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

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

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 and of proto-vowels before their combined use became universal. Almost 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 analysed 4,486 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 voiced vowel-like calls by nonhuman primates, implying similar biological functions. Selection regimes between proto-consonants and proto-vowels were thus probably indistinguishable at the dawn of spoken language evolution. Our findings suggest that the first proto-syllables or proto-words in our lineage probably constituted message reiterations, instead of messages of increasing intricacy.

Primate vocal behaviour 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'. By this, we mean utterances that feature vocal-fold action, namely the regular oscillation of vocal folds, as a 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 probably date back to a mammalian ancestor that lived around 125 million years ago[7], 80 million years before the last common ancestor of all primates some 45 million years ago[8].

Little theoretical attention and empirical effort have, however, been dedicated to voiceless calls in primates[9,10]. Voiceless calls (such as smacks, clicks and raspberries), unlike their voiced counterparts, do not result from vocal-fold action but instead from supralaryngeal manoeuvring. 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 beside 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 common ancestor of the great apes[1,2], dating back to about 10 million years ago[11,12]. The current state of knowledge thus raises a disquieting possibility: speech evolution theory may have remained incomplete until now, as 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 answers to critical questions about human behaviour 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 behaviour and spoken language evolution by examining how consonant-like calls were adaptively used by early human ancestors. Specifically, we ask whether the use of voiceless calls could have transmitted the same type(s) of communicative content as voiced vowel-like calls (despite the fact that their acoustics were fundamentally different from the latter). Notably, four main 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 voiceless calls by great apes will allow researchers to infer 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, like labial consonants in humans (for example /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 evolved to fulfil biological functions in this lineage[9,10,21]. Hence, this makes orangutan call repertoire an attractive model system to assess the selective forces shaping the emergence and use of

[1]Evolutionary Anthropology Research Group, Department of Anthropology, South Road, Durham DH1 3LE, UK. [2]Jalan S. Parman Gang Tomat, No. 18B, RT02, RW01, Sukaharja, Ketapang, West Kalimantan 78813, Indonesia. [3]Department of Anthropology, Boston University, 232 Bay State Road, Boston, Massachusetts 02215, USA. [4]School of Natural Science and Psychology, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK. [5]Faculty of Science, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. [†]Unaffiliated: raquelvicente22@gmail.com (R.V.); antoniovargas04@hotmail.com (A.A.); contact@madeleinehardus.com (M.E.H.). *e-mail: adriano.lameira@durham.ac.uk

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

Excluded variable	DF	Maximum frequency		Duration	
		χ^2	$P(>\chi)$	χ^2	$P(>\chi)$
Population*	1	7.0779	0.0078	19.788	<0.001
Size class†	2	51.652	<0.001	0.2382	0.8877
Individual*	1	583.95	<0.001	1,199.1	<0.001
Context†	4	1.8234	0.7682	45.737	<0.001

*Random variable. †Fixed variable. DF, degrees of freedom.

voiceless calls 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 therefore supports 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 the one that gave rise to proto-consonants in early humans. 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. Population, body-size class, individual ID and context were included as factors or 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: orangutan body-size class significantly affected the maximum frequency of orangutan kiss-squeaks, context affected the duration of the calls, 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 (the centres of distribution for each population/size class/individual/context). Group centroids were typically separated at each level by frequency differences of several hundreds of hertz and by time gaps of the order of 0.1 to 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 (that is, they were treated as random variables), and 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 by whom (size class and individual ID) the call was produced.

