despite
29 variations in size likely related to allometry (changes to bone size in proportion to body
30 size), natural and artificial selective pressures, and geographic and temporal heterogeneity,
31 the *shape* of horse teeth has changed surprisingly little over thousands of years across
32 Eurasia: the shapes of the premolars of prehistoric and historic domestic horses largely
33 resemble those of Pleistocene and recent wild horses. However, this changed dramatically
34 after the end of the Iron Age with an explosive increase in the pace and scale of variation in
35 the past two millennia, ultimately resulting in a two-fold jump in the magnitude of shape
36 divergence in modern breeds. Our findings indicate that the pace of change during
37 domestication may vary even within the same structure with shape, but not size, suggesting

38 a 'long-fuse' model of phenotypic modification, where an initial lengthy period of
39 invariance is followed by an explosive increase in the phenotypic change. These
40 observations support a testable model that is applicable to other traits and species, and add a
41 new layer of complexity to the study of interactions between humans and the organisms
42 they domesticated.

43

44 **Key words**

45 Artificial selection; disparity; morphology; phenotype; Procrustes; teeth.

46

47 **Introduction**

48 Horse domestication had a profound effect on humans (Clutton-Brock 1999; Anthony
49 2007; Olsen 2006). The timing of horse domestication around 3500 BCE (Outram et al.
50 2009), and the subsequent extirpation of wild progenitors, potentially correlates with
51 important environmental shifts i.e. expansion and contraction of forest cover (Warmuth et
52 al. 2011). However, reconstructing the history of this event has proved challenging
53 (Lippold et al. 2011; Wade et al. 2009; Steiner 2013); a deeper knowledge of the processes
54 that underpinned horse domestication has important implications for our understanding of
55 both the human groups that first undertook this particular domestication event (or events),

56 as well as the ecological setting within which domestication was initiated. Critically, with
57 the expansion of domestic populations our ability to distinguish between discrete groups of
58 wild and domestic horses diminishes to nil (Bendry 2012).

59

60 Maternally inherited mtDNA of modern horse breeds demonstrates a high degree of
61 variability, with an exponential population expansion starting 6000-8000 years ago
62 (Lippold et al. 2011). Genomic analyses confirm that horses did not undergo strong genetic
63 bottlenecks and indicate close relationships among breeds and even among domestic horses
64 and their closest living relative, the Przewalski wild horse, *Equus przewalskii* (Wade et al.
65 2009). The domestication process is strongly sex-biased (Wade et al. 2009): differences in
66 paternal-inherited DNA are small and all modern breeds share the same Y chromosome
67 haplotype (Lingren et al. 2004). Overall, the genetic evidence suggests that mares from
68 multiple populations contributed to the gene pool but that only a few stallions were
69 domesticated.

70

71 Archaeological data supporting western central Eurasia as a likely centre of early horse
72 domestication (the presence of mares-milk and carcass residues in ceramic vessels,
73 alongside pathological markers and bit-wear indicative of horse riding c.3500 BCE: Outram

74 et al. 2009) are now corroborated by autosomal genotype data (Warmuth et al. 2011; 2012).
75 These same archaeological lines of evidence point to an established and well-developed
76 economy that included the exploitation of a range of secondary products (Outram et al.
77 2009; Bendry 2011), not just the horse's ability to transport people and goods. Thus,
78 despite strong consensus that the horse was domesticated for riding, we still need a better
79 assessment of the overall process of domestication and how both the socio-economic and
80 environmental setting influenced this event over the course of time and across space.

81

82 *Investigating morphological responses*

83 As selection acts on phenotypes, the manner in which behaviour and morphology have been
84 shaped by human-animal interactions, and with the environment, is fundamental to our
85 understanding of the process of domestication. Although genetic approaches have
86 elucidated phenotypic features such as coat colour (Ludwig et al. 2009) and gait variability
87 (Andersson et al. 2012), we have yet to make strong connections between the behavioural
88 and anatomical correlates of genetic changes. Furthermore, at least some aspects of
89 caballine equid morphology potentially reflect adaptation: larger phalanges in glacial horses
90 might be an adaptation to heavy grounds (Bignon & Eismann 2012); shorter muzzles in
91 cold environments follow the prediction of Allen's rule (Eismann & Baylac 2000).

92

93 Teeth offer an excellent starting point to address adaptive phenotypic responses. They are
94 one of the most frequently recovered faunal elements within an archaeological setting, they
95 are shaped by natural selection to meet functional demands in relation to diet, and they
96 often mirror important features of a species' environmental setting (Kaiser & Schulz 2006;
97 Evans 2013). Indeed, the term 'dental ecology' has been used to illustrate the power of
98 teeth for informing on responses to the environment (Cuozzo & Sautner 2012). Using
99 geometric morphometrics (henceforth GMM: Adams et al. 2013), we explored premolar
100 variation in a unique dataset, in terms of spatio-temporal variation. The samples span
101 Eurasia, with the majority of the archaeological specimens dating to *c.* 4500-2000 years
102 ago (Fig. 1a; sample details in Supplementary Information). This period is well after initial
103 domestication, but during a time when populations of wild horses were still present
104 (Warmuth et al. 2011). GMM provides a robust framework within which to study both size
105 and shape data, resulting in crucial phenotypic evidence that can serve as an important
106 complement to molecular techniques such as DNA and isotope analysis. The method offers
107 other advantages: for example, it is non-destructive, and large samples can be assessed with
108 minimal expense, as the technique itself is low-cost. GMM offers a significant, and
109 relatively new, addition to the arsenal of techniques that capitalizes on an abundant primary

110 archaeological resource, faunal remains, to address key issues about human / animal /
111 environmental interactions. Thus, GMM offers a powerful toolkit to assess how much a
112 biometric marker (Houle et al. 2010), such as the premolar, has been shaped by interactions
113 with humans and the environment during domestication.

