A humanized version of Foxp2 does not affect ultrasonic vocalization in adult mice

K. Hammerschmidt[*,†]**, C. Schreiweis**[‡,1]**,**
C. Minge[†]**, S. Pääbo**[§]**, J. Fischer**[†,2]** and**
W. Enard[¶,2]

[†] *Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany ,* [‡] *Institut Necker-Enfants Malades (INEM), Centre de Médecine Moléculaire, CS Paris, France ,* [§] *Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, and* [¶] *Department of Biology II, Ludwig Maximilian University Munich, Martinsried, Germany*
[1] *Shared first author.*
[2] *Shared last author.*

Corresponding author: K. Hammerschmidt, Cognitive Ethology Laboratory, German Primate Center, D-37077 Göttingen, Germany. E-mail: khammerschmidt@dpz.eu

The transcription factor FOXP2 has been linked to severe speech and language impairments in humans. An analysis of the evolution of the *FOXP2* gene has identified two amino acid substitutions that became fixed after the split of the human and chimpanzee lineages. Studying the functional consequences of these two substitutions in the endogenous *Foxp2* gene of mice showed alterations in dopamine levels, striatal synaptic plasticity, neuronal morphology and cortico-striatal-dependent learning. In addition, ultrasonic vocalizations (USVs) of pups had a significantly lower average pitch than control littermates. To which degree adult USVs would be affected in mice carrying the 'humanized' Foxp2 variant remained unclear. In this study, we analyzed USVs of 68 adult male mice uttered during repeated courtship encounters with different females. Mice carrying the *Foxp2*[hum/hum] allele did not differ significantly in the number of call elements, their element structure or in their element composition from control littermates. We conclude that neither the structure nor the usage of USVs in adult mice is affected by the two amino acid substitutions that occurred in FOXP2 during human evolution. The reported effect for pup vocalization thus appears to be transient. These results are in line with accumulating evidence that mouse USVs are hardly influenced by vocal learning. Hence, the function and evolution of genes that are necessary, but not sufficient for vocal learning in humans, must be either studied at a different phenotypic level in mice or in other organisms.

Keywords: Communication, evolution, FOXP2, learning, mice, speech, ultrasound, vocalization

Humans are the only primates that have developed volitional control over their sound production. Monkeys and apes lack the ability to imitate sounds vocally. Further, nonhuman primates do not need auditory input to develop their species-specific vocalization structure (Ackermann *et al.* 2014; Egnor & Hauser 2004; Hammerschmidt & Fischer 2008; Jürgens 2009). Except for some few anecdotal findings, which suggest that captive apes may show some limited capacity for vocal production learning (Wich *et al.* 2012), the majority of studies showed that vocal learning does not play a role in the normal development of nonhuman primate vocalizations (Arriaga *et al.* 2012; Fitch 2010; Hammerschmidt & Fischer 2008). In terrestrial mammals and humans, the basic neurological and muscular systems necessary for vocalizations overlap to a large degree (Ackermann *et al.* 2014; Jürgens 2002). The present view is that a direct projection from the motor cortex to the laryngeal motor neurons, similar to the direct forebrain projection to vocal motor neurons in birds, is necessary for the ability to gain control over vocal production. Monkeys and apes lack such a direct projection (Arriaga *et al.* 2012; Hage *et al.* 2013; Simonyan 2014; Simonyan & Jürgens 2003).

One of the pressing questions is therefore to understand how the neurological basis for volitional vocal control emerged during evolution. Potentially, the tuning of cortico-striatal circuits, similar to the anterior forebrain pathway in birds with its specialized striatal region Area X, might have paved the way for recruiting neural circuits for vocal learning (Enard 2011; Murugan *et al.* 2013; Pfenning *et al.* 2014; Simonyan *et al.* 2012). In the search for the genetic basis that gave rise to such adaptations, *FOXP2* is a strong candidate gene. Humans that are heterozygous for a non-functional *FOXP2* allele suffer from an impairment that especially affects their speech and language development (Lai *et al.* 2001; Macdermot *et al.* 2005; Vargha-Khadem *et al.* 2005; Watkins 2011). Analyses of the evolution of the *FOXP2* gene showed two amino acid substitutions, which became fixed in the human lineage after its separation from the chimpanzee. It was suggested that these two substitutions underwent positive selection potentially due to their effect on speech and language (Enard *et al.* 2002; Zhang *et al.* 2002).

