Received: 26 January 2021 | Revised: 2 July 2021 | Accepted: 31 July 2021

DOI: 10.1111/gbb.12764

ORIGINAL ARTICLE

Genes, Brain and Behavior

A humanized version of *Foxp2* affects ultrasonic vocalization in adult female and male mice

Sophie von Merten[1,2] | Christine Pfeifle[2] | Sven Künzel[2] | Svenja Hoier[2] | Diethard Tautz[2]

[1]CESAM - Centro de Estudos do Ambiente e do Mar, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[2]Department for Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany

Correspondence
Sophie von Merten, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.
Email: sophievonmerten@gmail.com

Funding information
Fundação para a Ciência e a Tecnologia, Grant/Award Numbers: SFRH/BPD/118053/2016, UIDP/50017/2020+UIDB/50017/2020; Max Planck Society

Abstract

The transcription factor *FoxP2* is involved in setting up the neuronal circuitry for vocal learning in mammals and birds and is thought to have played a special role in the evolution of human speech and language. It has been shown that an allele with a humanized version of the murine *Foxp2* gene changes the ultrasonic vocalization of mouse pups compared to pups of the wild-type inbred strain. Here we tested if this humanized allele would also affect the ultrasonic vocalization of adult female and male mice. In a previous study, in which only male vocalization was considered and the mice were recorded under a restricted spatial and temporal regime, no difference in adult vocalization between genotypes was found. Here, we use a different test paradigm in which both female and male vocalizations are recorded in extended social contact. We found differences in temporal, spectral and syntactical parameters between the genotypes in both sexes, and between sexes. Mice carrying the humanized *Foxp2* allele were using higher frequencies and more complex syllable types than mice of the corresponding wildtype inbred strain. Our results support the notion that the humanized *Foxp2* allele has a differential effect on mouse ultrasonic vocalization. As mice carrying the humanized version of the *Foxp2* gene show effects opposite to those of mice carrying disrupted or mutated alleles of this gene, we conclude that this mouse line represents an important model for the study of human speech and language evolution.

KEYWORDS
communication, evolution, *FoxP2*, language, mice, song, speech, ultrasonic, USV, vocalization

1 | INTRODUCTION

A notable feature of humans is the capability for vocal learning, that is, to develop speech and language. Vocal learning is also well-known for many bird and some mammal species, including cetaceans, pinnipeds, and bats.[1–5] However, in none of these species a complex language is formed.

One gene that is considered to have played an important role in human speech and language acquisition is the transcription factor Forkhead BoxP2, *FoxP2*.[6] During the development of the mammalian brain, *FoxP2* is expressed in the deep layers of the cortex, medium spiny neurons of the basal ganglia and parts of the thalamus.[7,8] These neurons belong to brain circuits involved in the acquisition of motor skills and coordination, in procedural learning and sensory-motor integration.

Additionally, these brain circuits are crucial for the complex orofacial and laryngeal movements required during sound production.[9,10] Mutations or lowered expression levels of *FoxP2* are connected to language and speech disorders in humans[11] and hinder specification of learned songs in birds.[6,12] In bats, *FoxP2* also seems to play a role in vocal production learning[4] and echolocation,[13,14] although no functional studies have been conducted yet to confirm found correlations.

FoxP2 belongs to the most conserved genes in vertebrates. Between the human and the mouse gene, there are only three amino acid substitutions.[15] Two of these substitutions appeared after the split of humans and chimpanzees and it has been suggested that they underwent positive selection due to their effect on human speech and language.[16–18] However, this interpretation has been questioned based on the analysis of larger population samples.[19] Still, given the otherwise high degree of conservation of the respective substitutions, it is valid to analyze their possible functional role in a mouse model.

Female and male mice use complex ultrasonic vocalization (USV) in different contexts.[20–24] Mouse USV is composed of single vocal elements (syllables) arranged in non-random sequences, so-called songs.[25] Structural and temporal characteristics differ between individuals, populations, and laboratory strains, and according to an animal's age, sex, and social experience.[26–30] Deafening and cross-fostering studies suggest that most of the spectral features of USV are rather genetically inherited than learned.[31,32] Other studies found however that auditory feedback might be needed to maintain certain features of the mouse song and that mice have neuroanatomical features thought to be unique to humans and song-learning birds.[27,33]

Foxp2 plays an important role in vocalization and motor abilities in mice. Mouse pups with a disruption or mutation of the *Foxp2* gene emit USV calls at lower rates and, in heterozygous individuals of some mutant strains, at lower amplitudes and/or with changed call durations.[34] Homozygous individuals additionally suffer from severe motor impairment and die prematurely.[35,36] Consequently, the majority of studies has focused on pup USV.[34] Heterozygous mice of these mutant strains, however, can survive to adulthood. Adult male mice heterozygous for a *Foxp2*-knockout emit less USV calls, have a different syllable repertoire, and use an irregular rhythmic structure compared to their wildtype littermates.[37] Male mice heterozygous for different *Foxp2*-mutations have a syllable repertoire similar to wildtype strains.[38,39] However, depending on the type of mutation and the social context, they change the emission of complex syllable types[38,39] and show a reduced syntax complexity.[38]

Another interesting mouse model to study the influence of *Foxp2* on sound production and vocal learning is carrying the humanized allele of the murine *Foxp2* gene, $Foxp2^{hum/hum}$.[40] Mice carrying this gene are generally healthy and live to adulthood, but show some neuromorphological, neurochemical and neurophysiological abnormalities as well as behavioral differences, including accelerated learning. $Foxp2^{hum/hum}$ pups emit USV calls in a similar number and duration, however with lower mean peak frequencies as compared to their wildtype littermates, $Foxp2^{+/+}$.[40–43]

Despite all these effects of the $Foxp2^{hum/hum}$ on pup vocalization and behavior, as well as on adult learning and physiology, a previous study on adult vocalization found almost no difference between $Foxp2^{hum/hum}$ and $Foxp2^{+/+}$ mice.[44] It appears, however, that the approach used to obtain the USV recordings may have concealed actual differences. Hammerschmidt and colleagues conducted short recordings of pairs of mice in direct physical encounters, where pairs consisted of females and males of the same and of different genotypes (see Figure 1A in[44]). However, recorded USVs were not differentiated by sex and possible female vocalizations were not considered. It has been shown for many different strains and populations of mice that not only male but also female mice emit USV, in both same-sex and different-sex interactions and, depending on the context, in higher or lower numbers than males.[20–22,30,45–50] If, as can be expected and has been shown in the present study, female $Foxp2^{hum/hum}$ mice emit USV, the true differences between genotypes might have been masked in this way in the study by Hammerschmidt et al.[44] Another potential methodological constraint concerns the length of the recordings. In our previous studies of wildtype mice, we developed a scheme for extended recordings (12 hours; see also[30]), as this seems more natural than the standard stimulus-induced short recording times (2–5 min).

Here, we recorded and compared the USV of adult female and male $Foxp2^{hum/hum}$ and $Foxp2^{+/+}$ mice for a full night, taking care to record the **vocalization**s of female and male mice separately while they were allowed to have olfactory and auditory contact. We present and discuss the results of our analyses of temporal, spectral and syntactical characteristics of recorded USV.

2 | MATERIAL AND METHODS

2.1 | Animals and housing

The animals used in this study were obtained from the Max Planck Institute for Evolutionary Anthropology (Leipzig, Germany). We used animals homozygous for the humanized *Foxp2* gene ($Foxp2^{hum/hum}$, derived from the Bl6 ES cell 5H10, as described in[40]) and the corresponding C57BL/6 mice homozygous for the wildtype murine *Foxp2* gene ($Foxp2^{+/+}$). These lines were originally generated by Ozgene (Bentley, Australia) from two C57BL/6 embryonic stem-cell clones. The $Foxp2^{hum}$ allele is characterized by two particular nucleotide-substitutions in exon seven of the *Foxp2* gene.[40]

We recorded initially from a group of 32 mature virgin mice of age 10–16 weeks, eight female and eight male $Foxp2^{hum/hum}$ mice, and eight female and eight male $Foxp2^{+/+}$ mice (Table 1). However, after the completion of this dataset, we were notified that due to a backcross during the generation of the $Foxp2^{hum/hum}$ line (to a wildtype C57BL/6, Harlan Laboratories), a deletion of the gene *SNCA* (alpha-synuclein) has inadvertently segregated into this line. As we had not genotyped the initial 32 mice used in the comparison of $Foxp2^{hum/hum}$ and $Foxp2^{+/+}$ mice for the *SNCA* locus, we could not exclude a potential influence of the *SNCA* deletion on USV production. We thus genotyped additional $Foxp2^{hum/hum}$ mice for the *SNCA* locus and used them for a second group of recordings. In this group we recorded 32 mature virgin mice of age 10–16 weeks, eight female and eight male $Foxp2^{hum/hum\text{-}SNCADel}$ mice (carrying the deletion of the *SNCA*