114

115 >> Fig. 1. (a) Geographic distribution and age of samples. (b) Premolar landmark
116 configuration. (c) Number (N) of individuals per sample, box-plots of premolar centroid
117 size and icons (not to scale) emphasizing premolar size clusters: small (light grey), medium
118 (grey), large (black).

119

120 >> Tab. 1: Tests for mean size differences: the percentage of variance explained by
121 differences is under the main diagonal; P is above it. Two P values are shown: for the first
122 one, Student's t is tested using 10000 permutations, and, for the second one, the parametric
123 Student's t with no assumption of equal variances is used; significant Ps are in italics and
124 Bonferroni corrected significant Ps (<0.0018) are both in italics and underscored.
125 Comparisons between 'archaic' samples are emphasized using a light grey background.

126

127 **Materials and Methods**

128 *Samples summary*

129 Our comparative framework for the archaeological samples consisted of wild horses and
130 two modern breeds, Icelandic (ICE: n = 50) and Thoroughbred (THB: n = 18). The wild
131 specimens belong to an extinct Late Pleistocene population from, Šandalja, Croatia (CRO:
132 n = 19), dated to 40,000 to 8200 yrs uncal BP and potentially *Equus ferus* – although there
133 is no consensus on the number of wild horse species from this time period – and a small
134 sample of modern Przewalski horses (PRZ: n = 4). The latter is the only living wild horse
135 and likely retains primitive traits (Groves & Ryder 2000). The remaining archaeological
136 samples were predominantly of domestic individuals. These include: Bronze Age
137 specimens from Berel, in the Katonkaragaray Eastern Kazakh Oblast, excavated from
138 Kurgan burial mounds (KAZ: n = 22); Bell-Beaker period materials from
139 Szigetszentmiklós, Hungary (HUN: n = 10); samples from China dated to the East Zhou
140 Dynasty (alongside two Pleistocene specimens from this region) (CHI: n = 16) and Iron
141 Age materials from the site of Slepushka, (and one Bronze Age specimen from Ust'e),
142 Russia (RUS: n = 17). We use the term 'archaic' to refer to the archaeological specimens,
143 the Pleistocene wild horses, and the Przewalski horses. Anatomical landmarks are shown in
144 Figure 1b. Overall, 156 individual horses were studied, with an average sample size of 20
145 specimens per sample (Fig.1c and SI). Sex differences were negligible (Seetah et al. 2014).

146 All animals were adult with a similar degree of tooth wear. Because of variation in the
147 enamel folding between the apex of the crown and the cervical margin, an individual tooth
148 will show an apparent change in shape of the folding as the tooth wears. Tooth wear also
149 impacts on the proportions of the teeth themselves (Gidley 1901). For the latter, this
150 situation is compounded by the fact that tooth wear affects the individual teeth differently;
151 thus, tooth wear impacts on the proportions of a P4 differently to an M3. As we only used
152 P4 teeth, modifications to the proportions of individual teeth, consequential to wear, should
153 be uniform across our assemblage. Further, to the best of our ability we minimized
154 variations in enamel folding across the assemblage by selecting teeth with a similar crown
155 height (as per Gidley 1901: 97) and rejecting those that showed excess wear.

156

157 *Methods*

158 Size and shape variables were computed using a Procrustes superimposition (Rohlf & Slice
159 1990) on the raw landmark data (Seetah et al. 2014) digitized on high resolution digital
160 images in TPSDig (Rohlf 2015). Groups were tested using 10,000 permutations for mean
161 differences in size (absolute difference – i.e., the sign is not considered and the test
162 probability is two-tailed) and shape (Procrustes distance between mean shapes; two-tailed
163 test). Results were double-checked using t-tests for samples with unequal variance (size) or

164 their multivariate equivalent (shape) using James' statistics (Dryden 2013). Tests were
165 performed in MorphoJ (Klingenberg 2011), PAST (Hammer et al. 2001), NTSYSpc (Rohlf
166 2013) and R (R Development Core Team 2005). Allometry was tested in MorphoJ using a
167 multivariate regression of shape onto the natural logarithm of centroid size and significance
168 assessed with 10,000 permutations for the percentage of variance explained by size.
169 Variation within and among samples was summarized in PAST with box-plots for size and
170 between group principal component scatterplots for shape (BG-PCA - Seetah et al 2012). A
171 BG-PCA projects the specimens data onto the eigenvectors of the mean shapes variance-
172 covariance matrix.

173

174 Focusing exclusively on mean shapes, variation was summarized in R using principal
175 component analysis (PCA) with 95% confidence envelopes computed using 1000
176 bootstraps in NTSYSpc. For clarity, as two types of principal component analysis were
177 used (BG-PCA and 'standard' PCA), we stress here that, every time we refer to the between
178 group analysis, abbreviations are preceded by the acronym BG- (i.e., BG-PCA or BG-PCs);
179 if this acronym is missing, however, PCA and PCs simply refer to a principal component
180 analysis summarizing total variance in a sample regardless of groups. Size and shape
181 similarity relationships among sample means were also summarized in NTSYSpc with

182 distance-based trees (neighbour joining using Euclidean distances for size and Procrustes
183 distances for shape). To take into account uncertainties in estimates of means, node
184 repeatability was assessed by bootstrapping each sample, computing the corresponding
185 pseudo-means and the resulting tree, and finally computing a 50% majority rule consensus
186 tree from all 1000 bootstrapped trees. Group separation was also assessed by estimating
187 cross-validated classification accuracy using the first eight PCs of premolar shape (91.7%
188 of variance) in a discriminant analysis (DA). The general methodological framework used
189 in this study is described in Cardini (Cardini 2013) and exemplified in Viscosi & Cardini
190 (2011). The bootstrap procedures used to compute the confidence envelopes and the
191 percentages of node repeatability are detailed in Cardini & Elton (2008). The protocol used
192 to perform the DA, including the estimate of random chance thresholds and the sensitivity
193 analysis using a balanced design, is described in Evin et al. (2013).