To test the functional consequences of these two substitutions, Enard and colleagues established a mouse model, which carries these substitutions in the endogenous mouse *Foxp2* gene (Enard *et al.* 2009). Mice homozygous for this humanized *Foxp2* allele (*Foxp2*[hum/hum]) show specific alterations in the cortico-basal ganglia circuits, including changes to tissue dopamine levels, striatal synaptic plasticity, dendrite morphology and gene expression (Enard 2011; Enard *et al.* 2009; Reimers-Kipping *et al.* 2011; Schreiweis *et al.* 2014). Strikingly, pups carrying the humanized *Foxp2* gene
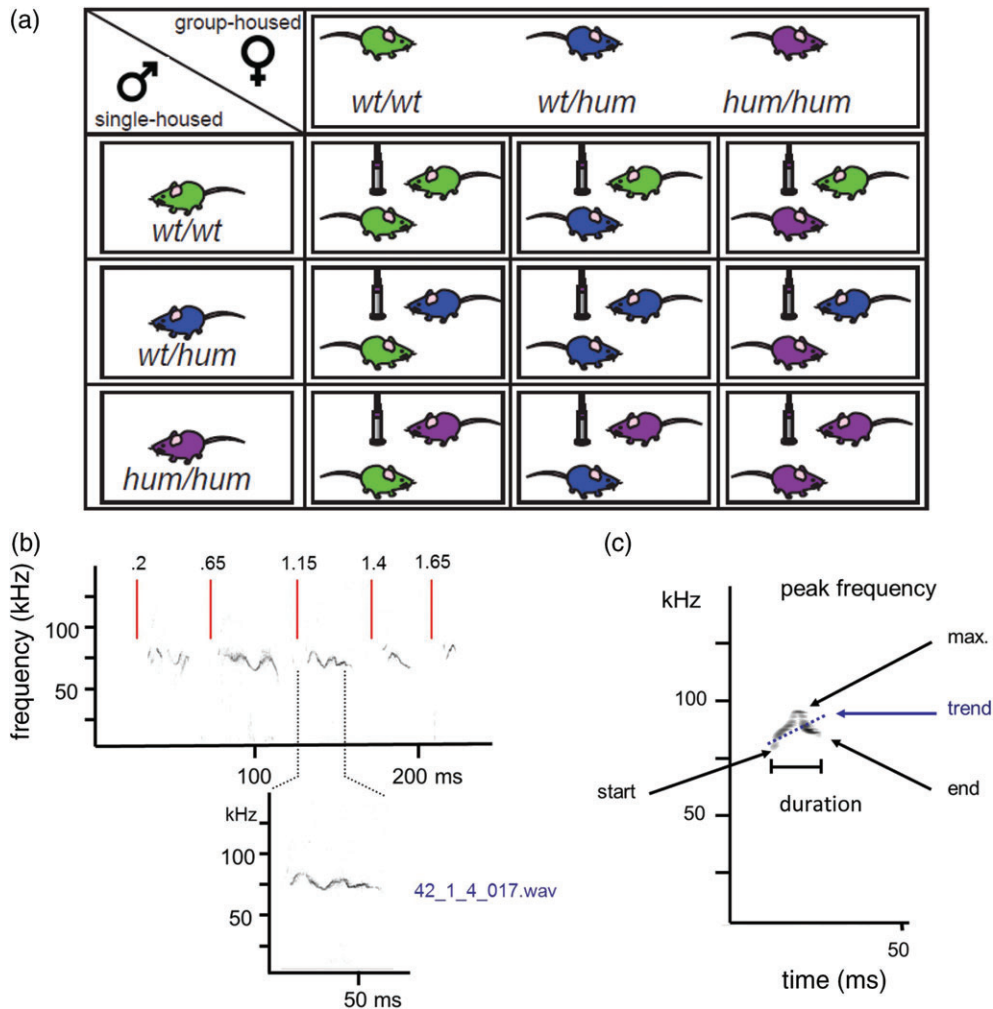
Figure 1: Measuring the effect of humanized Foxp2 on adult courtship songs. (a) Setup of recording experiment in which USVs of one male per genotype were recorded while encountering one female per genotype in a balanced random order. (b) Analysis of vocal behavior: temporal calculation and element selection. (c) Analysis of call element structure.

differed in their ultrasonic vocalizations (USVs). Although both *Foxp*$^{hum/hum}$ pups and their wild-type littermates were able to produce isolation calls within the normal strain-specific range and did not differ in their calling activity, the isolation calls themselves showed significant differences in the level of peak frequency (main energy), i.e. in the pitch of their USVs (Enard *et al.* 2009).

Importantly, the available evidence suggests that the structure of mouse vocalizations is largely innate and that auditory learning is not a prerequisite for the production of species (or strain) typical vocalizations (Ehret 2005; Hammerschmidt *et al.* 2012b, 2015). Nevertheless, it might be that the humanized version of *Foxp2* affects the motor patterns sub-serving vocal production. If this is the case, the differences between adult mice carrying the humanized *Foxp2* version and their wild-type littermates should be more pronounced because adult mice produce far more complex USVs than pups do

(Fischer & Hammerschmidt 2011; Holy & Guo 2005). To explore this possibility, we studied the courtship vocalization of adult males by exposing resident males of different *Foxp2* genotypes to encounters with females (Fig. 1a). We compared the structure of the emitted USVs among genotypes and specifically predicted more elaborate temporal and/or spectral patterns in *Foxp2*$^{hum/hum}$ male mice.

Materials and methods

Mice and recording setup
We used mice carrying the humanized *Foxp2* allele (*Foxp2*^{hum}) before the neomycin cassette was removed [derived from the Bl6 ES cell 5H10, as described in the study of Enard *et al.* (2009)] and mice carrying the humanized *Foxp2* allele after the neomycin cassette had been removed [*Foxp2*^{humΔneo}; 5H10 line as described in the study of Enard *et al.* (2009)]. We used these two lines as the presence of