Our results demonstrate that voiceless consonant-like calls in great apes exhibit rich acoustic variation and clear acoustic signatures. Namely, two prime acoustic parameters (maximum frequency and duration) in orangutan kiss-squeaks are significantly affected by population, size class, context and individual ID. These are the same main 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 35 million years (and 70 million years) 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 fulfilled by acoustic variation. Our results show that voiceless consonant-like calls display similar levels of variation to those 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 from endowing each call category with different types of message. To confirm call function directly, future playback experiments will need to verify which information orangutans extract from voiceless calls. Nevertheless, to our knowledge, primate calls do not exhibit variation to which conspecific receivers are insensitive or that they do not assess. It is strongly predicted that, if this level of variation exists in orangutan voiceless calls, then receivers will probably 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, rather than the result of vocal-fold action followed by a filter, as is the case in voiced calls[6]. This result indicates that both the 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 was derived from primate behaviors encompassing closed and open cycles of the mouth, associated with consonant and vowel production, respectively, with each full open–closed cycle corresponding to the production of a syllable. Our results, and previously described vocal behaviour in great apes[1], 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 closed–open mouth cycles[20]. It is therefore conceivable that the fast alternation of closed–open cycles seen in modern speech-production recruited, in the past, rapid mouth behaviours (such as lip-smacking[23] or suckling) in ancient primates, 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 moulded by selective regimes similar to those for 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 the two categories were originally equivalent in terms of variation and putative function. This view implies, however, that the reason for the first early human ancestors to combine proto-consonants and proto-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 categories ought to vary in ways the other did not, but this proposition is not supported by 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
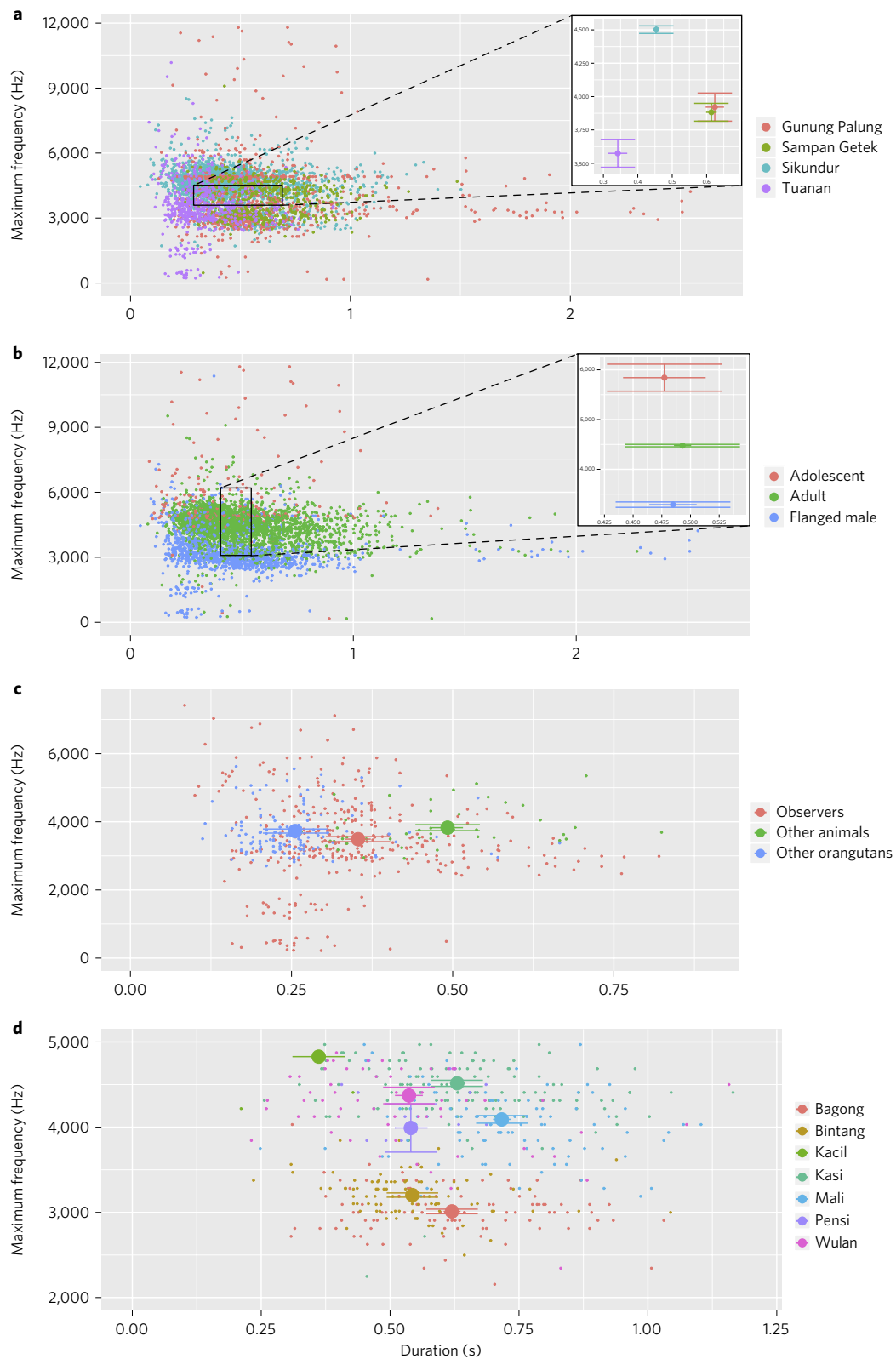
Figure 1 | Scatter plot representing orangutan kiss-squeaks along maximum frequency and duration. a–d, Per population (**a**); per size class (**b**); per context for the Tuanan population (**c**); and per individual for the Sampan Getek population (**d**). Large circles represent group centroids, with vertical and horizontal error bars representing the 95% confidence interval (represented in **a** and **b** in the magnified window).

