

1 **Proto-consonants were information-dense via identical bioacoustic**  
2 **tags to proto-vowels**

3  
4 Adriano R. Lameira<sup>1\*</sup>, Raquel Vicente<sup>2</sup>, António Alexandre<sup>2</sup>, Gail Campbell-Smith<sup>3</sup>,  
5 Cheryl Knott<sup>4</sup>, Serge Wich<sup>5,6</sup>, Madeleine E. Hardus<sup>2</sup>

6  
7 <sup>1</sup> Evolutionary Anthropology Research Group, Department of Anthropology, Durham  
8 University, UK

9 <sup>2</sup> Independent researcher

10 <sup>3</sup> International Animal Rescue, West Kalimantan, Indonesia

11 <sup>4</sup> Department of Anthropology, Boston University, USA

12 <sup>5</sup> School of Natural Science and Psychology, Liverpool John Moores University, UK

13 <sup>6</sup> Faculty of Science, University of Amsterdam, the Netherlands

14 \* Corresponding author: [adriano.lameira@durham.ac.uk](mailto:adriano.lameira@durham.ac.uk)

15  
16  
17 **Why did our ancestors combine the first consonant- and vowel-like utterances to**  
18 **produce the first syllable or word? To answer this question, it is essential to**  
19 **know what constituted the communicative function of proto-consonants vs.**  
20 **proto-vowels before their combined use became universal. Close to nothing is**  
21 **known, however, about consonant-like calls in the primate order<sup>1,2</sup>. Here, we**  
22 **investigate a large collection of voiceless consonant-like calls in nonhuman great**  
23 **apes – our closest relatives – namely orangutans (*Pongo* spp.). We analyzed 4486**  
24 **kiss-squeaks collected across 48 individuals in four wild populations. Despite**  
25 **idiosyncratic production mechanics, consonant-like calls displayed information-**  
26 **dense content and the same acoustic signatures found in nonhuman primate**  
27 **voiced vowel-like calls, implying similar biological functions. Selection regimes**  
28 **between proto-consonants and -vowels were, thus, probably indistinguishable at**  
29 **the dawn of spoken language evolution. Our findings suggest that the first proto-**  
30 **syllables or -words in our lineage probably constituted message reiterations,**  
31 **instead of messages of increasing intricacy.**

32  
33  
34 Primate vocal behavior is a cornerstone in the theory of speech evolution<sup>3</sup>. Vocal  
35 homologies between human and nonhuman primates provide potential paths for the  
36 evolution of spoken language in humans<sup>4</sup> and several vocal traits exhibit evolutionary  
37 continuity between human and nonhuman primate (hereafter *primate*) vocal systems<sup>5</sup>.  
38 Primate literature has hitherto focused almost exclusively on primate voiced calls, or  
39 “vocalizations”. That is, utterances that feature vocal fold action, namely vocal folds’  
40 regular oscillation as sound source<sup>6</sup>. Voiced calls characterize primate, and indeed  
41 mammalian repertoires as a whole, and they survive today in human speech  
42 predominantly in the form of vowels (as well as non-linguistic utterances, such as  
43 laughter and crying). Accordingly, voiced calls likely date back to a mammalian  
44 ancestor that lived some 125 million-years-ago<sup>7</sup>, 80Mya before the last common  
45 ancestor of all primates, some 45Mya<sup>8</sup>.

46 Marginal theoretical attention and empirical effort have, however, been  
47 dedicated to voiceless calls<sup>9,10</sup>. Voiceless calls (e.g. smacks, clicks, raspberries),  
48 unlike their voiced counterparts, do not result from vocal fold action but instead from  
49 supra-laryngeal maneuvering. This feature renders them homologous in terms of  
50 articulation and acoustics to voiceless utterances in humans, which primarily function

51 as consonants – the second basic building block of human spoken language besides  
52 vowels. Voiceless calls among primates are present in some Old World monkey  
53 species (in the form of lip-smacks) and in great apes. In great apes, voiceless calls  
54 have been reported in all genera, suggesting shared ancestry<sup>1</sup>. Accordingly, voiceless  
55 calls can be presumed to descend, *at least*, from the last great ape common ancestor,<sup>1,2</sup>  
56 dating back some 10 million-years-ago<sup>11,12</sup>. The current state of knowledge raises,  
57 hence, a disquieting possibility – speech evolution theory may have remained  
58 incomplete up until now, since it has strictly drawn on evidence on primate voiced  
59 calls, and thus, simply on aspects pertinent to vowel use and evolution. Only the  
60 integrated study of consonant-like primate calls will ultimately allow answering  
61 critical questions about human behavior and spoken language evolution. For instance,  
62 why were the first consonant- and vowel-like calls combined to generate the first  
63 syllable- and word-like utterance?