Table 1: Number of males per genotype and line

	Foxp2^{hum}	*Foxp2*^{humΔneo}	Sum
wild-type	11	10	21
knockin	12	11	23
heterozygous	12	12	24

a neomycin cassette can in principle have effects in addition to the two amino acid changes and to make maximal use of the available mice in the colony. A total of 144 virgin mice, aged 9–16 weeks, were grouped into 24 analysis units of six mice each (13 units/line *Foxp2*^{hum} and 11 units/line *Foxp2*^{humΔneo}). Each unit was composed of one homozygous, one heterozygous and one wild-type male and female mice. These were matched as much as possible for age and litter among the genotypes and housed in groups of two to five individuals (for number of males per genotype and line, see Table 1). Approximately 48 h prior to testing, the three females of one unit were group-housed in a clean cage and received dirty bedding from cages containing the males of the unit. Approximately 24 h prior to testing, males were single-housed. Vocalization recordings took place in a separate testing room after a short habituation period between 0800 and 1700 h. The lid, food grid and water bottle of the home cage containing the single-housed male were removed and the cage was put into a sound-proof plexiglass chamber (custom-made), which was cleaned with an anti-odorant disinfectant (Decosept, Borer Chemie, Zuchwil, Switzerland). To obtain a more consistent call quality, we reduced the home cage surface by inserting a plexiglass separator into the male home cage and positioned the ultrasound microphone (CM16) centrally above the remaining home cage surface. Ultrasonic vocalizations have a highly confined active space and the plexiglass separator limited the distance between the male and the microphone. After 120 seconds of habituation, the male was recorded for 30 seconds (control recording) before a female was added to the cage, the chamber was closed and ultrasonic courtship songs were recorded for 2 min. We decided to use this time frame because preliminary experience (pre-study) had indicated that subjects frequently fall silent after the first 2 min. The order of male/female encounters of each unit were pseudo-randomly assigned so that each male (i.e. of each genotype) encountered each female (i.e. of each genotype) once and no animal was tested twice consecutively (Fig. 1a). Units of each line were recorded on separate days, except for the last recording day, where two units of the *Foxp2*^{humΔneo} line and one unit of the *Foxp2*^{hum} line were recorded. For the recordings, the microphone was connected to a preamplifier (UltraSoundGate 116, hardware and software: Avisoft Bioacoustics, Glienicke, Germany), which was connected to a computer using the recording software AVISOFT SASLAB PRO v4.33 with a sampling frequency of 300 kHz (hardware and software: Avisoft Bioacoustics, Glienicke, Germany).

Mice were housed at a 10 h/14 h dark/light cycle under standard conditions in isolated ventilated racks (lights on at 0700 h). All animal work was performed in accordance with guidelines of the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, and federal regulations of Saxony, Germany.

Analysis of element structure

The sampling frequency of 300 kHz resulted in a frequency range of 150 kHz. To reduce background noise at lower frequencies, which could negatively influence the estimation of whistle tracking algorithm, we used a high-pass FIR filter of 35 kHz. We used the whistle tracking algorithm of AVISOFT SAS LAB PRO 5.2 (Avisoft Bioacoustics) with the following settings: monotonic, maximum change per step 8 pix = 4.7 kHz, minimum continuity = 8 milliseconds, hold time = 15 milliseconds. These criteria were compared with previous analyses of male mouse vocalizations (Fischer & Hammerschmidt 2011; Hammerschmidt *et al.* 2012a; Jamain *et al.* 2008). Based on these settings, we calculated the number of given call elements, inter-element interval (ICI: from the end

of an element to the start of the next element) and element duration.

For the analysis of call structure, we used the AVISOFT RECORDER 4.2 (Avisoft Bioacoustics) to store the single elements as separated wav files. We used the same whistle tracking algorithm with the same settings as described above. From the stored elements, we calculated spectrograms (frequency range: 150 kHz, frequency resolution: 293 Hz, time resolution: 0.21 milliseconds). The spectrograms were submitted to the custom software program LMA 2012 [developed by K. Hammerschmidt, see Fischer *et al.* (2013)] to estimate characteristic acoustic parameters (Fig. 1b,c). As mice typically concentrate the energy of their calls into one small frequency band, so-called whistles or pure tone-like sounds (see Figs. 1c, S1), we focused on the peak frequency, i.e. the loudest frequency in the spectrum, which corresponds in most cases to the fundamental frequency (F0). Mice often produce soft sounds, and minor head movements as well as close body contact can lead to strong amplitude fluctuations in USVs. Therefore, we used LMA to control the estimated peak frequency visually and excluded incorrectly estimated elements from the analysis.

For each call element, we determined the duration and the duration of amplitude gaps. We defined the start of an element when the sound energy of a time segment exceeded 10% of the mean maximum amplitude of the respective element. We used the same criterion to determine the end of an element. 'Duration' was calculated as the difference between these two measurements in milliseconds. An amplitude gap within an element is defined if the sound energy of a time segment falls below 10% of the mean maximum amplitude. The parameter 'amplitude gap' represents the sum of all time segments below the 10% threshold multiplied by the time resolution in milliseconds. Furthermore, we determined the start (PF start), maximum peak frequency (PF max), as well as the greatest difference in peak frequency between two consecutive 0.21-millisecond bins (so-called frequency jumps; PF jump). In addition, we calculated the location of the maximum frequency measured in relation to element duration ((1/duration) × location). To describe the element modulation, we calculated the slope of a linear trend ('element slope') through the peak frequencies of consecutive 0.21-millisecond bins.

To achieve a balanced distribution of elements per subject, we selected 150 elements per subject of the 120-second recordings (50 elements within every 40-second interval). In total, we analyzed 24 661 out of 85 635 recorded elements from 68 males. To measure structural differences between genotypes, we first compared the acoustic parameters across the entire vocal repertoire, i.e. irrespective of the specific element type ('unpartitioned vocal repertoire'). Next, we conducted the analysis at the level of different element types. We used a two-step cluster analysis (IBM SPSS STATISTICS 21, Armonk, New York, USA) with the log-likelihood distance measure and the Schwarz-Bayesian information criterion (BIC) to find the cluster solution with the best fit. First, we tested a three-cluster solution, which was the solution with the best BIC value, like in former mouse studies (Hammerschmidt *et al.* 2012a,2012b). In addition, we chose the six-cluster solution, which turned out to be the solution with the highest number of cluster by same silhouette coefficient ($S_C = 0.3$). Cluster solutions with a higher number of clusters dropped clearly under S_C of 0.3. Silhouette values ranged between −1.0 and 1.0 and values exceeding 0.5 are usually considered as solid solutions, silhouette values below 0.3 are seen as insufficient solutions (Rousseeuw 1987).