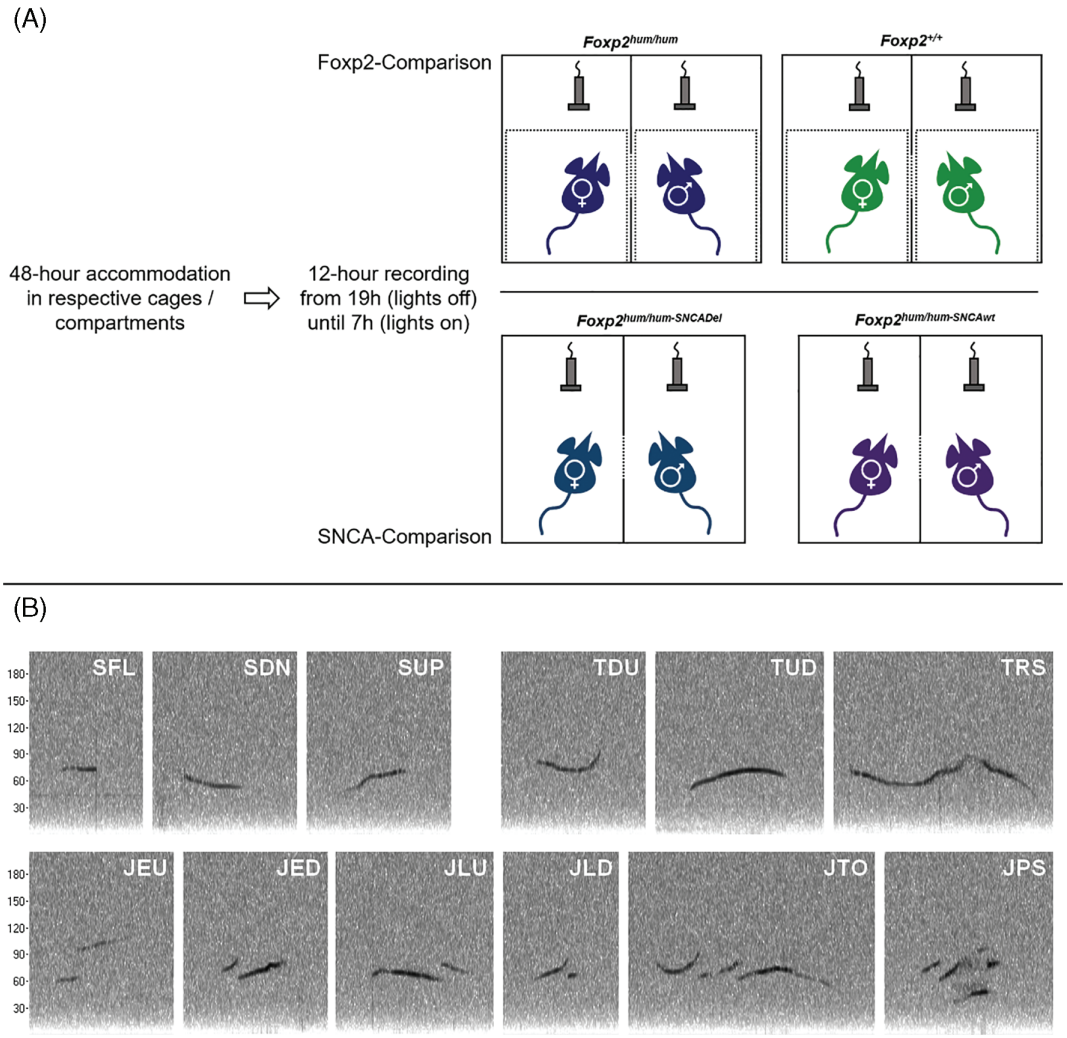

FIGURE 1 (A) Scheme of recordings of the Foxp2- and the SNCA-Comparison. Each one female and one male mouse of the same genotype were recorded in a setup that allowed sensory interaction between them through a grid, but a separate recording of their USV by having one microphone per side of the setup. For details of the setup of the Foxp2-Comparison see Figure S1; for details of the setup of the SNCAComparison see Figure 1 in.[30] (B) Spectrograms of the 12 syllable types use for statistical analysis. For an explanation of abbreviations see Table 3 and Method section Song and syllable parameters

locus), and eight female and eight male *Foxp2*$^{hum/hum\text{-}SNCAwt}$ mice (not carrying the deletion of the *SNCA* locus; Table 2) to assess a possible influence of the *SNCA* deletion on the USV pattern. In the following, we will use the terms Foxp2-Comparison and SNCA-Comparison to distinguish between the two datasets. Whenever not specified, methods were the same for both comparisons.

All mice were housed in family groups with one to four of their littermates (separated by sex) in one room under controlled standard conditions with a 12 h/12 h dark/light cycle (lights on at 0700 h). Mice were kept in standard laboratory cages (Type IV Bioscape, Germany), equipped with bedding material and environmental enrichment (cardboard box, paper stripes, wood wool, running plate). Standard diet (Altromin, Germany) and water was presented *at libitum*.

Different genotypes were kept in the same room, but, as ultrasonic sounds attenuate fast and thus do not travel far, events in which mice from one genotype could overhear USV of mice of another genotype would be extremely rare and if so, only faint. Further, it has been shown that mouse song is mostly innate and seems to be little influenced by learning.[32,51] Considering the unlikely possibility that the genotypes in our current study would influence the others' USV, most likely the USV of the two genotypes would be more similar and any differences we find would thus rather be an underestimation of the real differences found under strict auditory separation.

2.2 | SNCA genotyping

To identify the genotype *Foxp2*$^{hum/hum\text{-}SNCAwt}$ or *Foxp2*$^{hum/hum\text{-}SNCADel}$, we used specific primers to amplify the mouse *SNCA* exon 6: forward 5′-AAGACTATGAGCCTGAAGCCTAAG-3′, reverse: 5′-AGTGTGAAG CCACAACAATATCC-3′; 266-bp fragment) and to detect the *SNCA* deletion, we used the following primers: forward 5′-AGTTCCA

TABLE 1 Overview of mice recorded to test the influence of the humanized *Foxp2* gene. For each individual the sex and number of songs and syllables emitted during the full period of recording are given

Foxp2$^{hum/hum}$				*Foxp2*$^{+/+}$			
Mouse	**Sex**	**Songs**	**Syllables**	**Mouse**	**Sex**	**Songs**	**Syllables**
hum-F-1	f	57	528	mus-F-1	f	35	167
hum-F-2	f	35	122	mus-F-2	f	37	208
hum-F-3	f	108	641	mus-F-3	f	51	207
hum-F-4	f	34	243	mus-F-4	f	58	421
hum-F-5	f	6	71	mus-F-5	f	37	154
hum-F-6	f	8	11	mus-F-6	f	57	392
hum-F-7	f	0	0	mus-F-7	f	41	343
hum-F-8	f	40	148	mus-F-8	f	34	288
hum-M-1	m	56	328	mus-M-1	m	38	260
hum-M-2	m	2	12	mus-M-2	m	38	285
hum-M-3	m	44	196	mus-M-3	m	41	365
hum-M-4	m	37	288	mus-M-4	m	34	209
hum-M-5	m	32	109	mus-M-5	m	76	673
hum-M-6	m	36	288	mus-M-6	m	51	240
hum-M-7	m	3	3	mus-M-7	m	46	367
hum-M-8	m	0	0	mus-M-8	m	37	404
Total	16	498	2988	total	16	711	4983

TABLE 2 Overview of mice recorded to test the influence of the SNCA-deletion. For each individual the sex and number of songs and syllables emitted during the full period of recording are given

Foxp2$^{hum/hum\text{-}SNCADel}$				*Foxp2*$^{hum/hum\text{-}SNCAwt}$			
mouse	**Sex**	**Songs**	**Syllables**	**Mouse**	**Sex**	**Songs**	**Syllables**
SNCA_D1F1	f	70	770	SNCA_N1F1	f	51	97
SNCA_D2F1	f	76	312	SNCA_N2F1	f	43	79
SNCA_D3F1	f	87	765	SNCA_N3F1	f	69	574
SNCA_D4F1	f	73	393	SNCA_N4F1	f	71	482
SNCA_D5F1	f	101	1183	SNCA_N5F1	f	77	528
SNCA_D6F1	f	72	599	SNCA_N6F1	f	68	453
SNCA_D7F1	f	36	104	SNCA_N7F1	f	74	616
SNCA_D8F1	f	71	291	SNCA_N8F1	f	76	465
SNCA_D1M1	m	64	722	SNCA_N1M1	m	11	18
SNCA_D2M1	m	83	610	SNCA_N2M1	m	12	33
SNCA_D3M1	m	72	328	SNCA_N3M1	m	41	311
SNCA_D4M1	m	57	160	SNCA_N4M1	m	70	480
SNCA_D5M1	m	69	559	SNCA_N5M1	m	58	253
SNCA_D6M1	m	68	629	SNCA_N6M1	m	61	420
SNCA_D7M1	m	60	139	SNCA_N7M1	m	79	381
SNCA_D8M1	m	61	242	SNCA_N8M1	m	68	428
Total	16	1120	7806		16	929	5618

CTGTTCTGGCCAT-3′ and reverse 5′-GTAACAATACAGCAAGAGAT AC-3′ (174-bp fragment). DNA were extracted from ear clips. The primers (10 μM) were used in a standard PCR (95°C–30 s, 58°C–30 s, 72°C–30 s cycle 35) and the resulting fragments were resolved by 1,5% agarose gel-electrophoresis. The expected fragment sizes for WT were 266 bp, for the deletion 176 bp. Animals homozygous for the WT allele or the deletion allele were used for the SNCA-Comparison recordings.