explanation for the production of the first proto-syllables or -words. The combination of voiceless consonant-like calls and voiced vowel-like calls would have allowed better exploitation of the sound spectrum for the transmission of the same cue or bit

of information. Proto-syllables therefore probably represented message reiterations.

Our new research investigating voiceless calls in nonhuman great apes and their comparison with voiced calls refines our

understanding of consonant and vowel use by early human ancestors. This information will allow pertinent extrapolations to be drawn 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, Tuanan and Gunung Palung in Borneo (*P. pygmaeus wurmbii*), and Sikundur and Sampan Getek in Sumatra (*P. abelii*). This study comprised 2,510 observation hours at Tuanan, 1,520 at Gunung Palung, 1,132 at Sikundur and 498 at Sampan Getek, with a grand total of 5,660 observation hours.

Data collection. All orangutan kiss-squeaks were opportunistically recorded while following subjects, typically at distances of 7 to 30 m from the individuals. Only unaided kiss-squeak variants were addressed in the study because other variants are present in only some populations (that is, hand and leaf kiss-squeaks were not considered)[9,10]. Calls were recorded at Tuanan using a 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 to a RODE NTG-2 directional microphone (RØDE LLC, Sydney, Australia). Audio data were recorded under WAVE/WAV 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 behavioural 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 44.1 kHz. Kiss-squeaks were measured with Raven interactive sound analysis software (version 1.2.1, Cornell Lab of Ornithology, Ithaca, New York) using the spectrogram window (window type: Hann; 3-dB filter bandwidth: 124 Hz; grid frequency resolution: 2.69 Hz; 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 often the most informative among other parameters and at different levels of variation in primate voiced calls, including those of orangutans[15,16,25]. Third, both parameters can be extracted from voiced and voiceless calls, allowing a direct comparison in terms of levels of variation between the two call categories. Fourth, these parameters are extremely robust and resilient across different recording settings and equipment, whereas other parameters are not[19].

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 as the programming language[26] and using the function lmer of the R-package lme4[27]. Our two acoustic parameters — maximum frequency, and duration — represented the response variable of two separate models. The 'size class' factor comprised three classes (adolescent, adult, and large flanged-male morph) and 'context' five classes (towards other orangutans, other animals, observers, other humans, and predator models), and these were inserted in our models as fixed effects. Because individuals and populations were sampled repeatedly, these factors were considered random effects, with the 'population' factor exhibiting four levels (that is, four different populations) and 'individual' factor 48 levels (48 different individuals).

Our factor 'individual' was nested in 'population'. That is, no individual belonged simultaneously to two different populations. To structure our GLMM most accurately with regard to our data, we directly tested whether there was any difference between explicitly indicating the nested effect in our model and 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 (through "/" or "%in%") and a model that did not (see Supplementary Information). Therefore, for simplicity and because it had no effect 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 is one of the nearest to 1 among primates, particularly among great apes[28]. Second, sex differences in primate calls are often 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 owing to co-linearity.