194

195 Disparity (morphological variance in modern *versus* archaic horses) was tested using
196 methods modified from the protocol described by Drake & Klingenberg (2010; for an
197 extensive review on disparity see also Foote 1997). For univariate data, this implied a
198 simple test for the similarity of variances, which is the permutational version of the
199 Levene's test used by Nagorsen & Cardini (2009). The same test can be adapted to test

200 multivariate shape variances estimated by the mean squared Procrustes distance of each
201 individual to its group mean (D1) or equivalently the sum of variances of all shape
202 coordinates (Drake & Klingenberg 2010; Nagorsen & Cardini 2009). Informally, this is
203 akin to measuring the sum of the squared sides of the multivariate 'box' which contains the
204 data. An alternative test statistics (D2) is based on estimates of the squared volume of the
205 'box' occupied by the data. The product of the eigenvalues of the within-group shape
206 variance covariance matrix, which is the same as its determinant if all eigenvalues are used,
207 provides one way to do this estimate. However, similarly to Drake and Klingenberg (2010),
208 who used an alternative method based on convex hulls, we estimated the volume occupied
209 by either modern or archaic individuals within the sub-space of the first three PCs of the
210 total shape dataset. These PCs account for two thirds of the total variance and, by using
211 them to estimate the volumes of the shape space occupied by each group, we reduced issues
212 with numerical precision using very small numbers, such as those generated by the product
213 of eigenvalues from Procrustes shape data (Drake and Klingenberg 2010). D2 is therefore
214 computed in a subspace of the total shape space, whereas D1 uses all available shape
215 information (i.e., there is no dimensionality reduction). Thus, to summarize D2, its
216 computation meant performing two eigen-decompositions on the shape data: i) the first one
217 used all individuals regardless of group and was performed on shape coordinates (from a

218 common superimposition), as it simply aimed at building a subspace of the total shape
219 space in which to compute disparity as a 'volume'; ii) the second one took the shape
220 variables corresponding to the subspace built in i) (i.e., PC1, PC2 and PC3 scores); it was
221 performed within-group and used the product of the three eigenvalues of one group to
222 estimate its disparity in that subspace.

223

224 **Results**

225 Moderate premolar size differences were present despite large overlaps across samples
226 (Fig. 1c, Table 1). Using means, the resulting tree was well supported with three main size
227 subdivisions suggested by the longest branches of the dendrogram (Fig. 2): small (ICE,
228 HUN and RUS), medium (THB and KAZ) and large (CHI, CRO and PRZ). Bonferroni-
229 corrected tests were significant (with ca. 37% of variance explained on average) only when
230 they involved groups of small-toothed versus large-toothed animals, with the largest
231 premolars being about 15% bigger than the smallest ones. Size groups were made of a mix
232 of samples from different periods and localities. There was no clear pattern in terms of time
233 or geography: for instance, modern ICE were similar in size to Iron Age RUS and Bronze
234 Age HUN. With the exception of the large-toothed CHI sample, however, the largest
235 premolars belong to the two groups of wild horses (Pleistocene CRO and modern PRZ).

236 For CHI, the large mean size was not due to a bias in relation to the two Pleistocene
237 individuals in the sample, as their size was actually very close to but slightly (ca. 1 mm)
238 less than the mean size of Iron Age CHI.

239

240 >> Fig. 2. Summary of mean size variation: neighbour joining tree of mean sizes with
241 percentages of support for branches. Three main size subdivisions are emphasized using
242 different grey tones, as in Fig. 1. Arrows help visualizing pairwise comparisons significant
243 after a Bonferroni correction.

244

245 Shape had little covariation with size (Table 2). Thus, allometry seems to have had minor
246 and generally negligible effect both within and across samples. In contrast, in the
247 scatterplots capturing most of the shape variance (Fig. 3, 82.7% of between group variance,
248 corresponding to 50.7% of total sample variance; Fig. 4a, 91.0% of variance in means),
249 variation was highly structured and the pattern looked unexpectedly simple: modern breeds
250 showed very clear differences between each other and when compared to all archaic horses;
251 archaic horses largely overlapped with the four archaeological samples having similar
252 premolar shape and also resembling extinct (CRO) and living (PRZ) wild horses. Shape
253 components other than those shown in the scatterplots did not suggest any group

254 separation. In terms of the shape features that characterized the three main clusters (Fig. 3),
255 ICE had a relatively short and thick premolar; THB was comparatively long and narrow;
256 archaic samples had a somewhat intermediate shape.

257

258 >> Tab. 2: Tests for allometry within and among (using means) samples: multivariate
259 regression of shape onto the natural logarithm of centroid size (10000 permutation test for
260 the percentage of explained variance; significant Ps are in italics and in italics and
261 underscored if significant after a Bonferroni correction - $P < 0.0063$).

262

263 Tests of mean shape differences (Table 3, Fig. 4b) provided results in agreement with the
264 pattern suggested by the scatterplots. Virtually all Bonferroni-corrected comparisons
265 involving ICE and THB were significant. In contrast, out of 15 possible pairwise
266 comparisons between pairs of archaic samples, only five were significant. The variance
267 explained by differences with modern breeds (on average 27.7%) was almost twice as much
268 as that found in archaic horses (on average 15.6%). The average shape distance of either
269 ICE or THB to any of the archaic horses was more than double that which was noted
270 among pairs of archaic samples (respectively ≥ 0.0859 *versus* 0.0425 units of Procrustes
271 shape distance).