2.3 | USV recording

Recordings for the Foxp2-Comparison were conducted in two custom-made wooden boxes (each measuring 110 × 50 × 50 cm; Figure S1) placed inside the experimental room. To reduce reverberations and disturbances from outside, the inner walls of the boxes were faced with black acoustic foam. Each box was separated into two recording compartments by a wall of gray PVC. Two openings inside

this wall (contact windows, each measuring 20 cm in length and 2 cm in height and 2 cm above the ground of the box) enabled the mice in the two compartments of each box to have visual, olfactory, and acoustic contact, similar as described in von Merten et al.[30] These contact windows could be closed, e.g., during the habituation time of mice. In each compartment a commercial pet cage for mice (38 × 23 × 20 cm) was placed directly next to the separating wall. Mice housed in the room were acoustically separated from the experimental mice by the recording boxes.

For logistic reasons, recordings for the SNCA-Comparison were conducted in another type of setup, which we had used before in other recordings (see Figure 1 in[30]). This setup was only slightly larger (each mouse had 60 × 25 cm of floor space) and, similarly to the other setup, two pairs of mice were recorded at the same time. However, the wall of this setup was made from PVC walls that were not lined with acoustic foam. The possibly resulting echoes can overlap with the actual call, with a small temporal delay, which can lead to the calls appearing slightly longer and with an elevated amplitude in the recordings. These differences in recording setups do not allow a direct comparison of the obtained data, which were consequently analyzed separately.

In all recordings, one ultrasonic microphone (condenser ultrasound microphone CM16/CMPA, Avisoft Bioacoustics, Germany) was attached inside each recording compartment, in a central position 25 cm above the bottom of the respective cage/recording chamber. Microphones were connected to a multi-channel recording device (Avisoft UltraSoundGate 416Hm, 4-channel, operational up to 4 × 500 kHz). Recordings were made with a sampling rate of 500 kHz and a depth of 16bit (software: Avisoft USGH recorder) using the "whistle tracking" option to automatically detect and record mouse USV. Recordings were triggered by whistle-like calls ranging between 20 and 250 kHz and lasting at least 8 ms. To make sure the whole USV element was recorded, a pre-trigger of 1 ms was activated and the recording event lasted until 1 s after the last automatically detected sound.

2.4 | Recording schedule

Two days before a new recording session started, all recording cages/compartments were prepared with fresh bedding, a cardboard box, paper strips, food, and water. The four mice to be recorded in two pairs in the respective recording session were then individually introduced into these cages/compartments for habituation (Figure 1A). During this time, the contact windows were closed not allowing contact between mice. After habituation, the contact windows between the two members of each recording pair were opened and recordings started at 19:00 h (time of lights off) and lasted until 07:00 h on the next morning (lights on). Mice were thus recorded during the night, that is, including their periods of highest activity (see Figure S2 for vocal activity during the night).

Recording partners, the two mice in neighboring compartments, were always a male and a female mouse of the same genotype, that were unfamiliar to each other. In the days before recordings, male bedding was added to female cages to induce the estrus cycle of females. Estrus state of recorded females was not checked, as it was shown that this has no influence on male USV.[52]

Visual observation of the recorded vocalizations suggested a few possible simultaneous recordings in neighboring compartments. All such potential double recordings were detected by a custom script (by Bernhard Haubold, MPI for Evolutionary Biology) and not used for analysis.

2.5 | Sound analysis

For analysis of all recordings, the frequency-time course of the USV was extracted and a range of song and syllable parameters were calculated. We define a syllable as a single acoustic unit separated from other such units by at least 15 ms of silence, and song as a sequence of syllables, separated from other such syllable sequences by at least 500 ms of silence. We chose the respective time intervals after scrutinized visual and acoustic (slowed down 3×) inspection of many of the current recordings by two researchers experienced with mouse USV (SVM, SH).

We extracted the time-frequency course of each syllable semi-automatically using *Selena* (Department of Animal Physiology, University of Tübingen, Germany). For this purpose, recordings were displayed as color spectrograms using 1024 kHz Fast Fourier Transformation (FFT). These spectrograms were then enlarged to 4096 points, improving the spectral reading accuracy to 60 Hz. The frequency-time course of each single syllable was tracked by the software by marking the screen pixel with highest amplitude for each instantaneous FFT. Each of these markings was then checked visually and corrected manually if necessary. Afterwards, time, frequency and amplitude of each marked pixel were exported to a csv-file.

These csv-files were used to calculate several song and syllable parameters, using a custom-written MATLAB script (SVM, SH). For each individual we calculated the number of songs; for each song the number of syllables, its duration, and syllables per second (calculated per song, thus not considering the silence between songs); for each syllable the average, start, and minimum frequency, its frequency bandwidth, duration, and frequency slope. The average frequency of syllables was calculated as the centre of gravity (COG), a weighted average, where the relative amplitude of each frequency is taken into account. It is calculated as

$$\mathrm{COG} = \sum_{i=1}^{n} (Ferq * (Ampl_i + 100)) \div \sum_{i=1}^{n} (Ampl_i + 100)$$

where n is the number of frequency-amplitude pairs in each syllable, $Freq_i$ and $Ampl_i$ are the values for frequency and amplitude at i. In this model, frequency values with high relative amplitudes have a bigger influence on the average than frequency values with low relative amplitudes.

Based on these calculations, syllables were first assigned to 17 syllable types (Table S1), depending on their frequency slope and the existence of turning points and jumps. Syllable types were chosen

based on findings of Holy & Guo[25] and correspond to our previous work.[21,30] All syllable types can be grouped in three syllable groups: simple syllables (S), turn syllables (T) and jump syllables (J). Simple syllables are syllables without any frequency jumps or turning points that are either flat (SFL), upward frequency modulated (SUP) or downward frequency modulated (SDN), with modulations of at least 0.05 kHz per ms. Turn syllables are syllables with one or more turns of the direction of the frequency modulation, and are, depending on their spectrographic shape, further split into U-shaped syllables (turn-down-up TDU), inverse U-shaped syllables (turn-up-down TUD) and syllables with more than one turning point (TRS). Jump syllables contain one or more sudden frequency jumps with an at least 12 kHz change in less than four instantaneous FFT bins. Depending on the number and position of these jumps within the syllable, jump-syllables were sub-categorized. Syllables with one frequency jump were grouped into syllable types with one jump to a higher or lower frequency occurring in either the first (jump-early-up JEU, jump-early-down JED) or the second half (jump-late-up JLU, and jump-late-down JLD) of the syllable. Syllables with two frequency jumps were grouped into syllable types with two jumps occurring in either the first (jump-early-two JET) or the second half (jump-late-two JLT) of the syllable, syllable types with one jump up in each half of the syllable (jump-up-up JUU) or with one jump down in each half of the syllable (JDD jump-down-down), and syllable types with one jump up and one jump down in either the first or second half of the syllable respectively (jump-up-down JUD, jump-down-up JDU). All syllables with more than two jumps were classed into the remaining category (jumps JPS). Syllables with two frequency jumps were rarely observed (Table S1) and we thus had to group them into a higher class (two jumps JTO) for statistical analysis, resulting in 12 syllable types (Table 3, Figure 1B). To analyze the correlation between frequency bandwidth and number of jumps in syllables as well as for the visualization of the syntax analysis we used the three syllable groups (S, T, and J) instead of the 12 syllable types.

For some syllables, the assignment to one of the abovementioned syllable types was not possible, as their shape was not traceable at all. Such unstructured syllables have been described before.[53] Here, they were rare but observed in both genotypes. Due to their shape, it was not possible to extract their time-frequency course as it was done with other syllable types. Instead, start, stop, minimum and maximum frequency were marked and extracted. A possible explanation for these unstructured syllables might be found in the reported ongoing hearing loss of the C57BL/6 strain.[54,55] Deafening studies like that of Arriaga et al. found syllables with less spectral purity in the vocal repertoire of deafened mice.[33] Although it is reported that the hearing loss in the C57BL/6 strain usually sets on at the age of about 12 months,[55] it might be possible that already before this critical age (our mice were 10–16 weeks old), hearing abilities of mice were already somewhat limited. As statistical analysis did not reveal any difference between the *Foxp2*-genotypes, neither in the total amount of these syllables, nor in their structural measures, these syllables were not included into further analyses.

2.6 | Data analysis

All statistical analyses were carried out using R 3.4.3 (R[56]). Obtained *p* values were, if needed, corrected applying the Bonferroni correction

TABLE 3 Relative usage of the 12 syllable types used in the final analysis. For each group, the relative usage of syllable types in general (columns "use"), as a start ("sta") or as a stop syllable ("sto")

	female *Foxp2*$^{hum/hum}$			male *Foxp2*$^{hum/hum}$			female *Foxp2*$^{+/+}$			male *Foxp2*$^{+/+}$		
Type	use	sta	sto	use	sta	sto	use	sta	sto	use	sta	sto
SFL	0.002	0.003	0.007	0.002	0.010	0.000	0.004	0.000	0.006	0.003	0.006	0.006
SUP	0.322	0.354	0.401	0.297	0.314	0.340	0.428	0.426	0.441	0.392	0.424	0.436
SDN	0.090	0.087	0.087	0.094	0.067	0.067	0.112	0.094	0.117	0.091	0.069	0.097
TRS	0.116	0.153	0.129	0.083	0.095	0.096	0.128	0.163	0.123	0.120	0.186	0.142
TUD	0.070	0.076	0.063	0.070	0.086	0.096	0.097	0.126	0.103	0.087	0.080	0.083
TDU	0.047	0.038	0.073	0.042	0.024	0.067	0.040	0.034	0.069	0.049	0.053	0.064
JEU	0.014	0.014	0.003	0.015	0.019	0.010	0.010	0.011	0.009	0.008	0.006	0.008
JED	0.155	0.101	0.098	0.183	0.181	0.153	0.079	0.057	0.063	0.124	0.091	0.100
JLU	0.023	0.021	0.017	0.029	0.043	0.019	0.019	0.031	0.017	0.014	0.014	0.008
JLD	0.067	0.056	0.077	0.066	0.048	0.110	0.043	0.017	0.040	0.051	0.028	0.039
JTO[a]	0.066	0.059	0.031	0.074	0.057	0.038	0.028	0.029	0.006	0.045	0.036	0.011
JPS	0.028	0.038	0.014	0.045	0.057	0.005	0.012	0.011	0.006	0.016	0.008	0.006

Note: For a detailed description of the different syllable types, see Method section *Song and syllable parameters*.
Abbreviations: JEU, jump-early-up; JLU, jump-late-up; JED, jump-early-down; JLD, jump-late-down; JET, jump-early-two; JLT, jump-late-two; JUU, jump-up-up; JUD, jump-up-down; JDU, jump-down-up; JDD, jump-down-down; JTO, two jumps; JPS, more than two jumps; SFL, simple flat; SUP, simple up; SDN, simple down; TUD, inverted u-shaped turn; TDU, u-shaped turn; TRS, more than one turn.
[a]This syllable type contains all initial syllable types with two frequency jumps that had to be merged into this higher class for statistical analysis.

for multiple testing, using the formula $p' = p^*k$, where p' is the corrected p value and k is the number of tests used.