Element composition

To test whether mice carrying the humanized *Foxp2* differ in the structure of the calls, we analyzed the sequential composition. We did this at an even finer grained level of analysis, using a visual classification. We did this for two reasons: firstly, the automated characterization is only suited for calls with a high signal-to-noise ratio. The exclusion of all call elements that were disturbed by background noise, for instance, would have hampered the sequential analysis. Secondly, these finer categories made a type II error less likely. The visual classification was similar to previous descriptions of male mouse courtship vocalizations (Ey *et al.* 2012; Scattoni *et al.* 2008). We classified the elements of a call into 19 element types. In addition, we grouped these element types into three general element categories (Fig. S1). We calculated the percentage of given element types
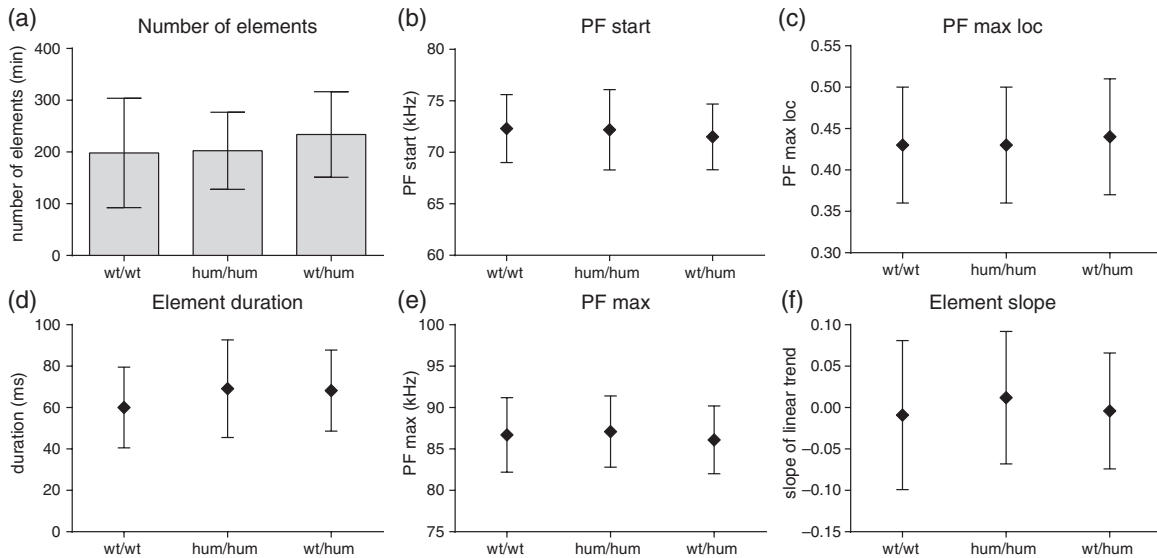
Figure 2: Influence of the humanized Foxp2 allele on call structures. number of given elements, their duration and four different frequency parameters are shown (mean ± SD).

(element type and general element category), transition rates of both categorizations, as well as chain lengths per subject and encounter.

Further statistical procedures

All data were analyzed using an LMM (IBM SPSS STATISTICS 21, USA), including genotype and line (*Foxp2^hum^* and *Foxp2^humΔneo^*) as fixed factors, and the interaction between genotype and line. For the description of general vocal behavior across all call types (unpartitioned repertoire), we included encounter order as a fixed factor. Whenever we used repeated recordings per subject, we included subject as random factor. We corrected all P-values for multiple testing using Simes correction.

Results

The 68 male subjects produced a total of 85 635 elements during 202 encounters, resulting in a mean element rate of 212 elements per minute. The number of elements did not differ among genotypes and between the two lines. We found no significant differences regarding genotype (*wt/wt*; *wt/hum* and *hum/hum*) or line (*Foxp2^hum^* and *Foxp2^humΔneo^*) in the number of given elements (genotype: $F_{2,62} = 0.88$, $P = 0.42$, line: $F_{1,62} = 0.013$, $P = 0.91$; Fig. 2), or in the mean ICI (genotype: $F_{2,57} = 0.64$, $P = 0.531$, line: $F_{1,77} = 0.151$, $P = 0.627$). We found no significant differences in mean element duration between genotypes ($F_{2,60} = 1.73$, $P = 0.187$; Fig. 2), but mice from the *Foxp2^humΔneo^* line produced significantly shorter elements than mice from the *Foxp2^hum^* line ($F_{1,60} = 12.16$; $P = 0.001$; mean ± SE: *Foxp2^hum^* $= 44 \pm 2$ milliseconds, *Foxp2^humΔneo^* $= 36 \pm 2$ milliseconds).

Furthermore, we found significant differences in relation to the order of encounters. All males were tested three times with three different females. Males produced significantly higher numbers of elements in the first

encounter (mean ± SE first encounter: 500.3 ± 18.8, second: 417.5 ± 18.8, third: 350.8 ± 18.9; $F_{2,126} = 35.9$, $P < 0.001$), with significantly shorter ICIs (mean ± SE first encounter: 255 ± 22 milliseconds, second: 318 ± 23 milliseconds, third: 340 ± 22 milliseconds; $F_{2,118} = 4.75$, $P = 0.01$), and a significantly shorter mean duration (mean ± SE first encounter: 33 ± 1 milliseconds, second: 40 ± 1 milliseconds, third: 46 ± 1 milliseconds; $F_{2,121} = 49.74$, $P < 0.001$). As the element frequency decreased, the element duration increased, such that the total amount of USVs (the amount of time a vocalization was produced) showed no significant differences ($F_{2,56} = 0.85$, $P = 0.433$).