272

273 >> Tab. 3: Pairwise tests (10000 permutations) for mean shape differences: the percentage
274 of variance explained by differences is under the main diagonal; P is above it. Two P values
275 are shown: for the first one, the Procrustes distance between sample means is the test
276 statistics, and for the second one, James' statistics, which does not assume
277 homoscedasticity, is used; significant Ps are in italics and Bonferroni corrected significant
278 Ps (<0.0018) are both in italics and underscored. Comparisons between 'archaic' samples
279 are emphasized using a light grey background.

280

281 Because sample size was heterogeneous and the PRZ sample was particularly small, results
282 may have been affected by unbalanced sampling and differences in statistical power. To
283 explore the sensitivity to sampling, we excluded PRZ and repeated all tests on shape using
284 a perfectly balanced design with subsamples of 10 random individuals from each original
285 sample. (N = 10 was chosen as this was the sample size of HUN, the second smallest
286 sample in the study.) Results from the balanced design are briefly summarized here, as
287 they show no appreciable difference compared to the analysis including all specimens: even
288 with a Bonferroni correction, all 11 tests involving modern breeds (ICE and THB) were
289 significant; however, with the same correction, only four of the 10 comparisons between

290 pairs of archaic samples showed significant differences.

291

292 Figure 4b shows a neighbour joining tree rooted using PRZ as an 'outgroup'. Phylogenetic
293 inference using Procrustes shape data is problematic (Adams et al. 2013) and was not our
294 aim. The tree is better seen as another type of graphical summary of mean shape similarity
295 relationships, a summary with a very high cophenetic correlation (0.971, an index of high
296 accuracy in reproducing the pattern of full multivariate shape distances) and the potential of
297 rooting the tree to explore directional change. When this was done, the tree topology
298 suggested a progressive trend in change from wild horses to modern breeds with
299 archaeological samples in between and modern breeds at the tips of very long branches.
300 However, only two clusters were well supported when inaccuracies in estimates of mean
301 shape were taken into account. The two clusters suggest the same sharp contrast between
302 modern and archaic animals as in the scatterplots and pairwise tests.

303

304 The magnitude and direction of group shape differences were further explored using cross-
305 validated discriminant analyses (Table 4). The baselines to assess classification accuracy
306 were the estimated average and the 95th percentile of correct classification by random
307 chance, which was respectively 26.3% and 30.1%. Modern breeds with the highest

308 accuracies (more than 83% of correctly classified individuals) were thus about three times
309 better than chance. Classification accuracy varied in archaic samples but it was lower and
310 averaged 43%, with a small increase to 56% if the two smallest samples (HUN and PRZ)
311 were not considered. Thus, although on average better than chance, about half of archaic
312 individuals were misclassified into other archaic samples; however, none were
313 misclassified as modern. With a perfectly balanced design (not shown) to control for
314 sample size heterogeneity (the same design as in the balanced permutation tests), modern
315 breeds would have a slightly lower average classification accuracy (75%) and archaic
316 samples a slightly higher one (58%), but the general outcome was similar and no archaic
317 individuals were misclassified as modern breeds. To put these percentages into context, if
318 they were interpreted using an arbitrary but common criterion employed by taxonomists for
319 wild populations, modern breeds would be consistently above the arbitrary 75% threshold
320 in classification accuracy for a valid subspecies (Patten et al. 2002) and archaic samples
321 would be mostly below it.

322

323 The overwhelming signal from shape data analysed and summarized with a multiplicity of
324 approaches is that modern breeds are highly distinctive with long and narrow premolars
325 (THB) or short and thick ones (ICE), whereas archaic populations tend to have an

326 intermediate, probably more primitive, shape, that has varied little over time and across
327 localities. These observations lead us to predict an increase in disparity before and after
328 modern breeds were selected.

329

330 >> Tab. 4: Cross-validated percentages of individuals classified using a DA on premolar
331 shape. Correctly classified individuals are on the main diagonal; percentages higher than
332 95% random chance are emphasized in red italics underscored.

333

334 >> Tab. 5: Disparity analysis: disparity 1 (D1) is based on variance (size) or sum of
335 variances (shape coordinates) and the test statistics is the F ratio; disparity 2 (D2) is the
336 product of the first 3 eigenvalues of the shape variance covariance matrix and the test
337 statistics is their absolute difference; D1 and D2 are tested using 10000 permutations.

338

339 With only two modern breeds in the dataset, our samples likely underestimate their
340 disparity and any test has to be considered preliminary and largely exploratory. Bearing this
341 caveat in mind, we performed a disparity analysis. First, we pooled modern breeds and
342 tested disparity differences between them, together, and the pooled sample of all archaic
343 individuals. Then, we compared disparity between archaic horses (pooled) and each of the

344 two modern breeds (one at a time). In this second series of tests, we expected similar
345 amounts of disparity if variation in archaic horses, regardless of geographic origin and
346 period, was comparable to that found within a single modern breed. For comparative
347 purposes, disparity was also tested, using the same design, on premolar size.

348

349 We found that disparity in modern breeds (pooled), estimated by the sum of all eigenvalues
350 (D1, Table 5), is 1.2 times larger than in archaic individuals. This is not significant, or only
351 marginally so ($P = 0.09$). Disparity estimated using the product of the first three
352 eigenvalues (D2, Table 4) is, however, highly significantly different ($P = 0.0016$) and 4.7
353 times larger in modern horses. As expected, when archaic horses are compared to one or
354 the other modern breed (one sample at a time), disparity is generally larger in the archaic
355 group but never reaches significance ($P > 0.05$). Results from shape data were at odds with
356 those on size, which indicated that archaic individuals vary in size three times or more than
357 modern breeds even when pooled ($P < 0.0001$, D1, Table 4). Thus, disparity analysis
358 strongly suggests a discordant pattern of change in size and shape, with size varying in the
359 archaic horses and shape being constrained within a relatively small region of the
360 morphospace until the appearance of highly derived modern breeds.