2.6.1 | Temporal and structural song and syllable parameters

Data of mice that had sung at least three songs (see Tables 1 and 2 for the number of songs of the Foxp2- and the SNCA-Comparison, respectively) was divided in two subsets per comparison, one containing three temporal song properties (song duration, syllables per song, and syllables per second), the other containing all structural syllable properties (average frequency, start frequency, minimum frequency, frequency bandwidth, duration of syllable, and frequency slope). The number of songs per individual formed the third data set, containing data of all mice, including those that had sung less than three songs. For an overview of the four temporal and six spectral parameters and their average values per group see Tables S3 and S4 (Foxp2- and SNCA-Comparison).

For all three data sets of both comparisons, we used generalized linear models (functions lmer and glmmadmb from the R packages lme4, lmerTest, and glmmADMB) with genotype, sex and their interaction as fixed factors and the above-mentioned parameters (Tables S3 and S4) as response variables. For each response variable, we chose the best fitting error distribution (confirmed visually with QQ-Plots), which were gamma for number of songs, syllables per song, and syllables per second; negative binomial for song duration and syllable duration; and gaussian for the rest. No data transformation was necessary. In datasets with more than one value per individual, we included individual as a random factor.

We found strong differences in frequency bandwidth and number of jump syllables between the genotypes of the Foxp2-Comparison. To analyze if those differences were related, we fit an additional generalized linear model for the response variable frequency bandwidth, including syllable group as a fixed factor.

An inspection of the data suggested a larger variance in the number of songs in the *Foxp2*$^{hum/hum}$ in contrast to the *Foxp2*$^{+/+}$ mice of the Foxp2-Comparison. We thus conducted a Levene's test for homogeneity of variance on the number of songs between the two genotypes.

To distinguish the influence of the different spectral parameters, including additionally the number of jumps and number of turns in emitted syllables, on separating the groups and to visualize data, we ran a linear discriminant function analysis (function lds from the R package mass) with a combination of genotype and sex as grouping variable and plotting the first two discriminant functions.

2.6.2 | Syntax analysis

To compare the usage of different syllable types between genotypes and sexes, we calculated for each group of animals from the Foxp2-Comparison four probability tables, containing (1) The general probabilities to emit any of the 12 syllable types in a song; the probabilities to (2) Start or (3) Stop a song with any of the 12 syllable types, and (4) The probabilities that any of the 12 syllable types follow any of the 12 syllable types (transition probability). The transition probability P_{ST} from a syllable type S to a syllable type T is calculated as

$$PST = \frac{occurrence\ of\ transition\ type\ S \rightarrow T}{occurrence\ of\ any\ transition\ type\ S \rightarrow X}$$

where S is any of the 12 syllable types, T is any of the 12 syllable types (including S), and X is either any syllable type (including S and T) or demarks the end of the sequence. All transition probabilities from one syllable type, including the ending of the sequence, add up to one.

We compared the four tables between genotypes and sexes using Chi-square tests. As the transition probability tables contained several values similar to zero, we reduced these tables for statistical analysis by (1) Removing the transitions from and to the syllable type involved in the least number of transitions (SFL, involved in less than 1% of all transitions; Table S2); and (2) Joining the transitions from and to the second and third rarest syllable types involved in transitions (JEU and JLU, involved in 2 or 3% of transitions respectively; Table S2), which resulted in transition tables without zeros.

In the next step, we analyzed if mice were arranging syllable types inside a song simply according to their general probability, or according to a more complex rule. We generated syllable sequences using two models, a Probability model (PM) and a Markov model (MM), and compared those to the recorded syllable sequences. While PMs use the probabilities of the occurrence of certain syllable types to generate sequences, MMs use the probabilities of transitions between certain syllable types[57]; see above formula for the calculation of transition probabilities. For both models, we generated 5000 sequences for each group of mice, applying a custom R-script to the recorded syllable sequences following von Merten et al.[30]

To generate syllable sequences after the PM, we used the general probabilities of syllable types to be used by the four groups of mice (columns "use" in Table 3). To generate syllable sequences after the MM, we used for each group a start matrix (the table containing the probabilities to start a sequence with a certain syllable type; columns "sta" in Table 3) and a transition matrix (containing the probabilities to transition from one syllable type to another; Table S2). The transition matrix we used to generate syllable sequences after the MM was the original table containing all possible transitions, not the one reduced for statistical analysis (see above). We set the maximum number of syllables per sequence for each group according to this group's upper limit of the 95% confidence interval of number of syllables per song recorded, that is, 22 for female *Foxp2*$^{hum/hum}$, 18 for male *Foxp2*$^{hum/hum}$, 21 for female *Foxp2*$^{+/+}$, 32 for male *Foxp2*$^{+/+}$. Sequences could end earlier if the algorithm stopped before (see[30]).

From the generated sequences as well as from the originally recorded sequences, we calculated the occurrence of syllable-type doublets (e.g., SUP→SUP or TRS→JPS) and syllable-type triplets (e.g., SUP→SUP→SDN or TRS→JPS→SUP). As an example, the

expected probability of occurrence of the triplet TRS→JPS→SUP would be calculated as $p_{TRS\rightarrow JPS\rightarrow SUP} = p_{TRS} * p_{JPS} * p_{SUP}$ after the PM and as $p_{TRS\rightarrow JPS\rightarrow SUP} = p_{TRS}{}^{start} * p_{TRS\rightarrow JPS} * p_{JPS\rightarrow SUP}$ after the MM.

To compare the fit of the syllable sequences generated by the two models with the recorded sequences, we calculated the absolute differences between the values of the respective parameters (i.e., occurrence of syllable-type doublets, occurrence of syllable-type triplets) from the generated syllable sequences and the recorded syllable sequences, summed over syllable types and levels of the respective parameter. E.g., the fit of the occurrence of syllable-type doublets as calculated from the PM-generated sequences is calculated as $d_{SUM\text{-}PM} = \sum\sum|PM(probability\ of\ doublet\ type) - original(probability\ of\ doublet\ type)|$; the fit of the occurrence of syllable type-triplets as calculated from the MM-generated sequences is calculated as $d_{SUM\text{-}MM} = \sum\sum|MM(probability\ of\ triplet\ type) - original(probability\ of\ triplet\ type)|$.

For a better visualization of transition probabilities, we constructed diagrams representing the Markov processes for each group. As Markov processes with many different states (syllable types, in our case) are complicated to read, we used the three syllable groups S, T, and J. The diagrams show the three syllable groups and all possible transitions between them, as well as the respective probabilities to start or to end a sequence with a syllable from the respective group. Please note that for statistical analyses, all 12 syllable types were used.

3 | RESULTS

We recorded from two groups of 32 mice each, termed the Foxp2-Comparison (Table 1) and the SNCA-Comparison (Table 2). The Foxp2-Comparison constitutes the main analysis, in which the humanized allele of *Foxp2* was compared with the murine wildtype allele of *Foxp2*. However, due to a backcross issue (see Methods), the humanized *Foxp2* allele was partially linked to an *SNCA* deletion allele. The SNCA-Comparison served therefore to assess whether this deletion allele would influence the USV call parameters.

3.1 | Temporal and spectral structure of USV of Foxp2$^{hum/hum}$ and Foxp2$^{+/+}$ mice

From the 32 mice of the Foxp2-Comparison we recorded 1209 songs, containing 7971 syllables over the 12 hours of recording. Females and males contributed almost equally, supporting the notion that both sexes should be analyzed for USVs.

We found no differences between genotypes or sexes in average values of temporal parameters (Tables 4 and S3, Figure 2A,B). While there was no difference in the average number of songs between genotypes, *Foxp2*$^{hum/hum}$ mice showed a significantly larger variance in this parameter than *Foxp2*$^{+/+}$ mice (*Foxp2*$^{hum/hum}$: mean 31.13, standard deviation 28.60, coefficient of variance 91.88; *Foxp2*$^{+/+}$: 44.44, 11.56, 26.02; Levene's Test: $F = 5.16$, $p = 0.03$; Table 1).