To exclude the possibility that acoustic differences would be restricted to elements with a short duration, as in the case of pup isolation calls (Enard *et al.* 2009), we conducted an additional analysis, focusing on short elements

Table 2: Comparison of the results of the statistical analyses for pups and adult mice. Acoustic parameters and the outcome of the statistical test (*P*-values) are given. *P*-values are corrected for multiple testing (Simes)

Acoustic parameters	Pup study (Enard *et al.* 2009)	Adult male mice
Duration (milliseconds)	n.s.	n.s.
Amplitude gap (milliseconds)	n.s.	n.s.
PF start (Hz)	<0.001	n.s.
PF max (Hz)	<0.001	n.s.
PF jump (Hz)	<0.05	n.s.
PF max loc	n.s.	n.s.
Slope of trend	<0.001	n.s.

n.s., non-significant.

Table 3: Significant differences of unpartitioned data set. Columns 2–4 show *P*-values for difference in genotype, line and encounter succession. *P*-values are corrected for multiple testing (Simes). Last three columns show means ± SEM

| Acoustic parameters	Genotype	Line	Succession of encounter	First encounter	Second encounter	Third encounter
Duration (milliseconds)	0.327	0.238	<0.001	53.5 ± 1.8	65.5 ± 1.8	78.8 ± 1.8
Amplitude gap (milliseconds)	0.112	0.966	<0.001	9.6 ± 0.5	12.9 ± 0.5	16.2 ± 0.5
PF start (kHz)	0.78	0.238	<0.001	73.2 ± 0.4	71.7 ± 0.4	70.9 ± 0.4
PF max (kHz)	0.78	0.238	<0.001	88.5 ± 0.4	86.1 ± 0.4	85.1 ± 0.4
PF jump (kHz)	0.112	0.966	<0.001	19.5 ± 0.5	19.4 ± 0.5	20.8 ± 0.5
PF max loc	0.78	0.238	<0.001	0.45 ± 0.01	0.44 ± 0.01	0.42 ± 0.01
Slope of trend	0.78	0.623	<0.001	0.019 ± 0.008	0.002 ± 0.008	−0.1 ± 0.008
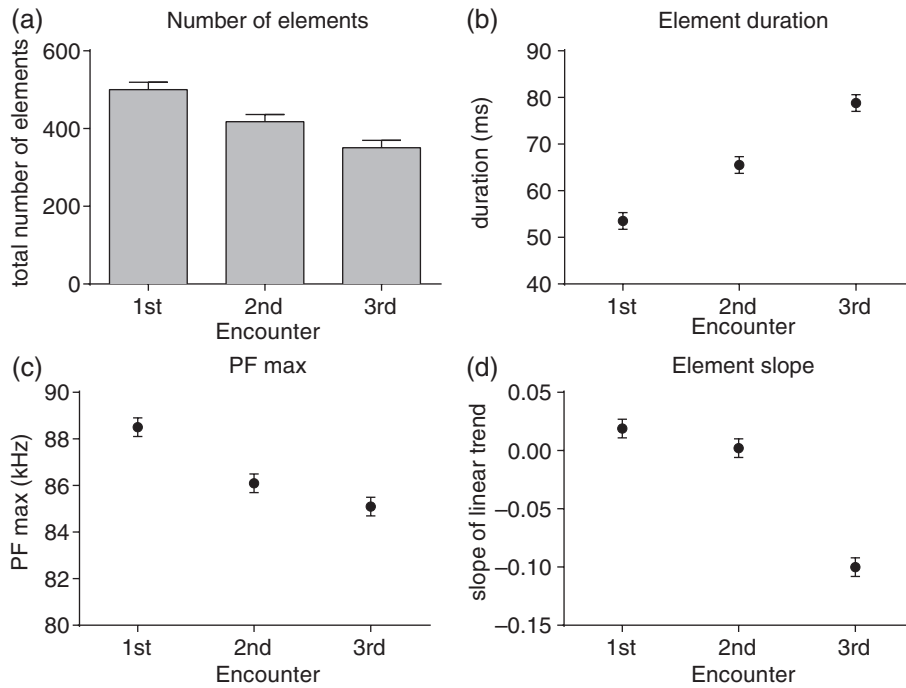
Figure 3: Influence of encounter repetitions on call structures. Number of given call elements, their duration and two different frequency parameters maximum of peak frequency and the linear slope are shown (mean ± SD).

(<50 milliseconds) only. The analysis did not show any significant differences in the structure of the calls between wild-type males (*wt*/*wt*) and males carrying the humanized version of the *Foxp2* gene (*hum*/*hum*, Table 2) either.

Difference in element structure

The analysis of the unpartitioned vocal repertoire showed no significant differences in relation to genotype, line or genotype/line interaction (Fig. 2, Table 3). We only found significant differences in relation to repetitive encounters. Dependent on whether the males had the first, second or third encounter with females, we found an increase in element duration, and a decrease in peak frequency (Fig. 3, Table 3).