361

362 **Potential sources of error**

363 Taxonomic and especially palaeontological analyses can be strongly affected by sampling
364 error because of the limited availability of individuals in one or more samples. In our study,
365 sample size is very heterogeneous, with ICE (N=50) and PRZ (N=4) falling at opposite
366 extremes in terms of number of specimens. We assessed the impact of sampling error in a
367 previous study using rarefaction analyses and resampling methods in this same horse teeth
368 dataset (Cardini et al. 2015). In that work, we concluded (p. 149) that “likely, centroid
369 size and shape variance require no less than 15–20 specimens to achieve a reasonable
370 degree of accuracy”. In the same study, largely in agreement with a previous one on
371 monkeys (Cardini and Elton 2008), we also showed that estimates of mean shapes may
372 require 20 or more specimens, while those of mean size can be fairly accurate with just 10
373 individuals. Thus, although most of our samples have sizes above or close to this minimum
374 putative requirement, a degree of caution must be exercised in the interpretation of results
375 and this is especially true for the smallest samples. Crucially, however, in terms of what the
376 populations we sampled allow us to say (see also below), the main conclusion of our study
377 seems robust. This is suggested by Figure 7a of Cardini et al. (2015) that indicates how,
378 despite a remarkable inflation of differences in the smallest random samples from the
379 rarefaction analysis of ICE, there is no overlap between ICE means and those of all other

380 groups. This is because differences are so large that small samples do not change the
381 general pattern, and this most likely holds also for THB, whose sample is relatively big and
382 whose shape is as distinctive as that of ICE. As for the analysis of mean shapes, results
383 from the disparity analysis are robust, because archaic samples are pooled, thus strongly
384 mitigating the problems with small samples.

385

386 Another source of inaccuracy is related to the inclusion of three allochronic individuals
387 (compared to the others from the same region) in two archaeological samples: two
388 Pleistocene specimens in CHI and one Bronze Age individual in RUS. This happened
389 because, at the time analyses were performed, we had no accurate information to age all
390 individuals and assumed that they were of similar age as the others from the same
391 geographical region. After we found out that CHI and RUS included 1-2 allochronic
392 individuals each, we investigated whether their inclusion in the analyses could have altered
393 results. To this aim, as detailed in the Supplementary Information, we showed that in terms
394 of both size and shape, the allochronic individuals were well within the range of the
395 corresponding samples and they did not appreciably alter the pattern of population
396 differences in means and variances (whose correlations before and after excluding those
397 three specimens are virtually 1). Thus, although the best choice would have been to leave

398 those three specimens out, we acknowledged the problem, demonstrated that it has no
399 practical consequences and kept them in the study. This avoided redoing all analyses,
400 which is a requirement when data are analysed in a common Procrustes shape space, which
401 is specific to the sample (and configuration) being analysed.

402

403 A third source of uncertainty is the limited number of samples in the analysis. Our samples
404 are representative of the whole Palearctic range, include the only living population of true
405 wild horse, the Przewalski horse, as well as an extinct one from the Pleistocene, and also
406 include two modern breeds and archaeological material from four different localities, over a
407 time span that corresponds to early and later stages of horse domestication. However,
408 clearly our results will have to be corroborated by future studies on other modern and
409 extinct populations. If, with additional samples from more regions, time periods and also
410 more modern breeds, the vast majority of extinct populations have small differences among
411 them and compared to Przewalski whereas most modern species have much more
412 distinctive shapes, then our hypothesis will be corroborated. If not, it will be rejected.

413

414 **Discussion**

415 With the provisos discussed above, our data indicate that three complimentary lines of

416 evidence (mean differences, classification accuracy and morphological disparity) suggest
417 that the *size* of horse premolars has changed to a variable degree over time and space in
418 both archaic and modern horses (Fig. 1c-2 and SI), but changes to *shape* have been modest,
419 with variation mostly overlapping amongst the archaic samples (Figs. 3 & 4) until the
420 development of modern breeds in recent centuries, when shape has become hugely
421 distinctive (Fig. 3 & 4). With an estimated time since the divergence of PRZ and the
422 lineage leading to domestic horses of more than 100,000 years (Steiner et al. 2013; Goto et
423 al. 2011; but see Der Sarkissian et al. 2015 for an alternative perspective), the data suggest
424 that conservativeness may have characterized premolar shape for almost 99% of the history
425 of *Equus caballus*, whereas in the last 1000-2000 years, under conditions of strong
426 selective breeding, shape differences have more than doubled compared to those observed
427 in archaic horses (Fig. 3). Thus, we propose, for premolar shape, an hypothesis of a 'long-
428 fuse' model of phenotypic change in domestication, whereby a long initial period of small
429 variation was followed by an explosive acceleration in the magnitude of shape change.

430

431 >> Fig. 3. Between group principal component analysis of shape (BG-PCA). Scatterplots of
432 the first two components (BG-PC1, 48.7%, and BG-PC2, 34.0%, of between group
433 variance) with groups emphasized using convex hulls. Mean shapes of ICE, THB and

434 archaic samples are shown using rendering of contours and deformation grids with
435 expansion factors, as in Viscosi & Cardini (2011) and Seetah et al. (2012).

436

437 >> Fig. 4. Summary of mean shape variation: (a) 3D scatterplot of the first three principal
438 components (91.0% of total variance) with 95% confidence ellipsoids around means; (b)
439 neighbour joining tree of mean shapes (using Procrustes distances and rooted at PRZ) with
440 percentages of support for branches and arrows to emphasize pairwise comparisons
441 significant after a Bonferroni correction.