Foxp2$^{hum/hum}$ mice emitted USV with significantly higher average and start frequencies, and larger frequency bandwidths than *Foxp2*$^{+/+}$ mice (Figure 2B,C, Table 4). The larger bandwidths are a consequence of the larger number of jump syllables used by *Foxp2*$^{hum/hum}$ mice (see subsection *Syllable types and syntax in songs of* Foxp2$^{hum/hum}$ *and* Foxp2$^{+/+}$ *mice*). Jump syllables have a larger bandwidth (mean [standard deviation]: 26.74 [10.40] kHz) than syllables containing no jumps (simple: 7.82 [5.72]; turn: 11.75 [6.74]), and hence the more frequent usage of syllable types with jumps led to a significantly larger bandwidth in *Foxp2*$^{hum/hum}$ as opposed to *Foxp2*$^{+/+}$ mice ($F = 4269.206$, $p < 0.001$; Table S5). We found no significant differences between the genotypes in the minimum frequency, the syllable duration, and the frequency slope (Table 4).

Female mice emitted USV syllables with smaller frequency bandwidths than male mice (Table 4). As we found no interactive effect of genotype and sex, found genotype differences can be attributed to both sexes, and sex differences to both genotypes.

The LDA analysis showed that the number of jump and turn syllables had the strongest influence on separating the groups (Table 5, Figure 3).

3.2 | Syllable types and syntax in songs of Foxp2$^{hum/hum}$ and Foxp2$^{+/+}$ mice

The vocal repertoire was the same for both genotypes, that is, they shared the same 12 syllable types, with simple upwards modulated syllables (SUP) being the most common (Table 3, Figure 4). However, genotypes differed significantly in the usage patterns of these types, that is, in their general use, in their use as start or stop syllable of a sequence (Table 6, Figure 4), and in the transitioning probabilities between syllable types (Table 6, Figure 5; please note that for better readability, Figure 5 only shows syllable transitions between three syllable groups, while statistical analysis was conducted on all 12 syllable types). *Foxp2*$^{hum/hum}$ mice used more complex syllables containing frequency jumps and less simple or turn syllables than *Foxp2*$^{+/+}$ mice (Table 3, Figure 4). The probability of emitting a certain syllable type in general was also different between the sexes: females used less jump syllables than males, fitting to the, on average, smaller frequency bandwidths of their calls (Tables 3 and 6, Figure 4).

To better understand the consequences of the differences in transitioning patterns between syllable types, we analyzed the occurrence of syllable type-doublets and -triplets (Figure 6). The most common transition type in both genotypes were doublets and triplets consisting of only SUP, the most common syllable type. However, *Foxp2*$^{hum/hum}$ mice additionally used many transitions containing jumps, mainly using the syllable type JED (jump early down; Figure 6).

For both genotypes and sexes the Markov model (MM) gave a better fit to the recorded syllable sequences than the Probability model (PM; Table 7, Figure 6). In both measured parameters, that is, the occurrence of syllable-type doublets and syllable-type triplets, the summed absolute differences between the respective values calculated from the MM generated syllable sequences were smaller than

TABLE 4 Results of statistical analysis for all temporal and spectral parameters of the Foxp2-Comparison and the SNCA-Comparison

		Foxp2-comparison		SNCA-comparison	
Parameter	**Factor(s)**	**F**	***p'***	**t/z**	***p'***
Temporal parameters					
Number of songs	gt	1.301	0.792	−0.925	0.364
	sex	0.3968	1.6023	−0.955	0.349
	gt*sex	0.3491	1.6788	−0.241	0.812
Syllables per song	gt	0.6284	1.2843	1.256	0.209
	sex	2.4698	0.3492	1.111	0.267
	gt*sex	2.3925	0.3669	−0.356	0.722
Song duration	gt	0.1298	1.4374	−0.366	0.718
	sex	0.0051	1.886	−0.249	0.805
	gt*sex	1.3605	0.441	−0.105	0.918
Syllables per second	gt	3.2802	0.0938	0.093	0.926
	sex	1.4219	0.3648	**2.018**	**0.044**
	gt*sex	0.1873	1.3306	0.105	0.917
Spectral parameters					
Average frequency	**gt**	**10.13**	**0.0304**	1.075	0.293
	sex	0.37	4.388	0.175	0.863
	gt*sex	0.15	5.64	−0.292	0.772
Start frequency	**gt**	**16.39**	**0.0032**	1.321	0.199
	sex	1.97	1.3784	0.408	0.687
	gt*sex	1.78	1.5464	−0.507	0.616
Minimum frequency	gt	1.64	1.6936	1.246	0.226
	sex	0.18	5.4024	−0.014	0.989
	gt*sex	0.17	5.4392	−0.189	0.851
Frequency bandwidth	**gt**	**102.81**	**<0.0008**	−0.97	0.33
	sex	**46.527**	**<0.0008**	0.22	0.83
	gt*sex	3.904	0.4712	−0.15	0.88
Syllable duration	gt	0.0002	7.9088	−0.72	0.47
	sex	1.3054	2.0264	0.28	0.78
	gt*sex	1.8637	1.3776	0.04	0.97
Frequency slope	gt	5.2838	0.2384	−0.468	0.644
	sex	0.6833	3.328	−0.372	0.714
	gt*sex	1.8173	1.5144	−0.136	0.893

Note: The results for each statistical comparison between the effects of genotype (gt), sex and their interaction (gt*sex) are presented. *p* values are corrected for multiple testing, significant effects are printed in bold. For summary statistics all temporal and spectral parameters of the Foxp2-Comparison and the SNCA-Comparison see Tables S3 and S4, respectively.

those calculated from the PM generated syllable sequences (Table 7, Figure 6).

3.3 | Temporal and spectral structure of USV of Foxp2$^{hum/hum}$ mice with SNCA-deletion

From the 32 mice of the SNCA-Comparison we recorded 2049 songs, containing 13,424 syllables over the 12 hours of recording. The individual number of songs per mouse in these 12 hours ranged from 11 (a male *Foxp2*$^{hum/hum\text{-}SNCAwt}$) up to 101 songs (a female *Foxp2*$^{hum/hum\text{-}SNCADel}$) (Table 2). Note that the different number and variance of calls numbers of the Foxp2- and the SNCA-Comparison was likely influenced by the different recording setups used (see Material and Methods section *USV recording*).

We found no differences in neither temporal nor spectral structure of songs or syllables between *Foxp2*$^{hum/hum\text{-}SNCADel}$ and *Foxp2*$^{hum/hum\text{-}SNCAwt}$ mice (Tables 4 and S4). The only significant difference we found in this data set, was a higher syllable rate in female than male mice (Tables 4 and S4).

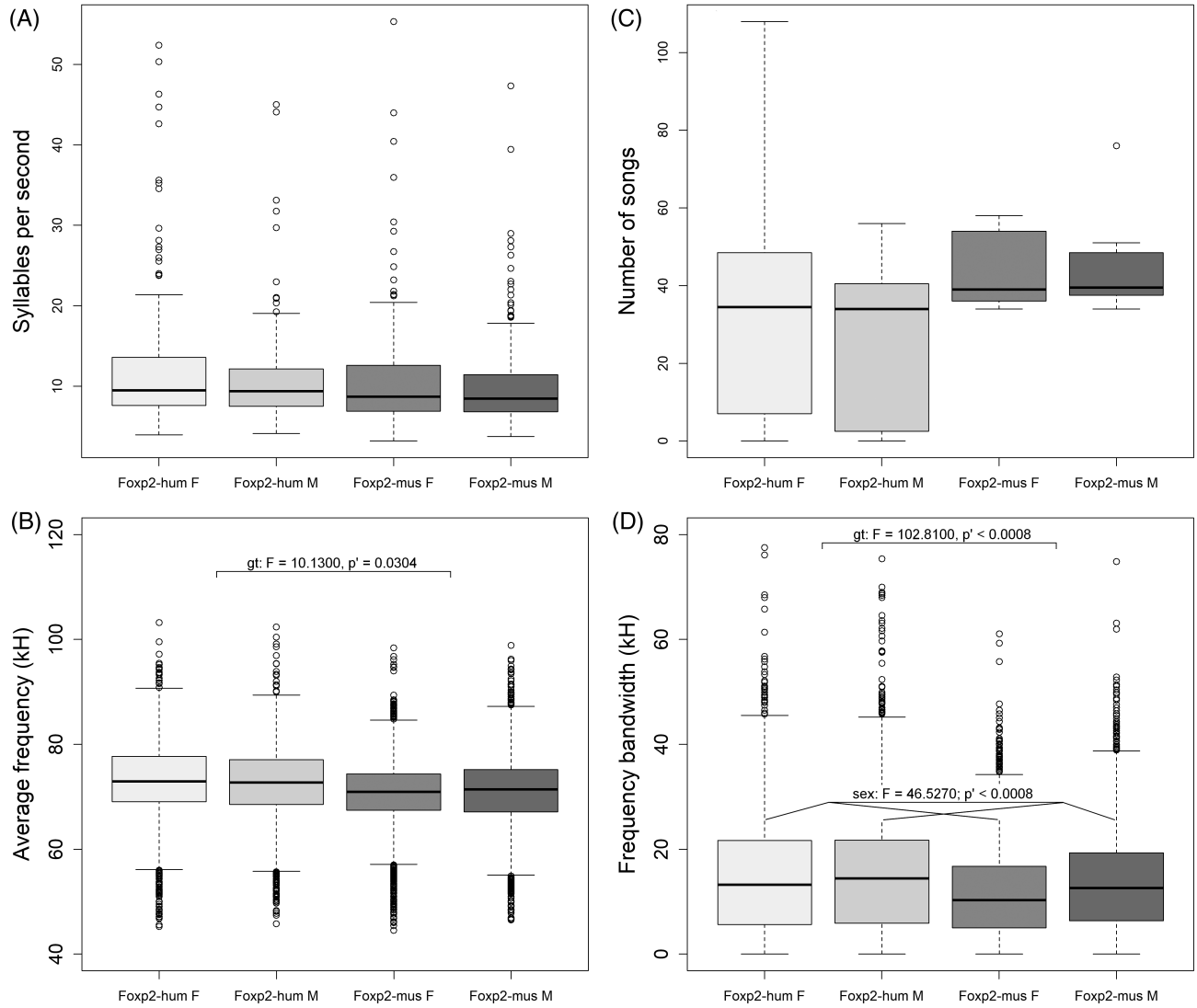

FIGURE 2 Boxplots showing (A) number of the syllables emitted per each second of song, (B) the number of songs emitted over the whole night, (C) the average frequency of syllables, and (D) their frequency bandwidth. Groups are: Foxp2-hum F = female Foxp2hum/hum, Foxp2-hum M = male Foxp2hum/hum, Foxp2-mus F = female Foxp2+/+, Foxp2-mus M = male Foxp2+/+. Result of comparison of the syllables per second between genotypes is given; all other comparison were non-significant