64 Here, we address this gap in our knowledge within the theoretical edifice of  
65 human behavior and spoken language evolution by examining how early human  
66 ancestors adaptively used consonant-like calls. Specifically, we ask whether the use of  
67 voiceless calls could have transmitted the same type(s) of communicative content as  
68 voiced vowel-like calls (insofar their acoustics were fundamentally different from the  
69 latter). Notably, four major types of acoustic variation have been described in primate  
70 voiced calls. Primate voiced calls may function to transmit information on population  
71 membership<sup>13</sup>, individual body size<sup>14</sup>, individuality (ID)<sup>15</sup> and call context<sup>16</sup>.  
72 Ultimately, assessing the presence of these levels of acoustic variation in great ape  
73 voiceless calls will allow inferring the selective regimes and, tacitly, the potential  
74 biological functions that underpinned the evolution of proto-consonants within the  
75 human lineage in comparison with proto-vowels.

76 Orangutans (*Pongo* spp.), the earliest diverging great ape lineage, provide an  
77 ideal model species to address these open questions. Orangutans are unique among  
78 nonhuman primates in that the predominant call type produced across populations –  
79 the “kiss-squeak” – is voiceless<sup>9,17</sup>. These calls rely exclusively on lip and airflow  
80 coordination for vocal production, alike labial consonants in humans (e.g. /p/). Kiss-  
81 squeaks represent alarm calls<sup>9,17</sup> and the lack of apparent voiceless homologues in  
82 other nonhuman great apes<sup>18</sup> suggests that they probably represent derived calls in the  
83 orangutan lineage. Additionally, orangutans exhibit an overall repertoire of voiceless  
84 calls richer than what has been so far described in other nonhuman great apes<sup>17,19,20</sup>.  
85 These data suggest recurrent events of voiceless call emergence in *Pongo*, suggesting  
86 that voiceless calls may have indeed evolved to fulfill biological functions in this  
87 lineage<sup>9,10,21</sup>. Hence, this makes orangutan call repertoire an attractive model system  
88 to assess the selective forces shaping voiceless call emergence and use in hominids.  
89 Moreover, kiss-squeaks in orangutans are often combined with a voiced alarm call  
90 (the “grumph”) to produce a voiceless-voiced call combination<sup>17</sup>. This configuration  
91 is in direct articulatory parallel with human consonant-vowel syllables and supports,  
92 therefore, the view that these voiceless calls provide a desirable empirical window  
93 into proto-consonant use in human ancestors. We do not propose evolutionary  
94 continuity between orangutan kiss-squeaks and any specific human consonant.  
95 Instead, we investigate kiss-squeaks as model calls homologous to the precursors of  
96 consonants. We assume that these calls in orangutans have stemmed from an  
97 evolutionary process equivalent to that that gave rise to proto-consonants in early  
98 humans in the past. We are specifically interested in the moment in speech evolution  
99 when consonant-like and vowel-like calls were available within our lineage but not  
100 yet predominantly used in combination.

101 We conducted generalized linear models to examine the informational content  
 102 of orangutan kiss-squeaks. All levels of acoustic variation (population, body size  
 103 class, individual ID and context) were included as factors/variables in two models. In  
 104 either model, the response variable corresponded to one of two measured acoustic  
 105 parameters that summarized voiceless calls along the frequency and time axes:  
 106 maximum frequency (Hz) and duration (s), respectively. Results revealed that each  
 107 variable produced a significant effect on our response variables. Namely, orangutan  
 108 body size class significantly affected orangutan kiss-squeaks' maximum frequency,  
 109 context affected the calls' duration, and population membership and individual ID  
 110 affected both acoustic parameters simultaneously (Table 1). Figure 1 shows the data  
 111 distribution per level of variation and respective group centroids (i.e. the centers of  
 112 distribution for each population/size class/individual/context). Group centroids were  
 113 typically separated at each level by frequency differences in the order of several  
 114 hundreds of Hz and by time gaps in the order of 0.1 and 0.01 seconds. Along both  
 115 frequency and time axes, confidence intervals for each group centroid rarely  
 116 overlapped with those of another group.