442

443 We anticipated large phenotypic changes in modern breeds consequential to artificial
444 selection, although not specifically targeting tooth morphology. Differences in premolars
445 are in fact small in size but remarkably large in shape when ICE and THB are compared
446 (Fig. 3). Although of smaller magnitude, a degree of variation in premolar size across all
447 groups was also predictable across all samples based simply on variability in body mass
448 (e.g., the small PRZ and ICE vs. the medium-large THB), geographical distances, time
449 heterogeneity and likely genetic differences. These factors should, however, also affect
450 shape. In contrast, we found a clear disconnection between patterns of size and shape
451 variation: while size varies over time and space, showing no clear trend except a moderate

452 degree of reduction in most domesticated samples compared to wild horses, the amount of
453 shape change of archaic horses in almost 100,000 years is overall modest and especially
454 small if compared to the massive differences shown by modern breeds. This relative
455 phenotypic conservativeness in shape, despite size changes, is unexpected and potentially
456 might provide insight into both ecological and anthropological factors.

457

458 Dramatic ecological changes occurred after the end of the last glaciation *c.* 11,500 years
459 ago. Over the next 2000-3000 years, during the Early Holocene, temperature rose, forests
460 expanded and open habitats favoured by horses shrank (Warmuth et al. 2011). Humans
461 hunted horses and, with the expansion of agriculture in the mid Holocene, *c.* 8000-5000
462 years ago, contributed to modifying the environment. Regardless of when and how
463 domestication occurred, horses had to cope with a variety of selective pressures, whose
464 combined effect was profound for the wild counterpart, resulting in the extinction of those
465 populations. Further, mtDNA diversity was greater before *c.* 2800 years ago rather than
466 after (Lippold et al. 2011). Thus, as we anticipated, both environmental and genetic factors
467 suggest that change would be expected, an expectation that is in agreement with our
468 findings regarding size but at odds with those for shape. In fact, as premolar size varies in
469 several archaic samples, allometry on its own, with its typically pervasive effect on bone

470 shape (Klingenberg 1998), would predict concomitant shape differences, which we have
471 not found.

472

473 Large shape changes seem to appear suddenly (in geological terms) in populations dated to
474 after the first millennium BCE. This might be consistent with a *refining* phase during
475 domestication that profoundly impacted on breeding practices because of important socio-
476 economic developments in horse husbandry and greater management of individual
477 populations. The almost complete extinction of wild horses in much of Eurasia may have
478 been approximately concomitant with this stage: a reduction in wild populations, resulting
479 from increased habitat encroachment by humans and a warming climate, potentially
480 catalysed a reassessment of the symbolic and economic value of the domesticated form of
481 the horse. Technological developments may also have played a role in promoting change,
482 for example, with the development of the metal bit (Bendry 2011). In this complex scenario
483 of cultural innovation and environmental change, highly distinct modern morphological
484 types emerged with a major contribution, according to genetic evidence, likely due to an
485 independent second wave of domestication in Western Europe (Lindgren et al. 2004;
486 Achilli et al. 2012; Cieslak et al. 2010).

487

488 **Conclusion**

489 Domestic animals offer some of the best examples of how strong directional selection can
490 change morphology (Drake & Klingenberg 2010). Our results support the notion that
491 domestication is a highly dynamic process (Larson et al. 2014; Marshall et al. 2014), with
492 morphological changes occurring over an extended period of time (Dobney & Larson
493 2006). However, the amount of change can vary remarkably over time and more
494 importantly may vary sharply depending on the trait or even just the aspect of the character
495 being studied, as shown by our tooth size and shape data. An 'explosion' of morphological
496 disparity is expected as artificial selection became more intense in the last few centuries
497 Interestingly, however, we also show that, if horses were first domesticated around 6000
498 years ago (Outram et al. 2009), the shape of horse teeth then likely remained unchanged for
499 over half of the subsequent history of the *domestic* horse, until transformations that took
500 place over the past 2000-3000 years. To our knowledge, this has never been reported and
501 suggests a new layer of complexity, in which shape variation is decoupled from size, in the
502 study of human-animal interactions. These outcomes reinforce the singular nature of horse
503 domestication *per se*, and support the notion that the special relationship between horses
504 and humans has existed for some considerable time. Within a well-established and
505 powerful statistical framework (Adams et al. 2013), our work provides a testable

506 hypothesis, which we term the 'long-fuse' model of phenotypic change in domestication,
507 and which can be verified in new samples and on other anatomical structures, and different
508 species.

509

510 **References**

511 Adams D., Rohlf F. & Slice D. A field comes of age: geometric morphometrics in the 21st
512 century. *Hystrix, the Italian J. of Mammalogy* **24**(1), 7-15 (2013).

513

514 Achilli A. *et al.* Mitochondrial genomes from modern horses reveal the major haplogroups
515 that underwent domestication. *Proc. Natl Acad. Sci. USA* **109**(7), 2449–2454
516 doi:10.1073/pnas.1111637109 (2012).

517

518 Andersson L. S. *et al.* Mutations in *DMRT3* affect locomotion in horses and spinal circuit
519 function in mice. *Nature* **488**, 642-646 (2012).