TABLE 5 Results of the linear discriminant analysis for the structural data. For each parameter the coefficients of the three linear discriminants (LD) are given. For each linear discriminant the loading is given at the bottom

Parameter	LD1	LD2	LD3
Average frequency	0.0574	−0.0662	0.0273
Start frequency	0.0324	−0.1185	0.0645
Minumum frequency	−0.1507	0.1528	−0.0135
Frequency band	−0.0818	−0.0171	0.0063
Syllable duration	0.0271	−0.0156	−0.0120
Frequency slope	0.2004	−0.4018	−0.0943
Number of jumps	−0.5526	0.4675	−1.4984
Number of turns	−0.3985	0.2629	0.3622
Loadings of discriminant functions:	0.8844	0.0838	0.0318

4 | DISCUSSION

4.1 | Ultrasonic vocalization of mice with a humanized Foxp2 gene

In this study, we analyzed the ultrasonic vocalization of adult female and male mice carrying a humanized version of the murine *Foxp2* gene (*Foxp2*$^{hum/hum}$), which differs by two amino acid substitutions from the natural murine *Foxp2* (*Foxp2*$^{+/+}$). We found significant differences in the temporal, structural and syntactical composition of USV between *Foxp2*$^{hum/hum}$ and *Foxp2*$^{+/+}$ mice. *Foxp2*$^{hum/hum}$ mice had a larger variation in the number of songs emitted by individuals, used syllables with higher average and starting frequencies, larger frequency bandwidths, and more complex syllables (i.e., containing frequency jumps or turns) than *Foxp2*$^{+/+}$ mice. As syllables containing jumps usually

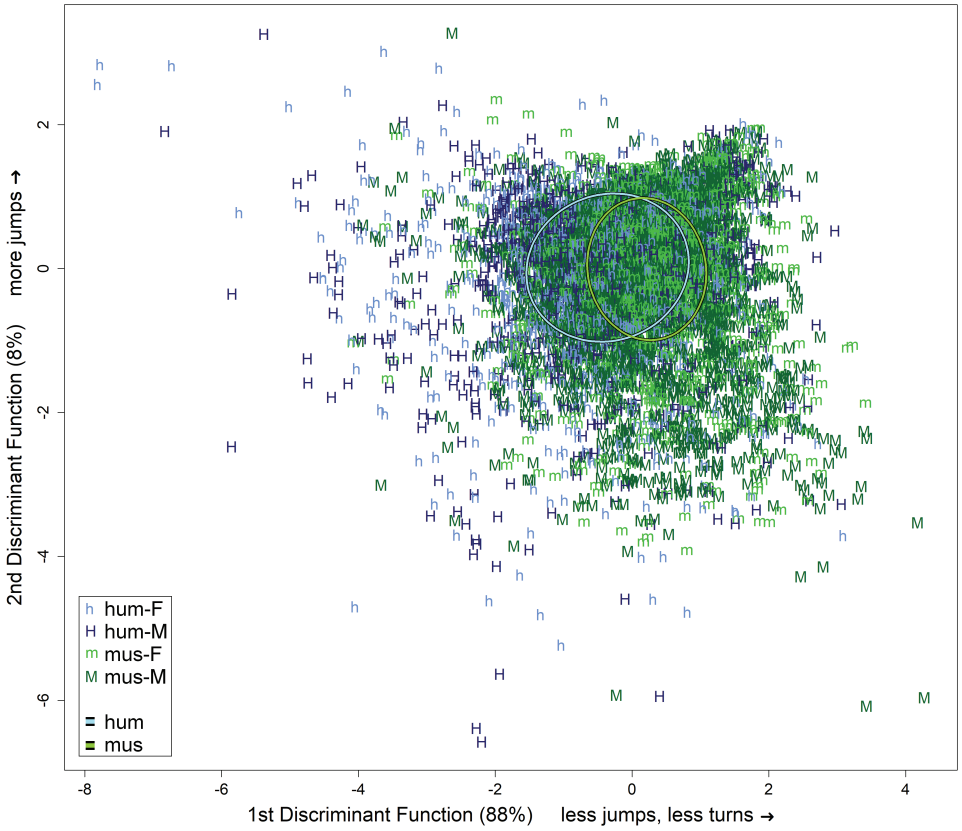

FIGURE 3 LDA plot of the first two linear discriminants for syllable parameters. Each point represents one syllable and is colored and shaped according to genotype and sex of the singing animal. Ellipses show the standard deviation for genotypes. Groups are Foxp2hum/hum females and males (hum-F: the letter h, light blue; hum-M: the letter H, dark blue; standard deviation: blue ellipse) and Foxp2+/+ females and males (mus-F: the letter m, light green; mus-M: the letter H, dark green; standard deviation: green ellipse). The loading of each linear discriminant function is given as percentage at the respective axis

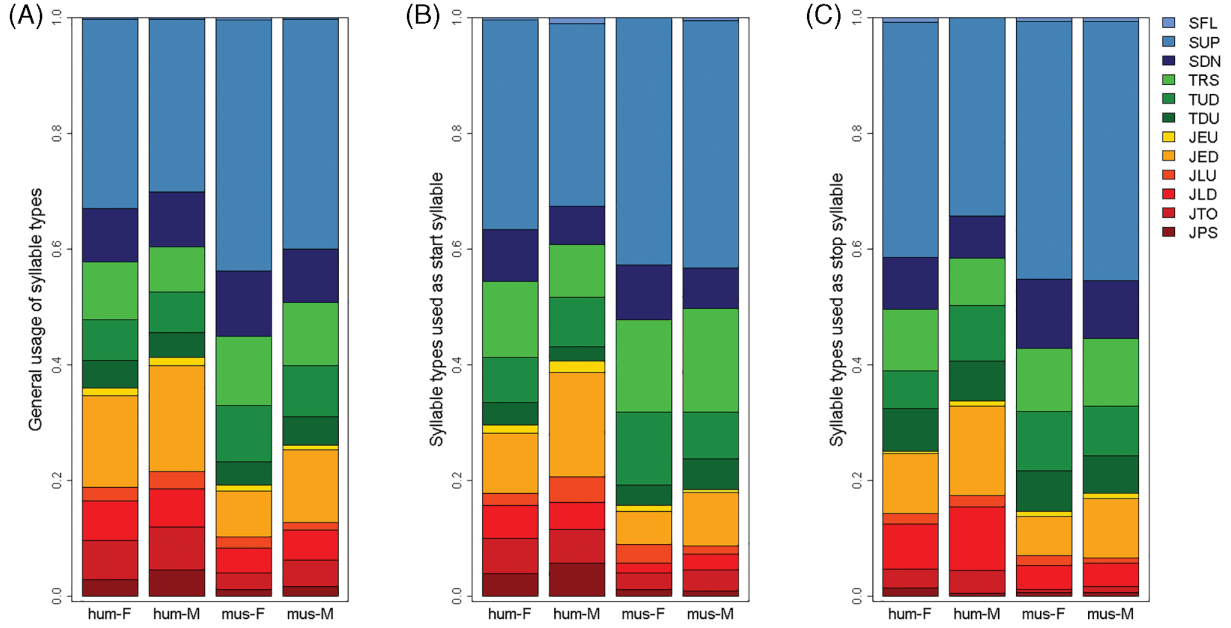

FIGURE 4 Probability of the four groups to use any of the 12 syllable types. The three panels show the probability to use syllable types (A) in general, (B) as start or (C) as stop syllable of a sequence of syllables. Groups are: hum-F = female Foxp2hum/hum, hum-M = male Foxp2hum/hum, mus-F = female Foxp2+/+, mus-M = male Foxp2+/+. Syllable types are labeled with abbreviations that are further explained in Table 3 and Method section Song and syllable parameters

have larger frequency bandwidths, the higher number of syllables containing jumps in *Foxp2*$^{hum/hum}$ mice explains the larger bandwidths used by this genotype.