117 These models were controlled for repeated sampling of call recordings from  
 118 the same individuals and populations (i.e. they were treated as random variables), for  
 119 the nested effect of individuals within population, and the models were offset for the  
 120 effect of recording distance between the microphone and the subject. Results indicate  
 121 that orangutan voiceless calls exhibit frequency and time signatures directly resulting  
 122 from biologically meaningful factors indicating where (population), when (context)  
 123 and who (size class and individual ID) produced the call.  
 124

125 Table 1. Comparison of the full model (with all fixed and random variables) to  
 126 reduced models (each excluding one variable).

| Excluded variable       | Maximum frequency |                |                  | Duration       |                  |
|-------------------------|-------------------|----------------|------------------|----------------|------------------|
|                         | Df                | X <sup>2</sup> | Pr (>Chi)        | X <sup>2</sup> | Pr (>Chi)        |
| Population <sup>1</sup> | 1                 | 7.0779         | <b>0.0078</b>    | 19.788         | <b>&lt;0.001</b> |
| Size class <sup>2</sup> | 2                 | 51.652         | <b>&lt;0.001</b> | 0.2382         | 0.8877           |
| Individual <sup>1</sup> | 1                 | 583.95         | <b>&lt;0.001</b> | 1199.1         | <b>&lt;0.001</b> |
| Context <sup>2</sup>    | 4                 | 1.8234         | 0.7682           | 45.737         | <b>&lt;0.001</b> |

127 <sup>1</sup>Random variable

128 <sup>2</sup>Fixed variable

129

130 Our results demonstrate that voiceless consonant-like calls in great apes  
 131 exhibit rich acoustic variation and clear acoustic signatures. Namely, two prime  
 132 acoustic parameters (max frequency and duration) in orangutan kiss-squeaks are  
 133 significantly affected by population, size class, context, and individual ID. These  
 134 results show that the acoustic profile of voiceless consonant-like calls in primates can  
 135 be loaded with biologically meaningful information. These same four factors  
 136 constitute the major levels along which voiced vowel-like calls vary. This parallel  
 137 indicates that consonant-like calls are potentially as adaptive as vowel-like calls,  
 138 despite being *at least* 35Mya (and 70Mya) younger among primates (and mammals).  
 139 In other words, consonant-like calls and variation therein most likely allowed early  
 140 human ancestors to adaptively use voiceless consonant-like calls much as they would  
 141 use voiced vowel-like calls.

142 In bioacoustics, communicative function is subserved by acoustic variation.  
 143 Our results show that voiceless consonant-like calls display similar levels of variation  
 144 known for voiced vowel-like calls. Therefore, we tentatively propose that the

145 communicative functions of both call categories are probably equal. Since consonant-  
146 like calls vary along the same levels as vowel-like calls, individuals are in fact  
147 prevented of endowing each call category with different types of message. In order to  
148 directly confirm call function, future playback experiments will need to verify if  
149 orangutans extract information from the different levels of variation in voiceless calls.  
150 Nevertheless, to our knowledge, it has never been demonstrated that primate calls  
151 exhibit variation that conspecific receivers are not sensitive to or do not assess. It is  
152 strongly predicted that, if this level of variation exists in orangutan voiceless calls,  
153 then, receivers will likely gauge it in a functional way in some measure.

154 The parallel found between variation in voiceless consonant- and voiced  
155 vowel-like calls was detected even though consonant-like calls exhibit distinct  
156 production mechanisms. Specifically, orangutan kiss-squeaks are the result of lip and  
157 air flow control, other than the result of vocal fold action followed by a filter, as is the  
158 case in voiced calls<sup>6</sup>. This result indicates that the both laryngeal and the supra-  
159 laryngeal anatomy of the primate vocal tract can independently imprint the same  
160 acoustic signatures onto their respective acoustic output.