520

521 Anthony D. W. *The Horse, the Wheel and Language*. Princeton University Press. NJ
522 (2007).

523

524 Bendrey R. Identification of metal residues associated with bit-use on prehistoric horse
525 teeth by scanning electron microscopy with energy dispersive X-ray microanalysis. *J. of*
526 *Arch. Sci.* **38**, 2989-2994 (2011).
527
528 Bendrey R. From wild horses to domestic horses: a European perspective. *World Arch.*
529 **44**(1), 135–157. doi:10.1080/00438243.2012.647571 (2012).
530
531 Bignon O. & Eisenmann V. in *Equids in Time and Space* (ed Mashkour, M.) (Oxbow
532 Books, pp.161-171, 2002)
533
534 Cardini A. in *Encyclopedia of Life Support Systems* (eds UNESCO-EOLSS Joint
535 Committee) (EOLSS 2013).
536
537 Cardini A. & Elton S. Does the skull carry a phylogenetic signal? Evolution and modularity
538 in the guenons. *Bio. J. of the Linnean Society* **93**(4), 813–834. doi:10.1111/j.1095-
539 8312.2008.01011.x (2008).
540
541 Cieslak M. *et al.* Origin and history of mitochondrial DNA lineages in domestic horses.

542 *PLoS One* **5**(12), e15311 (2010).

543

544 Clutton-Brock J. *A Natural History of Domesticated Mammals*. Cambridge University
545 Press, (1999).

546

547 Cuzzo F. P., Sauter M.L. What is dental ecology? *Am. J. Phys. Anthropol.* 148(2): 163–
548 170. doi:10.1002/ajpa.21656 (2012).

549

550 Der Sarkissian C. *et al.* Mitochondrial genomes reveal the extinct Hippidion as an outgroup
551 to all living equids. *Biol. Lett.* 11: 20141058. <http://dx.doi.org/10.1098/rsbl.2014.1058>.

552

553 Dobney K. & Larson G. Genetics and animal domestication: new windows on an elusive
554 process. *J. of Zool.* **269**(2), 261–271. doi:10.1111/j.1469-7998.2006.00042.x (2006).

555

556 Drake A. G., & Klingenberg C. P. Large scale diversification of skull shape in domestic
557 dogs: disparity and modularity. *The American Naturalist* **175**(3), 289–301 (2010).

558

559 Dryden I. L. shapes: Statistical shape analysis (Version 1.1-9). Retrieved from <http://cran.r->

560 project.org/web/packages/shapes/index.html (2013).

561

562 Eisenmann V. & Baylac M. Extant and fossil Equus (Mammalia, Perissodactyla) skulls: a
563 morphometric definition of the subgenus Equus. *Zoologica Scripta* **29**(2), 89–100.

564 doi:10.1046/j.1463-6409.2000.00034.x (2000).

565

566 Evin A. *et al.* The long and winding road: identifying pig domestication through molar size
567 and shape. *J. of Arch. Sci.* doi:10.1016/j.jas.2012.08.005 (2013)

568

569 Evans A. R. Shape descriptors as ecometrics in dental ecology. *Hystrix, the Italian J. of*
570 *Mammalogy* **24**(1), 133-140. (2013).

571

572 Foote M. The evolution of morphological diversity. *Ann. Rev. of Ecol. and Systematics*
573 **28**,129-152 (1997).

574

575 Gidley J. W. Tooth characters and revision of the North American species of the Genus
576 *Equus*. *Bull. Am. Mus. Nat. Hist.* **14**, 91-141 (1901).

577

578 Goto H. *et al.* A massively parallel sequencing approach uncovers ancient origins and high
579 genetic variability of endangered Przewalski's horses. *Genome Biol. & Evol.* **3**, 1096–1106.
580 doi:10.1093/gbe/evr067 (2011).

581

582 Groves C. P. & Ryder O. A. in *The Genetics of the Horse* (eds Bowling, A.V. & Ruvinsky,
583 A.) (CABI Publishing, pp. 1-24, 2000).

584

585 Hammer O. Harper D. & Ryan P. PAST: Paleontological statistics software package for
586 education and data analysis. *Paleontol. Electron.* **4** (1): 1–9. (2001).

587

588 Houle D. Govindaraju, D. R. & Omholt, S. Phenomics: the next challenge. *Nature Reviews*
589 */ Genetics.* **11**, 855-866 (2010).

590

591 Kaiser, T. M. & Shulz. E. Tooth wear gradients in zebras as an environmental proxy - a
592 pilot study. *Mitt. hamb. zool. Mus. Inst.* 103, S187-210.

593

594 Klingenberg C. P. Heterochrony and allometry: the analysis of evolutionary change in
595 ontogeny. *Biological Reviews* **73**(1), 79–123. doi:10.1111/j.1469-185X.1997.tb00026.x

596 (1998).
597
598 Klingenberg C. P. MorphoJ: an integrated software package for geometric morphometrics.
599 *Mol. Ecol. Res.* **11**(2), 353–357. doi:10.1111/j.1755-0998.2010.02924.x (2011).
600
601 Larson G *et al.* Current perspectives and the future of domestication studies. *Proc. Natl*
602 *Acad. Sci. USA* **111**(17), 6139-6146 doi/10.1073/pnas.1323964111 (2014).
603
604 Lindgren G. *et al.* Limited number of patriline in horse domestication. *Nature Genetics*
605 **36**(4), 335–336. doi:10.1038/ng1326 (2004).
606
607 Lippold S. *et al.* Whole mitochondrial genome sequencing of domestic horses reveals
608 incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.*
609 **11**(1), 328. doi:10.1186/1471-2148-11-328 (2011).
610
611 Ludwig A. *et al.* Coat color variation at the beginning of horse domestication. *Science*
612 **324**(5926), 485–485. doi:10.1126/science.1172750 (2009).
613

614 Marshall F. B. *et al.* Evaluating the roles of directed breeding and gene flow in animal
615 domestication. *Proc. Natl Acad. Sci. USA* **111**(17), 6153-6158
616 doi/10.1073/pnas.1323964111 (2014).