Differences between *Foxp2*$^{hum/hum}$ and *Foxp2*$^{+/+}$ mice have already been found for the innate USV calls of mouse pups.[40] Pups carrying the humanized version of the murine *Foxp2* gene and their

	genotype			sex		
	Chi^2	df	p'	Chi^2	df	p'
General syllable usage	**226.47**	**11**	**<0.001**	**27.08**	**11**	**0.004**
Usage as start syllable	**55.88**	**11**	**<0.001**	12.99	11	0. 294
Usage as stop syllable	**35.59**	**11**	**<0.001**	8.05	11	0.709
Syllable transitions	**413.76**	**99**	**<0.001**	115.96	99	0.234

Note: Significant results, printed in bold, indicate that the probability to use a certain syllable type, to start or stop a sequence with a certain syllable type, or to use a certain syllable type transition differed between the tested groups. *p* values are corrected for multiple testing.

TABLE 6 Results of the syntax analysis, showing the comparison of the usage of the 12 syllable types between genotypes and between sexes

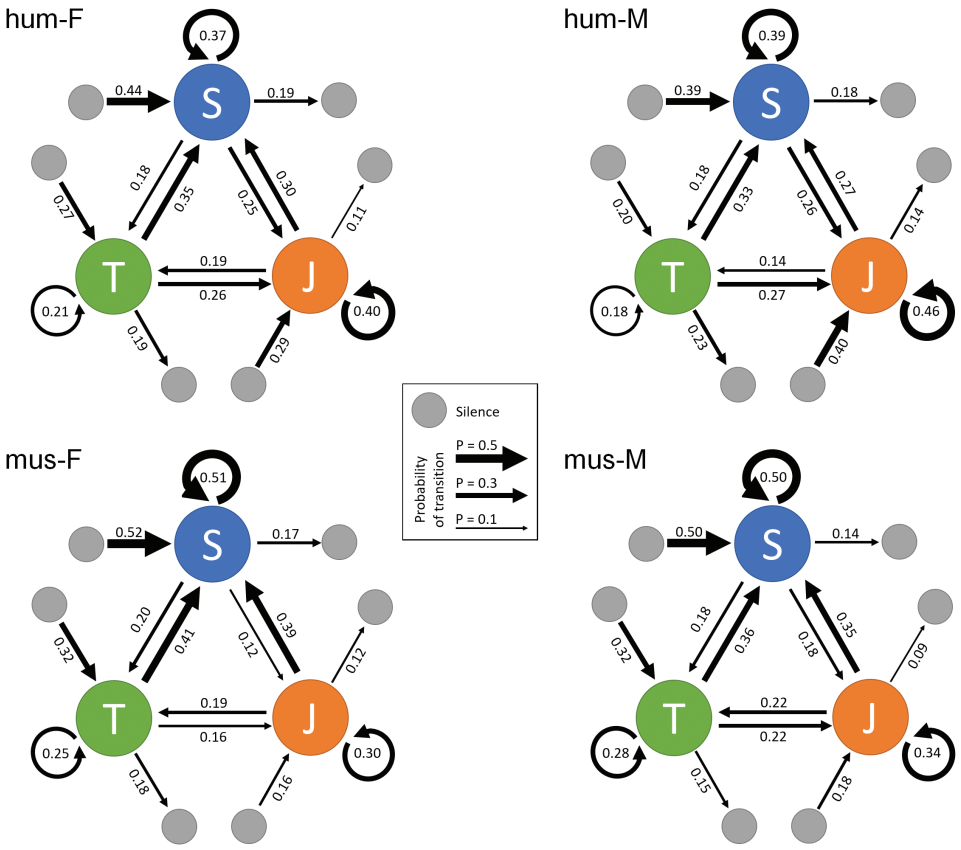

FIGURE 5 Transition probabilities between syllable type groups. For better readability, the 12 syllable types used in this study were grouped into three syllable groups: simple syllable types (blue S), syllable types containing one or more turns (green T) and syllable types containing one or more jumps (orange J). Please note that for all statistical analyses, all 12 syllable types were used. Gray circles represent the beginning or end of a sequence. Numbers show the probability to move from one state (i.e., beginning or syllable type group) to another state (i.e., syllable type group or end). Arrows show the direction of the respective transition and arrow thickness is relative to the probability of this transition. For each group of mice, one diagram is shown. Groups are: hum-F = female Foxp2hum/hum, hum-M = male Foxp2hum/hum, mus-F = female Foxp2+/+, mus-M = male Foxp2+/+

wildtype littermates emit typical isolation calls, but *Foxp2*$^{hum/hum}$ pups used significantly lower average calling frequencies (74 ± 0.75 kHz) than *Foxp2*$^{+/+}$ pups (78 ± 0.76 kHz; see Supplemental Table S7 in[40]). In contrast, we found higher calling frequencies in adult *Foxp2*$^{hum/hum}$ mice. As isolation calls of pups and USV of adult mice serve a different purpose and might ontologically not be related, these differences are not contradictory.

Our results are partly in contrast to those published by Hammerschmidt et al.[44] They had recorded adult mice of the same strains, but found almost no difference between the genotypes in any of the tested USV parameters. However, in one of their analyses they found a hint that *Foxp2*$^{hum/hum}$ mice produced syllables with slightly more pronounced frequency jumps and a slightly earlier frequency maximum than *Foxp2*$^{+/+}$ mice. Interestingly, this tendency coincides with our result that both female and male *Foxp2*$^{hum/hum}$ mice emitted more jump syllables.

There were some methodological differences between the earlier and our current study that seem important to be discussed in detail. In contrast to our study, Hammerschmidt et al.[44] had recorded female and male mice, of the same and different genotypes, together. Female mice of different strains, including the current ones, have been shown to vocalize in many contexts (Venerosi et al.[49]; Moles et al. [22]; Hammerschmidt et al.[20]; von Merten etal.[30]; Hoier et al.[21]; Warren et al.[50]; the current study) including mating.[48] By recording different genotypes together, some of the recordings likely contained a mix of USV from *Foxp2*$^{hum/hum}$ and *Foxp2*$^{+/+}$ mice, thus masking the true differences between genotypes. Further, Hammerschmidt et al.[44] had recorded USV for only 2 min. This short duration was chosen due to

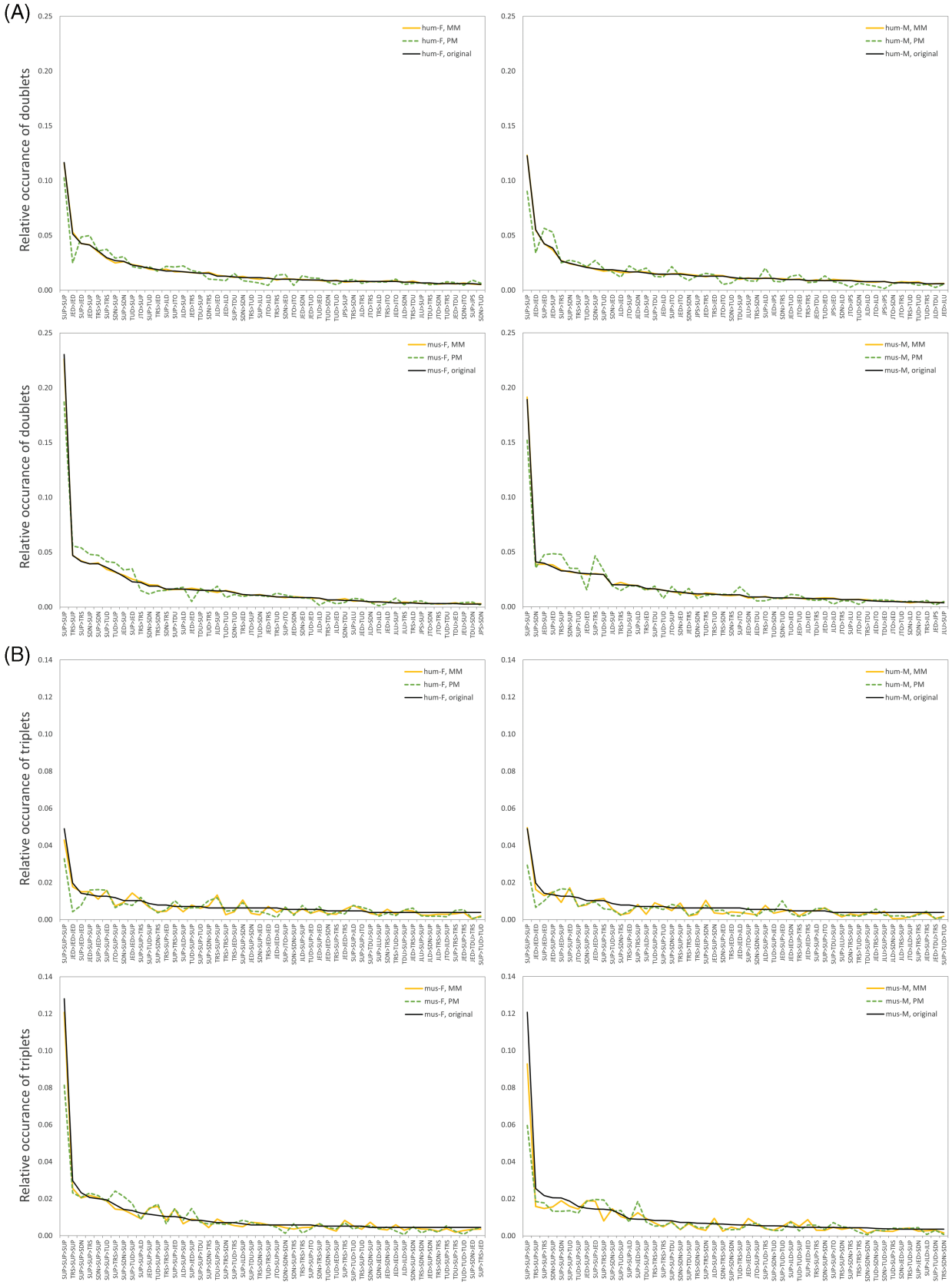

FIGURE 6 Relative occurrence of (A) syllable type-doublets and (B) syllable type-triplets. Doublets and triplets are sorted according to decreasing probability in the real data. Graphs are separated by groups of mice: hum-F = female Foxp2hum/hum, hum-M = male Foxp2hum/hum, mus-F = female Foxp2+/+, mus-M = male Foxp2+/+. Solid black lines show the distribution of doublets and triplets in the observed syllable sequences (original); dotted green and dashed yellow lines show the distribution of the respective doublets and triplets in the syllable sequences calculated with the Probability model (PM) and the Markov model (MM), respectively