The two-step cluster analysis identified a three-cluster solution as the best solution. The first cluster comprised long elements with major frequency jumps. The second cluster consisted of elements with medium duration and a general descending frequency course. The third element cluster contained short elements without frequency jumps. The analysis showed no significant differences in relation to genotypes or line in any of the different call types (Table 4).

In addition, we tested a similarly good cluster solution with six clusters to compare the obtained results with a more differentiated data set. We found significant differences between genotypes only in one cluster (CT5), in two of the acoustic parameters: PF jump [*wt*/*wt* = 30.1 ± 0.79 kHz (mean ± SEM); *hum*/*wt* = 32.9 ± 0.67 kHz; *hum*/*hum* = 33.9 ± 0.68 kHz], PF max loc (*wt*/*wt* = 0.55 ± 0.018; *hum*/*wt* = 0.51 ± 0.14; *hum*/*hum* = 0.48 ± 0.014). This suggests that males homozygous for humanized *Foxp2* produced call elements with slightly more pronounced frequency jumps and a slightly earlier frequency maximum (Table S1).

Differences in element composition

We tested 65 males, for which we had a sufficient number of elements to conduct a meaningful sequence analysis. The

Table 4: Significant differences in acoustic variables split by cluster (three-cluster solution). Percentage and number of element per cluster are given, *P*-values are corrected for multiple testing (Simes)

Acoustic parameters	Cluster 1 (21.4%, $N = 5280$) long elements, high jumps		Cluster 2 (44%, $N = 10\,863$) medium elements, descending		Cluster 3 (35.5%, $N = 8518$) short elements	
	Genotype	Line	Genotype	Line	Genotype	Line
Duration (milliseconds)	0.335	0.67	0.704	0.278	0.871	0.343
Amplitude gap (milliseconds)	0.335	0.634	0.256	0.479	0.862	0.841
PF start (Hz)	0.335	0.144	0.887	0.215	0.862	0.841
PF max (Hz)	0.335	0.144	0.704	0.07	0.474	0.841
PF jump (Hz)	0.846	0.187	0.063	0.278	0.308	0.343
PF max loc	0.335	0.634	0.704	0.278	0.084	0.841
Slope of trend	0.819	0.585	0.887	0.189	0.308	0.841

Table 5: Differences in sequence composition (chain length and number of transitions). *P*-values are corrected for multiple testing (Simes)

Acoustic parameters	Genotype	Line
Number of transition	0.658	0.537
Number of main transition	0.659	0.255
Mean chain length	0.479	0.423
Maximum chain length	0.479	0.423
Mean main chain length	0.479	0.389
Maximum main chain length	0.657	0.585

mean number of elements per sequence was 376.4 ± 13.7 (mean $\pm$ SEM), minimum $= 218$, maximum $= 777$ elements. We tested the chain length, how many times successive elements fell into the same element type and the transition rate, i.e. how often a subject changed from one element type to another. We did these comparisons at two levels: at the level of the 19 single element types and with the three general element categories to compare changes and transition of short, medium and long elements. We found no significant differences in any of these measurements between mice with the humanized *Foxp2* gene, heterozygous or wild types (Table 5). We also found no significant differences in element type usage (Table S2).

Discussion

We did not find a significant influence of the humanized *Foxp2* allele on USVs of adult mice. This result was independent from the way we classified the elements, i.e. whether we tested the unpartitioned data set containing all elements or focused on single element types. We also found no differences in our sequential analysis. Wild-type mice and mice carrying the humanized *Foxp2* gene did not differ in the percentage of element types or in their sequential order. In all of these comparisons, we found only one significant difference in one element type (ET5) that was identified via a cluster solution with a total of six clusters. In this cluster, we found a minor difference in the amount of frequency jump and the

location of the maximum peak. Given the number of statistical tests carried out in this study, it appears possible that this difference constitutes a false positive. Whether or not this change is of biological significance remains unclear in any case.

The obvious problem with interpreting negative results is that we cannot exclude whether an alternative method would have picked up some differences that our method may have missed. In a pilot study, we had determined many more acoustic variables, but these were highly correlating with the variables in the present analysis. We therefore decided to use the present reduced set of acoustic variables, to reduce the likelihood of type 2 error in the subsequent statistical tests. This set of parameters has proved to be sensitive in other studies (Enard *et al.* 2009; Fischer & Hammerschmidt 2011; Hammerschmidt *et al.* 2012a), and they picked up differences in relation to the number of encounters. We are therefore relatively confident that the lack of a significant difference between genotypes is not simply due to an inefficient method. In addition, mice have a very limited frequency resolution in the frequency range of their courtship vocalizations (Portfors *et al.* 2011); therefore, it is unlikely that further acoustic analysis or playback experiments could show a positive result. Ultimately, this would be an empirical question.

For now, our present results strongly suggest that the two amino acid substitutions in *Foxp2* do not have a substantial effect on adult mouse vocalization. Yet, we found that repeated encounters had a significant influence on song structure. This variation in vocal output, specifically the decreased number of call elements, could reflect motivational changes. Behavioral observations would be needed to clarify whether the mice explored less, for instance, and therefore produced not only fewer, but also longer vocalizations.

Apparently, the two amino acid substitutions are not sufficient to affect adult mouse vocalization patterns. This may be due to the fact that mice are not obligate vocal learners. The finding that male courtship songs of mice have a high structural complexity (Holy & Guo 2005) and that the structure of calls in pups change with age and social context (Grimsley *et al.* 2011) suggested that mice have a considerable amount of vocal flexibility. However, to what extent these variations in vocal repertoires are influenced by cortical control, motivational status and/or maturation processes is not yet clear.