Before running the models, we verified whether recording distance (metres) from the orangutan individuals affected our response variables. These analyses were strictly exploratory. For both maximum frequency and duration, we observed a significant effect of recording distance (Spearman test, maximum frequency: $n = 4,447$, $rho = -0.211$, $P < 0.001$; duration: $n = 4,426$, $rho = 0.307$, $P < 0.001$). 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 Cogn. Sci.* **18,** 60–62 (2014).
2. Lameira, A. R. The forgotten role of consonant-like calls in theories of speech evolution. *Behav. Brain Sci.* **37,** 559–560 (2014).
3. Fitch, W. T. The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4,** 258–267 (2000).
4. Fitch, W. T. & Zuberbühler, K. in *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (eds Altenmüller, E., Schmidt, S. & Zimmermann, E.) 26–48 (Oxford Univ. Press, 2013).
5. Lemasson, A., Ouattara, K. & Zuberbühler, K. in *The Evolutionary Emergence of Language* (eds Botha, R. & Everaert, M.) 181–203 (Oxford Univ. Press, 2013).
6. Taylor, A. M. & Reby, D. The contribution of source–filter theory to mammal vocal communication research. *J. Zool.* **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? *Proc. R. Soc. B* **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 (*Pongopygmaeus wurmbii*) indicates functional arbitrariness. *PLoS ONE* **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 Res.* **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. *Proc. Natl Acad. Sci. USA* **112,** 4877–4884 (2015).
13. Wich, S. A., Schel, A. M. & de Vries, H. Geographic variation in Thomas langur (*Presbytis thomasi*) loud calls. *Am. J. Primatol.* **70,** 566–574 (2008).
14. Fitch, W. T. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* **102,** 1213–1222 (1997).
15. Lameira, A. R. & Wich, S. Orangutan long call degradation and individuality over distance: a playback approach. *Int. J. Primatol.* **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: Geographic Variation in Behavioral Ecology and Conservation* (eds Wich, S., Mitra Setia, T., Utami, S. S. & Schaik, C. P.) 49–60 (Oxford Univ. Press, 2009).
18. Salmi, R., Hammerschmidt, K. & Doran-Sheehy, D. M. Western gorilla vocal repertoire and contextual use of vocalizations. *Ethology* **119,** 831–847 (2013).
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. *J. Acoust. Soc. Am.* **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. *J. Exp. Biol.* **218,** 907–914 (2015).
22. MacNeilage, P. F. The frame/content theory of evolution of speech production. *Behav. Brain Sci.* **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. *Curr. Biol.* **22,** 1176–1182 (2012).
24. Waser, P. M. & Brown, C. H. Habitat acoustics and primate communication. *Am. J. Primatol.* **10,** 135–154 (1986).

25. Delgado, R. A. *et al.* in *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds Wich, S., Mitra Setia, T., Utami, S. S. & Schaik, C. P.) 215–224 (Oxford Univ. Press, 2009).
26. R Development CoreTeam. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2010); http://www.gbif.org/resource/81287
27. Bates, D., Maechler, M. & Dai, B. *lme4: Linear-Mixed Effects Models Using S4 Classes* (2008); http:// cran.r-project.org/web/packages/lme4/lme4.pdf
28. Puts, D. A. *et al.* Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proc. R. Soc. B* **283,** 20152830 (2016).

Author contributions
A.R.L. conceived and designed the study. A.R.L., R.V., A.A. and M.E.H. collected data. A.R.L., R.V., A.A. and M.E.H. analysed data. A.R.L., G.C.-S., C.K. and S.W. contributed collection of materials and data, and analysis tools. A.R.L., G.C.-S., C.K., S.W. and M.E.H. wrote the manuscript.

Additional information
Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.R.L.

How to cite this article: Lameira, A. R. *et al.* Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels. *Nat. Hum. Behav.* **1,** 0044 (2017).

Competing interests
The authors declare no competing interests.