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

Adriano R. Lameira[1]*, Raquel Vicente[2], António Alexandre[2], Gail Campbell-Smith[3], Cheryl Knott[4], Serge Wich[5,6], Madeleine E. Hardus[2]

[1] Evolutionary Anthropology Research Group, Department of Anthropology, Durham University, UK
[2] Independent researcher
[3] International Animal Rescue, West Kalimantan, Indonesia
[4] Department of Anthropology, Boston University, USA
[5] School of Natural Science and Psychology, Liverpool John Moores University, UK
[6] Faculty of Science, University of Amsterdam, the Netherlands
* Corresponding author: adriano.lameira@durham.ac.uk

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

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

Marginal theoretical attention and empirical effort have, however, been dedicated to voiceless calls[9,10]. Voiceless calls (e.g. smacks, clicks, raspberries), unlike their voiced counterparts, do not result from vocal fold action but instead from supra-laryngeal maneuvering. This feature renders them homologous in terms of articulation and acoustics to voiceless utterances in humans, which primarily function as consonants – the second basic building block of human spoken language besides vowels. Voiceless calls among primates are present in some Old World monkey species (in the form of lip-smacks) and in great apes. In great apes, voiceless calls have been reported in all genera, suggesting shared ancestry[1]. Accordingly, voiceless calls can be presumed to descend, *at least*, from the last great ape common ancestor,[1,2] dating back some 10 million-years-ago[11,12]. The current state of knowledge raises, hence, a disquieting possibility – speech evolution theory may have remained incomplete up until now, since it has strictly drawn on evidence on primate voiced calls, and thus, simply on aspects pertinent to vowel use and evolution. Only the integrated study of consonant-like primate calls will ultimately allow answering critical questions about human behavior and spoken language evolution. For instance, why were the first consonant- and vowel-like calls combined to generate the first syllable- and word-like utterance?

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

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

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

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

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

		Maximum frequency		Duration	
Excluded variable	Df	X^2	Pr (>Chi)	X^2	Pr (>Chi)
Population[1]	1	7.0779	**0.0078**	19.788	**<0.001**
Size class[2]	2	51.652	**<0.001**	0.2382	0.8877
Individual[1]	1	583.95	**<0.001**	1199.1	**<0.001**
Context[2]	4	1.8234	0.7682	45.737	**<0.001**

[1] Random variable

[2] Fixed variable

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

In bioacoustics, communicative function is subserved by acoustic variation. Our results show that voiceless consonant-like calls display similar levels of variation known for voiced vowel-like calls. Therefore, we tentatively propose that the communicative functions of both call categories are probably equal. Since consonant-like calls vary along the same levels as vowel-like calls, individuals are in fact prevented of endowing each call category with different types of message. In order to directly confirm call function, future playback experiments will need to verify if orangutans extract information from the different levels of variation in voiceless calls. Nevertheless, to our knowledge, it has never been demonstrated that primate calls exhibit variation that conspecific receivers are not sensitive to or do not assess. It is strongly predicted that, if this level of variation exists in orangutan voiceless calls, then, receivers will likely gauge it in a functional way in some measure.

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

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

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

Conversely, elaboration and redundancy are common mechanisms of adaptation in animal acoustic systems that ensure effective communication[24]. Fulfilling effective vocal communication could therefore pose a parsimonious and proximate explanation for the production of the first proto-syllables or -words. Namely, the combination of voiceless consonant-like calls and voiced vowel-like would have allowed better exploiting the sound spectrum for the transmission of the same cue or bit of information. Proto-syllables probably represented, therefore, message reiterations.

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

Methods

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

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

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

These two parameters were chosen for four main reasons. First, they capture the general profile of a call along the time and frequency domains, respectively. Second, these two parameters have demonstrated to be highly informative, indeed oftentimes the *most* informative among other parameters and at different levels of variation in primate voiced calls, including orangutans'[15,16,25]. Third, both parameters are extractable from voiced and voiceless calls, allowing a direct comparison in terms of levels of variation between the two call categories. Forth, these parameters are extremely robust and resilient across different recording settings and equipment, whereas other parameters are not[19].

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

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

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

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

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

References

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

Sciences **4,** 258–267 (2000).