Interestingly, Arriaga *et al.* (2012) identified a weak direct cortical projection to brainstem vocal motor neurons suggesting that mice should be able to have a limited degree of volitional control. However, several other studies show that deaf mice or cross-fostered mice still develop their strain-specific vocalizations (Hammerschmidt *et al.* 2012b; Kikusui *et al.* 2011; Mahrt *et al.* 2013). Hence, auditory feedback – a key feature for vocal learning – does not appear to be required for mouse vocalizations. In a recent study, *Emx1-CRE;Esco2*^{fl/fl} male mice lacking the hippocampus and large parts of the cortex show no differences in structure and complexity of courtship vocalization in comparison to control males (Hammerschmidt *et al.* 2015). This shows that cortical areas are not necessary for the production of USVs in mice irrespective of the putative direct cortical projection to brainstem vocal motor neurons (Arriaga *et al.* 2012). Hence, mouse vocalizations appear to depend much less on cortical influence and structures than human speech. Thus, if the humanized *Foxp2* affects cortico-basal ganglia circuits as suggested (Enard *et al.* 2009), they may indeed have no effects on mouse adult vocalizations.

As the adult vocalizations appear to be unaffected by the human-specific amino acid substitutions, their previously described effects on element pitch and peak frequency modulation of pup isolation calls (Enard *et al.* 2009) are presumably transient. Both studies have the same methodological background and the same recording time. In both cases, we controlled for multiple testing. An interesting question is how such small differences (e.g. mean peak frequency of 77 ± 0.7 and $73.7 \pm 0.7 \, kHz$ for wt/wt and hum/hum littermates, respectively) may be caused. In principle, a large range of even small developmental effects of humanized Foxp2 on neural and non-neural structures affecting pup vocalizations could be responsible. In our view, such effects would be very difficult to discover and study. Thus, we do not currently view the pup vocalization phenotype as a promising route to further study the effects of humanized Foxp2.

A more promising approach may be to study the genetic and neural basis of human speech in other model organisms that do show vocal learning such as song birds (Wohlgemuth *et al.* 2014) or develop new model species such as bats (Morell 2014). Nevertheless, the practical and scientific advantages of studying mice are huge and there is accumulating evidence that they have validity for studying at least some aspects of the genetic basis of human evolution (Enard 2014). If the cortical and subcortical effects on vocalizations are difficult or even impossible to study in mice, one may be able to study them indirectly by measuring other correlated phenotypes or endophenotypes, i.e. phenotypes which are more directly related to the genotype than vocalization. More work needs to be carried out to identify such phenotypes and clarify if they have validity for human speech. Motor learning paradigms have been suggested to be valid to decipher the effects of heterozygous loss-of-function mutations in Foxp2 (French & Fisher 2014; French *et al.* 2012; Groszer *et al.* 2008). Some aspects of motor learning are indeed affected in mice carrying humanized *Foxp2* (Schreiweis *et al.* 2014) as these mice learn stimulus–response associations faster than wild-type littermates potentially because the balance of activity between striatal regions is altered as indicated by measuring dopamine levels, gene expression patterns and an NMDA (N-methyl-D-aspartate) receptor-dependent form of long-term depression (Schreiweis *et al.* 2014). The study of such phenotypes may hopefully be used to better understand the human-specific function of FOXP2.

References

Ackermann, H., Hage, S. & Ziegler, W. (2014) Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav Brain Sci* **37**, 529–546.

Arriaga, G., Zhou, E. & Jarvis, E.D. (2012) The mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS One* **7**, e46610.

Egnor, S.E.R. & Hauser, M.D. (2004) A paradox in the evolution of primate vocal learning. *Trends Neurosci* **27**, 649–654.

Ehret, G. (2005) Infant rodent ultrasounds - a gate to the understanding of sound communication. *Behav Genet* **35**, 19–29.

Enard, W. (2011) FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr Opin Neurobiol* **21**, 415–424.

Enard, W. (2014) Mouse models of human evolution. *Curr Opin Genet Dev* **29**, 75–80.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S.L., Wiebe, V., Kitano, T., Monaco, A.P. & Paabo, S. (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872.

Enard, W., Gehre, S., Hammerschmidt, K. *et al.* (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* **137**, 961–971.

Ey, E., Yang, M., Katz, A.M., Woldeyohannes, L., Silverman, J.L., Leblond, C.S., Faure, P, Torquet, N., Le Sourd, A.M., Bourgeron, T. & Crawley, J.N. (2012) Absence of deficits in social behaviors and ultrasonic vocalizations in later generations of mice lacking neuroligin4. *Genes Brain Behav* **11**, 928–941.

Fischer, J. & Hammerschmidt, K. (2011) Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav* **10**, 17–27.

Fischer, J., Noser, R. & Hammerschmidt, K. (2013) Bioacoustic field research: a primer to acoustic analyses and playback experiments with primates. *Am J Primatol* **75**, 643–663.

Fitch, W.T. (2010) *The Evolution of Language*. Cambridge University Press, Cambridge.

French, C.A. & Fisher, S.E. (2014) What can mice tell us about Foxp2 function? *Curr Opin Neurobiol* **28**, 72–79.

French, C.A., Jin, X., Campbell, T.G., Gerfen, E., Groszer, M., Fisher, S.E. & Costa, R.M. (2012) An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. *Mol Psychiatry* **17**, 1077–1085.

Grimsley, J.M., Monaghan, J.J. & Wenstrup, J.J. (2011) Development of social vocalizations in mice. *PLoS One* **6**, e17460.

Groszer, M., Keays, D.A., Deacon, R.M. *et al.* (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol* **18**, 354–362.