161 Our results align with the frame/content theory, perhaps the most renowned  
162 hypothesis granting equivalent roles to consonant and vowel production in the process  
163 of speech evolution<sup>22</sup>. This hypothesis poses that speech derived from primate  
164 behaviors encompassing close and open cycles of the mouth, associated with  
165 consonant and vowel production, respectively, with each full open-close cycle  
166 corresponding to the production of a syllable. Previously described great ape vocal  
167 behavior<sup>1</sup> and our results suggest that both consonant- and vowel-like calls were  
168 already in use separately before their concatenation to form syllables and words. For  
169 example, previous evidence from an orangutan who learned a new voiced and  
170 voiceless call shows that both categories can be produced at a speech-like rhythm of  
171 close-open mouth cycles<sup>20</sup>. As such, it is conceivable that the fast alternation of close-  
172 open cycles during speech-production *today*, recruited in the past fast ancient primate  
173 mouth behaviors (such as lip-smacking<sup>23</sup> or suckling) as a means of greatly  
174 accelerating the delivery of consonant- and vowel-like calls already present in the  
175 species' repertoire.

176 If similar selection pressures acted on communication in early humans and  
177 early orangutans, our findings suggest that, at the dawn of speech evolution, proto-  
178 consonants were information-dense. They were molded by similar selective regimes  
179 as proto-vowels and are predicted to have fulfilled similar communicative functions.  
180 Since both call categories evolved to become the two building blocks of all the  
181 world's spoken languages, it is perhaps unsurprising that both categories were  
182 originally equivalent in terms of variation and putative function. This view implies,  
183 however, that the reason of the first early human ancestors for having combined  
184 proto-consonants and -vowels to generate the first proto-syllable or -word was not  
185 based on functional disparity. That is, a consonant-vowel combination would have  
186 served poorly to transmit two different bits of information. To transmit different  
187 messages, one of the two categories ought to vary in ways the other did not, but such  
188 proposition did not find support in our results.

189 Conversely, elaboration and redundancy are common mechanisms of  
190 adaptation in animal acoustic systems that ensure effective communication<sup>24</sup>.  
191 Fulfilling effective vocal communication could therefore pose a parsimonious and  
192 proximate explanation for the production of the first proto-syllables or -words.  
193 Namely, the combination of voiceless consonant-like calls and voiced vowel-like  
194 would have allowed better exploiting the sound spectrum for the transmission of the

195 same cue or bit of information. Proto-syllables probably represented, therefore,  
196 message reiterations.

197 New research investigating nonhuman great ape voiceless calls and their  
198 comparison with voiced calls allows refining our understanding of consonant and  
199 vowel use by early human ancestors. This information will allow drawing pertinent  
200 extrapolations about the evolutionary drives and synergies that played out between  
201 speech building blocks before and after the emergence of the first syllables and  
202 words.

203

## 204 **Methods**

205

206 **Study sites.** This study was conducted across four research stations, two in Borneo  
207 (*P. pygmaeus wurmbii*) – Tuanan and Gunung Palung – and two in Sumatra (*P.*  
208 *abelii*) – Sikundur and Sampan Getek. This study comprised 2510 observation hours  
209 at Tuanan, 1520 at Gunung Palung, 1132 at Sikundur and 498 at Sampan Getek, with  
210 a grand total of 5660 observation hours.

211

212 **Data collection.** All orangutan kiss-squeaks were opportunistically recorded while  
213 following subjects, typically at 7 to 30 meters distance from the individuals. Only  
214 kiss-squeaks unaided variants were addressed in the study because other variants are  
215 only present in some populations (i.e. hand and leaf kiss-squeaks were not  
216 considered)<sup>9,10</sup>. Calls were recorded at Tuanan using Marantz Analogue Recorder  
217 PMD222 (Marantz, Corporation, Kenagawa, Japan) in combination with a Sennheiser  
218 Microphone ME 64 (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany)  
219 or a Sony Digital Recorder TCD-D100 in combination with a Sony Microphone  
220 ECM-M907 (Sony Corporation, Tokyo, Japan). In all remaining sites, calls were  
221 recorded using a Marantz Analogue Recorder PMD-660 or a ZOOM H4next Handy  
222 Recorder (ZOOM Corporation, Tokyo, Japan), both connected with a RODE NTG-2  
223 directional microphone (RØDE LLC, Sidney, Australia). Audio data were recorded  
224 under Wave format at 16 bit. No meaningful differences in audio input were expected  
225 to result from different professional microphones (see below). Audio recordings were  
226 collected simultaneously with complete focal behavioral data on the focal animals and  
227 other conspecifics when in association. Data collection involved no interaction with  
228 or handling of the animals and strictly followed the Indonesian law.