617

618 Nagorsen D. W., & Cardini A. Tempo and mode of evolutionary divergence in modern and
619 Holocene Vancouver Island marmots (*Marmota vancouverensis*) (Mammalia, Rodentia). *J.*
620 *of Zoo. Syst. and Evol. Res.* **47**(3), 258–267. doi:10.1111/j.1439-0469.2008.00503.x (2009).

621

622 Olsen, S.L. in *Domestication: New Genetic and Archaeological Paradigms* (eds Zeder, M.,
623 Bradley, D.G., Emshwiller, E., Smith, B.D.) (University of California Press, pp.245-269,
624 2006).

625

626 Outram A. K *et al.* The earliest horse harnessing and milking. *Science* **323**, 1332-1335
627 (2009).

628

629 R Development Core Team. R: a language and environment for statistical computing.
630 (<http://www.r-project.org>) (2005).

631

632 Rohlf F. J, & Slice, D. Extensions of the Procrustes method for the optimal superimposition
633 of landmarks. *Syst. Zoo.* **39**(1), 40–59. doi:10.2307/2992207 (1990).
634

635 Rohlf F. The tps series of software. *Hystrix.* **26**(1), 9-12 (2015).
636

637 Rohlf F. NTSYSpc: Numerical taxonomy system, ver. 2.1. *Setauket, New York: Exeter*
638 *Publishing, Ltd* (2013).
639

640 Seetah T. K. Cardini A, & Miracle P. T. Can morphospace shed light on cave bear spatial-
641 temporal variation? Population dynamics of *Ursus spelaeus* from Romualdova pećina and
642 Vindija, (Croatia). *J. of Arch. Sci.* **39**(2), 500–510. doi:10.1016/j.jas.2011.10.005 (2012).
643

644 Seetah K. Cucchi, T., Dobney, K., & Barker, G. A geometric morphometric re-evaluation
645 of the use of dental form to explore differences in horse (*Equus caballus*) populations and
646 its potential zooarchaeological application. *J. of Arch. Sci.* **41**, 904–910 (2013).
647

648 Steiner C. C. Makova, K. D. & Ryder, O. A. Mitochondrial genome: Clues about the
649 evolution of extant equids and genomic diversity of horse breeds. *Equine Genomics* 311-

650 321 (2013).

651

652 Viscosi V. & Cardini A. Leaf morphology, taxonomy and geometric morphometrics: a
653 simplified protocol for beginners. *PLoS ONE* **6**(10), e25630.

654 doi:10.1371/journal.pone.0025630 (2011).

655

656 Wade C. M. *et al.* Genome sequence, comparative analysis, and population genetics of the
657 domestic horse. *Science* **326**(5954), 865–867. doi:10.1126/science.1178158 (2009).

658

659 Warmuth, V. *et al.* European domestic horses originated in two Holocene refugia. *PLoS*
660 *ONE* **6**(3), e18194. doi:10.1371/journal.pone.0018194 (2011).

661

662 Warmuth V. *et al.* Reconstructing the origin and spread of horse domestication in the
663 Eurasian steppe. *Proc. Natl Acad. Sci. USA* **109**(21), 8202–8206.

664 doi:10.1073/pnas.1111122109 (2012).

665

666 **Conflicting interests**

667 We have no conflicting interests.

668

669 **Author contributions**

670 KS & AC designed the study. GB oversaw the study. KS coordinated and collected
671 samples, and performed initial landmarking. AC performed statistical analysis and
672 validation. KS & AC wrote the article with essential input from GB. All authors gave final
673 approval for publication.

674

675 **Funding statement**

676 Funding was provided to GB from the Leverhulme Trust project grant scheme (F/09 757/B)
677 and to KS & AC from the Lang Fund for Human-Environmental Anthropology, Dept. of
678 Anthropology, Stanford.

679

680 **Acknowledgments**

681 We are extremely grateful to the following for facilitating access to materials and providing
682 essential input for specific samples: ICE and PRZ: F. Mayer (Museum für Naturkunde,
683 Berlin); THB and PRZ: R. Sabin & L. Tomsett (Natural History Museum, London); CHI:
684 T. Yaqi & H. Songmei (Shaanxi Archaeological Research Institute, Xi'an), KAZ: Z.
685 Samashev & K. Kashkanbajev (Institute of Oriental Studies, Almaty), RUS: P. Kotsinsev

686 (Institute of Ecology of Plants and Animals, Yekaterinburg) & B. Hanks (Department of
687 Archaeology, Pittsburgh); CRO: D. Brajkovic (Institute for Quaternary Palaeontology and
688 Geology, Zagreb) & P. Miracle (Department of Archaeology, Cambridge) and HUN: A.
689 Choyke (Department of Medieval Studies, Budapest) & A. Endrődi (Budapest History
690 Museum, Budapest). We thank N. Vibla and J. Li for assistance in recovering samples and
691 translation; A. Evin for providing a set of duplicate samples from Berlin and T. Cucchi and
692 K. Dobney for essential analytical input. We are deeply grateful to I. Dryden (University of
693 Nottingham), P. Gunz (Max Planck Institute for Evolutionary Anthropology, Leipzig), L.
694 Monteiro (Universidade Estadual do Norte Fluminense), S. Schlager (University of
695 Freiburg) and other MORPHMET subscribers for help and suggestions with R scripts and
696 packages; to A. Drake for discussing with us methodological and conceptual issues about
697 morphological disparity in dogs; to P. Mitteroecker (University of Vienna) for his
698 important feedback on how to estimate multivariate variance; to P.D. Polly (Indiana
699 University, Bloomington) for discussions on teeth evolution and shape analysis which
700 greatly improved our interpretations. Finally, we are sincerely grateful to P.D. Polly and an
701 anonymous referee for insightful comments on an early draft of this manuscript.