Hage, S.R., Gavrilov, N. & Nieder, A. (2013) Cognitive control of distinct vocalizations in rhesus monkeys. *J Cogn Neurosci* **25**, 1692–1701.

Hammerschmidt, K. & Fischer, J. (2008) Constraints in primate vocal production. In Oller, K. & Griebel, U. (eds), *The Evolution of Communicative Creativity: From Fixed Signals to Contextual Flexibility*. The MIT Press, Cambridge, MA, pp. 93–119.

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H. & Fischer, J. (2012a) The structure and usage of female and male mouse ultrasonic vocalizations reveal only minor differences. *PLoS One* **7**, e41133.

Hammerschmidt, K., Reisinger, E., Westekämper, K., Ehrenreich, H., Strenzke, N. & Fischer, J. (2012b) Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci* **13**, 40.

Hammerschmidt, K., Whelan, G., Eichele, G. & Fischer, J. (2015) Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning. *Sci Rep* **5**, 8808.

Holy, T.E. & Guo, Z. (2005) Ultrasonic songs of male mice. *PLoS Biol* **3**, 2177–2186.

Jamain, S., Radyushkin, K., Hammerschmidt, K., Granon, S., Boretius, S., Varoqueaux, F., Ramanantsoa, N., Gallego, J., Ronnenberg, A., Winter, D., Frahm, J., Fischer, J., Bourgeron, T., Ehrenreich, H. & Brose, N. (2008) Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc Natl Acad Sci* **105**, 1710–1715.

Jürgens, U. (2002) Neural pathways underlying vocal control. *Neurosci Biobehav Rev* **26**, 235–258.

Jürgens, U. (2009) The neural control of vocalization in mammals: a review. *J Voice* **23**, 1–10.

Kikusui, T., Nakanishi, K., Nakagawa, R., Nagasawa, M., Mogi, K. & Okanoya, K. (2011) Cross fostering experiments suggest that mice songs are innate. *PLoS One* **6**, e17721.

Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. & Monaco, A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523.

MacDermot, K.D., Bonora, E., Sykes, N., Coupe, A.M., Lai, C.S., Vernes, S.C., Vargha-Khadem, F., McKenzie, F., Smith, R.L., Monaco, A.P. & Fisher, S.E. (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet* **76**, 1074–1080.

Mahrt, E.J., Perkel, D.J., Tong, L., Rubel, E.W. & Portfors, C.V. (2013) Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. *J Neurosci* **33**, 5573–5583.

Morell, V. (2014) When the bat sings. *Science (New York, NY)* **344**, 1334–1337.

Murugan, M., Harward, S., Scharff, C. & Mooney, R. (2013) Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron* **80**, 1464–1476.

Pfenning, A.R., Hara, E., Whitney, O. *et al.* (2014) Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science (New York, NY)* **346**, 1256846.

Portfors, C.V., Mayko, Z.M., Jonson, K., Cha, G.F. & Roberts, P.D. (2011) Spatial organization of receptive fields in the auditory midbrain of awake mouse. *Neuroscience* **193**, 429–439.

Reimers-Kipping, S., Hevers, W., Paabo, S. & Enard, W. (2011) Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience* **175**, 75–84.

Rousseeuw, P. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* **20**, 12.

Scattoni, M.L., Gandhy, S.U., Ricceri, L. & Crawley, J.N. (2008) Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS One* **3**, e3067.

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., Goyal, S., Rea, E., French, C.A., Puliyadi, R., Groszer, M., Fisher, S.E., Mundry, R., Winter, C., Hevers, W., Pääbo, S., Enard, W. & Graybiel, A.M. (2014) Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc Natl Acad Sci* **111**, 14253–14258.

Simonyan, K. (2014) The laryngeal motor cortex: its organization and connectivity. *Curr Opin Neurobiol* **28**, 15–21.

Simonyan, K. & Jürgens, U. (2003) Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Res* **974**, 43–59.

Simonyan, K., Horwitz, B. & Jarvis, E.D. (2012) Dopamine regulation of human speech and bird song: a critical review. *Brain Lang* **122**, 142–150.

Vargha-Khadem, F., Gadian, D.G., Copp, A. & Mishkin, M. (2005) FOXP2 and the neuroanatomy of speech and language. *Nat Rev Neurosci* **6**, 131–138.

Watkins, K. (2011) Developmental disorders of speech and language: from genes to brain structure and function. *Prog Brain Res* **189**, 225–238.

Wich, S.A., Krützen, M., Lameira, A.R., Nater, A., Arora, N., Bastian, M.L., Morrogh-Bernard, H.C., Utami Atmoko, S.U., Pamugkas, J., Perwitasari-Farajallah, D., Hardus, M.E., Noordwijk, M. & van Schaik, C.P. (2012) Call cultures in Orang-Utans? *PLoS One* **7**, e36180.

Wohlgemuth, S., Adam, I. & Scharff, C. (2014) FoxP2 in songbirds. *Curr Opin Neurobiol* **28**, 86–93.

Zhang, J.Z., Webb, D.M. & Podlaha, O. (2002) Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics* **162**, 1825–1835.

Acknowledgments

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Figure S1: Examples of calls describing the categories used for visual classification.

Table S1: Significant differences between genotypes' six-cluster solution. Results are given as *P*-values and are corrected for multiple testing (Simes). Significant differences are marked in bold.

Table S2: Differences in element type usage. ET, element type; GEC, general element category; *P*-values show differences in relation to genotype and line and are corrected for multiple testing (Simes). Significant differences are marked in bold.