229

230 **Data analyses.** Recordings were transferred to a computer with a sampling rate of  
231 44.1kHz. Kiss-squeaks were measured with Raven Interactive Sound Analysis  
232 Software (version 1.2.1, Cornell Lab of Ornithology, Ithaca, NY) using the  
233 spectrogram window (Window type: Hann; 3 dB filter bandwidth: 124Hz; grid  
234 frequency resolution: 2.69Hz; grid time resolution: 256 samples). Two acoustic  
235 parameters were measured following previous studies<sup>9,15</sup>: maximum frequency (Hz)  
236 and duration (seconds). Maximum frequency represented the frequency with the  
237 highest amplitude (dB) in the call. Duration represented the time difference between  
238 the off and onset of the call. Both parameters were extracted directly from the  
239 spectrogram window by drawing a selection encompassing the complete call from  
240 onset to offset.

241

242 These two parameters were chosen for four main reasons. First, they capture  
243 the general profile of a call along the time and frequency domains, respectively.  
244 Second, these two parameters have demonstrated to be highly informative, indeed  
oftentimes the *most* informative among other parameters and at different levels of

245 variation in primate voiced calls, including orangutans<sup>15,16,25</sup>. Third, both parameters  
246 are extractable from voiced and voiceless calls, allowing a direct comparison in terms  
247 of levels of variation between the two call categories. Forth, these parameters are  
248 extremely robust and resilient across different recording settings and equipment,  
249 whereas other parameters are not<sup>19</sup>.

250 In order to establish the presence of each type of variation (between  
251 populations, size classes, contexts and individuals) potentially present in orangutan  
252 voiceless calls, we conducted generalized linear mixed model analyses (GLMM)  
253 using R<sup>26</sup> and the function lmer of the R-package lme4<sup>27</sup>. Our two acoustic parameters  
254 – max frequency and duration – represented the response variable of two separate  
255 models. “Size class” factor comprised 3 levels (i.e. adolescent, adult, large flanged-  
256 male morph) and “context” 5 levels (i.e. towards other orangutans, other animals,  
257 observers, other humans, and predator models), and were inserted in our models as  
258 fixed effects. Because individuals and populations were sampled repeatedly, these  
259 factors were considered random effects, with “population” factor exhibiting 4 levels  
260 (i.e. 4 different populations) and “individual” factor 48 levels (i.e. 48 different  
261 individuals).

262 Our factor “individual” was nested in “population”. That is, no individual  
263 belonged simultaneously to two different populations. In order to structure our  
264 GLMM most accurately with regards to our data, we directly tested whether there was  
265 any difference between explicitly indicating the nested effect in our model or not.  
266 These test models simply included our response variable as predicted by individual ID  
267 and population. There was a null difference between a model that explicitly indicated  
268 the nested effect (via “/” or “%in%”) and a model that did not (Supplementary  
269 material). As such, for a matter of simplicity and because this had no effect  
270 whatsoever on model performance, our full model did not explicitly indicate the  
271 nested effect of “individual” within “population”.

272 Variation between sexes was not considered in our analyses for two reasons.  
273 Male/female ratio in frequency (Hz) in orangutan calls has been shown to be one of  
274 the nearest to 1 among primates, particularly among great apes<sup>28</sup>. Second, oftentimes,  
275 sex differences in primate calls are primarily the result of body size differences and  
276 our model already included body size as a fixed effect. Had we included sex and body  
277 size simultaneously, this would have disrupted model performance due to co-linearity.

278 Before running the models, we verified whether recording distance (meters)  
279 from the orangutan individuals affected our response variables. These analyses were  
280 strictly exploratory. For both max frequency and duration, we observed a significant  
281 effect of recording distance (Spearman test, maximum frequency: n = 4447, rho = -  
282 0.211, p < 0.001; duration: n = 4426, rho = 0.307, p < 0.001). For this reason, we  
283 inserted recording distance in both models as an offset variable.

284  
285 **Data availability.** The data that support the findings of this study are available from  
286 the corresponding author upon request.