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

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

Acknowledgements

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

Contributions

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

Competing interests

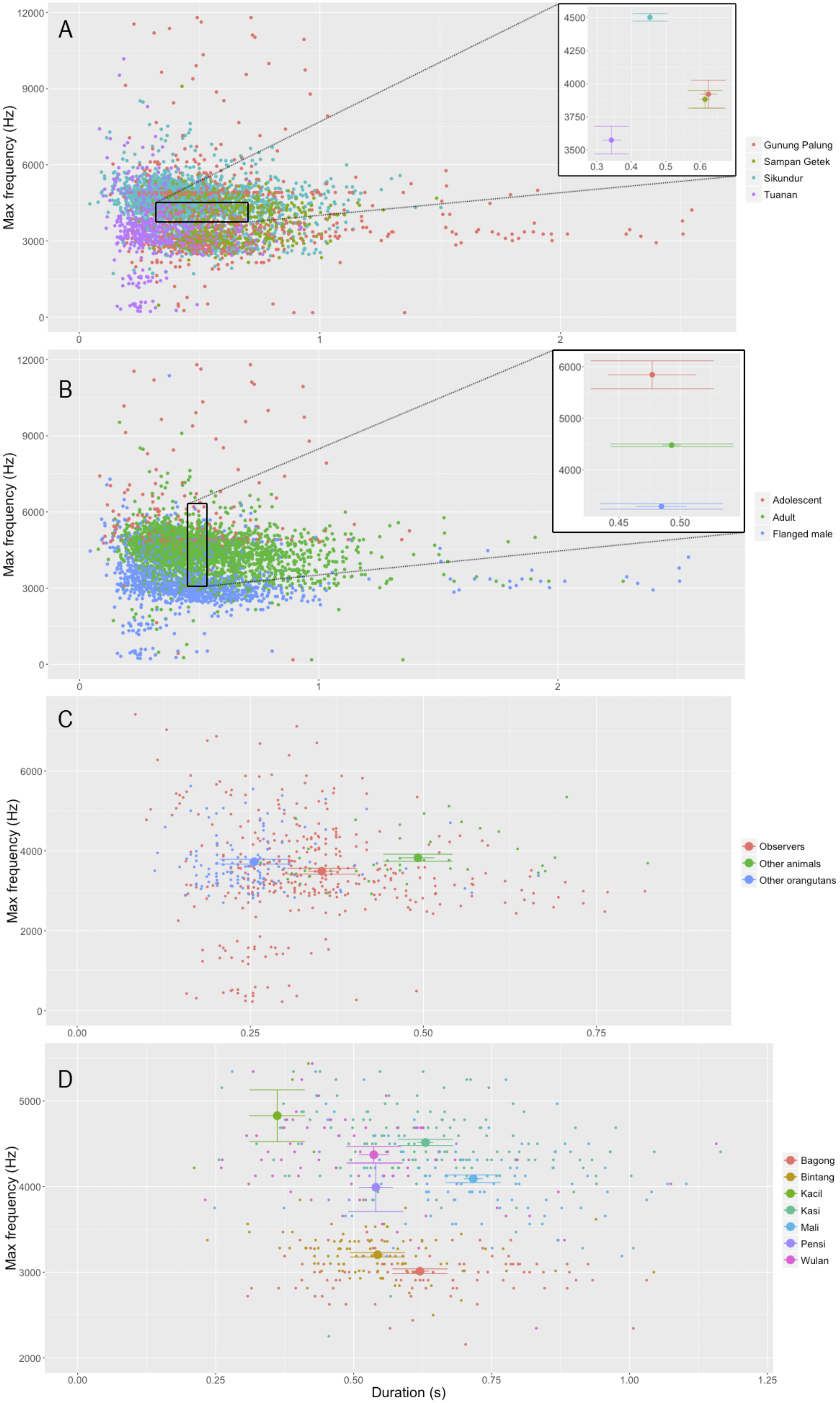
The authors reveal no conflicting interests.

Corresponding author

Correspondence to Adriano R. Lameira.

Figure legends

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


A
Max frequency (Hz)
12000
9000
6000
3000
0
0
1
2
4500
4250
4000
3750
3500
0.3
0.4
0.5
0.6
Gunung Palung
Sampan Getek
Sikundur
Tuanan
B
Max frequency (Hz)
12000
9000
6000
3000
0
0
1
2
6000
5000
4000
0.45
0.50
Adolescent
Adult
Flanged male
C
Max frequency (Hz)
6000
4000
2000
0
0.00
0.25
0.50
0.75
Observers
Other animals
Other orangutans
D
Max frequency (Hz)
5000
4000
3000
2000
0.00
0.25
0.50
0.75
1.00
1.25
Duration (s)
Bagong
Bintang
Kacil
Kasi
Mali
Pensi
Wulan