287  
288

## 289 **References**

290

- 291 1. Lameira, A. R., Maddieson, I. & Zuberbuhler, K. Primate feedstock for the  
292 evolution of consonants. *Trends in cognitive sciences* **18**, 60–62 (2014).
- 293 2. Lameira, A. R. The forgotten role of consonant-like calls in theories of speech  
294 evolution. *Behavioral and Brain Sciences* **37**, 559–560 (2014).
- 295 3. Fitch, W. T. The evolution of speech: a comparative review. *Trends in Cognitive*

- 296 *Sciences* **4**, 258–267 (2000).
- 297 4. Fitch, W. T. & Zuberbu hler, K. in *The Evolution of Emotional Communication:*  
298 *From Sounds in Nonhuman Mammals to Speech and Music in Man* (eds. Altenm u  
299 ller, E., Schmidt, S. & Zimmermann, E.) 26–48 (Oxford University Press, 2013).
- 300 5. Lemasson, A., Ouattara, K. & Zuberbuhler, K. in *The Evolutionary Emergence of*  
301 *Language* (eds. Botha, R. & Everaert, M.) 181–203 (Oxford University Press,  
302 2013).
- 303 6. Taylor, A. M. & Reby, D. The contribution of source–filter theory to mammal vocal  
304 communication research. *Journal of Zoology* **280**, 221–236 (2010).
- 305 7. Luo, Z.-X., Ji, Q., Wible, J. R. & Yuan, C.-X. An Early Cretaceous Tribosphenic  
306 Mammal and Metatherian Evolution. *Science* **302**, 1934–1940 (2003).
- 307 8. Ni, X. *et al.* The oldest known primate skeleton and early haplorhine evolution.  
308 *Nature* **498**, 60–64 (2013).
- 309 9. Hardus, M. E., Lameira, A. R., van Schaik, C. P. & Wich, S. A. Tool use in wild orang-  
310 utans modifies sound production: a functionally deceptive innovation?  
311 *Proceedings of the Royal Society B: Biological Sciences* **276**, 3689–3694 (2009).
- 312 10. Lameira, A. R. *et al.* Population-specific use of the same tool-assisted alarm call  
313 between two wild orangutan populations (*Pongo pygmaeus wurmbii*) indicates  
314 functional arbitrariness. **8**, e69749 (2013).
- 315 11. Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H. & Mailund, T. Incomplete  
316 lineage sorting patterns among human, chimpanzee, and orangutan suggest  
317 recent orangutan speciation and widespread selection. *Genome Research* **21**,  
318 349–356 (2011).
- 319 12. White, T. D., Lovejoy, C. O., Asfaw, B., Carlson, J. P. & Suwa, G. Neither chimpanzee  
320 nor human, *Ardipithecus* reveals the surprising ancestry of both. *Proceedings of*  
321 *the National Academy of Sciences of the United States of America* **112**, 4877–4884  
322 (2015).
- 323 13. Wich, S. A., Schel, A. M. & de Vries, H. Geographic variation in Thomas langur  
324 (*Presbytis thomasi*) loud calls. *American journal of primatology* **70**, 566–574  
325 (2008).
- 326 14. Fitch, W. T. Vocal tract length and formant frequency dispersion correlate with  
327 body size in rhesus macaques. *Journal of the Acoustical Society of America* **102**,  
328 1213–1222 (1997).
- 329 15. Lameira, A. R. & Wich, S. Orangutan Long Call Degradation and Individuality Over  
330 Distance: A Playback Approach. *International Journal of Primatology* **29**, 615–625  
331 (2008).
- 332 16. Spillmann, B. *et al.* Acoustic properties of long calls given by flanged male orang-  
333 utans (*Pongo pygmaeus wurmbii*) reflect both individual identity and context.  
334 *Ethology* **116**, 385–395 (2010).
- 335 17. Hardus, M. E. *et al.* in *Orangutans* (eds. S. Wich, T. Mitra Setia, S.S. Utami & Schaik,  
336 C. P.) 49–60 (Oxford University Press, 2009).
- 337 18. Salmi, R., Hammerschmidt, K. & Doran-Sheehy, D. M. Western Gorilla Vocal  
338 Repertoire and Contextual Use of Vocalizations. *Ethology* n/a–n/a (2013).  
339 doi:10.1111/eth.12122
- 340 19. Lameira, A. R. *et al.* Orangutan (*Pongo* spp.) whistling and implications for the  
341 emergence of an open-ended call repertoire: A replication and extension. *Journal*  
342 *of the Acoustical Society of America* **134**, 1–11 (2013).
- 343 20. Lameira, A. R. *et al.* Speech-like rhythm in a voiced and voiceless orangutan call.  
344 **10**, e116136 (2015).
- 345 21. De Boer, B., Wich, S. A., Hardus, M. E. & Lameira, A. R. Acoustic models of  
346 orangutan hand-assisted alarm calls. *The Journal of experimental biology* **218**,  
347 907–914 (2015).
- 348 22. MacNeilage, P. F. The frame/content theory of evolution of speech production.  
349 *Behavioral and Brain Sciences* **21**, 499–511 (1998).

- 350 23. Ghazanfar, A. A., Takahashi, D. Y., Mathur, N. & Fitch, W. T. Cineradiography of  
351 monkey lip-smacking reveals putative precursors of speech dynamics. *Current*  
352 *biology : CB* **22**, 1176–1182 (2012).
- 353 24. Waser, P. M. & Brown, C. H. Habitat acoustics and primate communication.  
354 *American journal of primatology* **10**, 135–154 (1986).
- 355 25. Delgado, R. A. *et al.* Geographical variation in orangutan long calls. 215–224  
356 (2009).
- 357 26. R Team, D. C. R: *A Language and Environment for Statistical Computing*. (R  
358 Foundation for Statistical Computing, 2010).
- 359 27. Bates, D., Maechler, M. & Dai, B. *Lme4: Linear-Mixed Effects Models Using S4*  
360 *Classes*. (2008).
- 361 28. Puts, D. A. *et al.* Sexual selection on male vocal fundamental frequency in humans  
362 and other anthropoids. *Proceedings. Biological sciences / The Royal Society* **283**,  
363 20152830 (2016).
- 364

### 365 **Acknowledgements**

366 We are grateful to the Indonesian Institute of Science (LIPI), the Indonesian Ministry  
367 of Research and Technology (RISTEK), the Indonesian Directorate General of Forest  
368 Protection and Nature Conservation (PHKA), Gunung Palung National Park Bureau  
369 (BTNGP) and Gunung Leuser National Park (TNGL) and Leuser Ecosystem  
370 Management Authority (BPKEL) for authorization to carry out research in Indonesia.  
371 We thank Universitas Nasional (UNAS), Tanjungpura University (UNTAN) and  
372 Universitas Sumatera Utara (USU) for supporting the project and acting as counter-  
373 partner. Bornean Orangutan Survival (BOS, Palangka Raya, Central Kalimantan),  
374 Sumatran Orangutan Conservation Programme (SOCP, Medan, North Sumatra) and  
375 Gunung Palung Orangutan Project (GPOCP, Ketapang, West Kalimantan) acted as  
376 sponsors. We thank Marie-Claire Pagano for technical support. Roger Mundry and  
377 Jeremy Kendal provided valuable input on the design of the generalized linear mixed  
378 models, as well as Heidi Collieran and Séan Roberts at the 1<sup>o</sup> Quantitative Methods  
379 Spring School 2016 at the Max Plank Institute for the Science of Human History,  
380 Jena, Germany. The funders had no role in study design, data collection and analysis,  
381 decision to publish, or preparation of the manuscript.

### 382 **Contributions**

383 AL conceived and designed the study. AL, RV, AA and MH collected data. AL, RV,  
384 AA and MH analyzed data. AL, GC, CK, SW contributed materials/data collection  
385 and analysis tools. AL, GC, CK, SW and MH wrote the manuscript.

### 386 **Competing interests**

387 The authors reveal no conflicting interests.

### 388 **Corresponding author**

389 Correspondence to Adriano R. Lameira.

390

### 391 **Figure legends**

392

393 Fig 1. Scatterplot representing orangutan kiss-squeaks along maximum frequency  
394 (Hz) and duration (s) (A) per population, (B) per size class, (C) per context for the  
395 Tuanan population, and (D) per individual for the Sampan Getek population. Large  
396 circles represent group centroids with vertical and horizontal error bars representing  
397 the 95% confidence interval (represented in A and B in magnified window).