TABLE 7 Results of the syntax analysis, showing the comparison of the fit of the syllable-sequences generated by the Probability model (PM) or the Markov model (MM) to the recorded syllable-sequences

	female *Foxp2*$^{hum/hum}$		male *Foxp2*$^{hum/hum}$		female *Foxp2*$^{+/+}$		male *Foxp2*$^{+/+}$	
	$d_{SUM\text{-}PM}$[a]	$d_{SUM\text{-}MM}$[a]	$d_{SUM\text{-}PM}$	$d_{SUM\text{-}MM}$	$d_{SUM\text{-}PM}$	$d_{SUM\text{-}MM}$	$d_{SUM\text{-}PM}$	$d_{SUM\text{-}MM}$
Doublets	0.2454	**0.0426**	0.3340	**0.0416**	0.2662	**0.0445**	0.2653	**0.0389**
Triplets	0.6919	**0.5627**	0.8396	**0.6662**	0.5973	**0.4491**	0.5829	**0.4297**

Note: Fits are given as distances between model data and real date, that is, the larger the number the worse the fit. Better fits are printed in bold.
[a]$d_{SUM\text{-}PM}$ and $d_{SUM\text{-}MM}$: distance between original data and those generated by the PM or MM, respectively. See Methods section *Syntax analysis* for formulas.

their observation that mice fall silent after this time. While we have also observed a reduction of vocalizations over time, our mice produced repeated bouts of vocalization along the night (see Figure S2). Mice have been shown to change both courtship and vocal behavior over time, likely a consequence of an increase in information gathered, as well as possible changes of their internal state and/or external circumstances over time.[21,46,58,59] Recording vocalization over extended periods of time thus likely allows to collect vocalizations that are less biased by short-term effects. Hammerschmidt et al.[44] had recorded, during 2 min, from each pair of mice on average 424 syllables. Even taking into account that two mice were recorded together, this is considerably more than we recorded in a comparable time interval, that is, on average 17 syllables during the first 2 min of vocalization per mouse. One possible explanation for the higher number of USV in the previous study, is the direct physical contact between females and males, which was not possible in our setup. Ideally, a setup should allow direct contact but consider vocalization of all recorded mice, like recently shown by Sangiamo et al.[59] Such a setup would also allow to compare the effect of different recording times and analyze a possible change of vocalization over time.

Apart from mice carrying the humanized version of the murine *Foxp2*, other mouse lines with alterations, that is, disruptions or deleterious mutations, of the *Foxp2* gene have been studied.[34–39] As homozygotes of these lines usually do not survive to adulthood, heterozygotes are used to analyze adult USV. The respective studies found that such heterozygous mice (here termed *Foxp2*$^{+/alt}$) tend to emit shorter songs with a lower syllable rate than their wildtype littermates.[37–39] Interestingly, while *Foxp2*$^{+/alt}$ mice had a lower syllable rate, we found the syllable rate rather higher in *Foxp2*$^{hum/hum}$ mice as opposed to the wildtype. While this difference was non-significant, it still complements very well previous results on changes in *Foxp2*$^{hum/hum}$ mice as opposed to changes in mice heterozygous for non-functional *Foxp2* alleles: When compared to the respective wildtype lines, *Foxp2*$^{hum/hum}$ mice often show the opposite effects to *Foxp2*$^{+/alt}$ mice, e.g., in exploratory behavior and dopamine levels.[40] Our results thus confirm the notion that *Foxp2*$^{hum/hum}$ mice are interesting candidates for the study of the evolution of human speech and language.

As opposed to the mentioned studies on the USV of adult mice with heterozygous alterations of the *Foxp2* gene, we additionally found differences in the frequency of USV between genotypes. The differences in frequency values between genotypes and sexes in this study were only between 1 and 3 kHz, and thus seem quite small in relation to the overall range of USV (from about 25 up to 145 kHz). While an older study suggested that mice are able to discriminate frequency changes of 1 kHz,[60] a newer study showed a rather limited frequency resolution in mice.[61] Nevertheless, as *Foxp2*$^{hum/hum}$ mice are an artificial creation,[40] even a small difference to their wildtype littermates can hint at differences in the vocal production of these genetically engineered mice. Thus, even if they were not biologically relevant in nature, they can still serve as an interesting model system.

Foxp2$^{hum/hum}$ mice showed a larger variation in the number of songs than *Foxp2*$^{+/+}$ mice and had both the highest and lowest song emission rate. Results from other studies suggest that the amount of vocalization uttered by a mouse is an individual characteristic, with some mice being very vocal, while others hardly vocalize at all.[21,29,30] Our results show that *Foxp2*$^{hum/hum}$ mice are possibly more diverse in this characteristic. It might be worthwhile to record individuals of both genotypes repeatedly to assess the influence of individuality on emission rate.[62]

The SNCA-Comparison dataset showed that neither *SNCA*, nor possible other genetic background differences that could have occurred during the backcross of another WT strain have a major influence on the USV patterns. *SNCA* codes for a protein required for synaptic activity and has been implicated in Parkinson's disease,[63] but also in other cognitive functions, such as musical performance and perception.[64,65,66] Also in birds, the corresponding gene synelfin has been correlated with plasticity in the developing song control system.[67] *SNCA* could thus have been a good candidate for also being involved in USV emission or perception in mice, but our data provide no direct evidence for this, at least in combination with the humanized *Foxp2* allele.

4.2 | Syllable-type usage and syntax in mice with a humanized Foxp2 gene

The syllable repertoire was similar in both genotypes, that is, all syllable types emitted by one genotype were also emitted by the other and vice versa. Simple upwards modulated syllables were the most frequent syllable type, which is typical for wild and laboratory mice.[30,53] The usage pattern of those syllable types, however, differed between genotypes. *Foxp2*$^{hum/hum}$ mice used more complex jump

syllables and less simple syllables than *Foxp2*$^{+/+}$ mice. Indeed, the number of jumps had the strongest influence on separating the four groups (two genotypes and two sexes).

The genotypes differed not only in the proportional usage of syllable types but also in their sequencing within a song. Fitting to the general usage of syllable types, *Foxp2*$^{hum/hum}$ mice showed more transitions from and to jump syllables than *Foxp2*$^{+/+}$ mice, including repetitions of jump syllables. We could, however, not show a difference in the complexity of syntax between *Foxp2*$^{hum/hum}$ and *Foxp2*$^{+/+}$ mice. Both genotypes showed a sequencing of syllable types that was more complex than the mere probability of syllable usage or a simple Markov model suggested.

These results are in concordance with previous studies analyzing syntax of *Foxp2* mutant mice: Adult mice heterozygous for deleterious alterations of the *Foxp2* gene (*Foxp2*$^{+/alt}$) use the same syllable repertoire as their wildtype littermates but show quantitative differences in the usage of these syllable types and changes in their sequencing, that is, the syntax.[37–39] Like in the general USV differences discussed above, the differences between *Foxp2*$^{hum/hum}$ and *Foxp2*$^{+/+}$ mice are opposite to those between *Foxp2*$^{+/alt}$ and *Foxp2*$^{+/+}$ mice. While the mice carrying the humanized version of the *Foxp2* gene used more jump syllables and repeated them in higher numbers, mice carrying a disruption or deleterious mutation of this gene used less complex syllable types.[37,38]

It is not well understood which factors determine the probability of uttering certain syllable types. Strain specific patterns can play a role[26] as well as the type of social interaction, both within and between females and males.[21–23,45,47,59,68] Additionally, the complexity of syllables seems to play a role in mate attraction. Male mice adjust the production of complex syllables, depending on the olfactory stimulus with which they are confronted[23] and the sex of an intruding animal.[68] It has further been suggested that complex syllables might facilitate the attraction of females.[69] Indeed, while both sexes vocalize in a mating context, males emit more songs than females, likely to evoke their receptivity.[48] Fitting to this, the male mice in our study emitted more complex syllables with larger frequency bandwidths than females. As complex syllables and high syllable rates are considered more costly to produce,[70] the number of complex syllables in courtship songs could therefore serve as a signal for a male's individual fitness.

In the current study, *Foxp2*$^{hum/hum}$ mice emitted not only more complex syllable types with larger frequency bandwidths than *Foxp2*$^{+/+}$ mice, but also showed the tendency to have a higher call rate, although this result was not significant. Vocalizing animals usually face a trade-off between the frequency bandwidth (complexity) of their calls and the call rate, that is, the number of vocalizations produced within a certain time interval: the larger the bandwidth, the lower the call rate. This is known from vocalizations of different species, among them songbirds, bats, and mice,[70–72] and can mainly be explained by limitations in vocal production. The finding that *Foxp2*$^{hum/hum}$ mice emitted more large-bandwidth, complex syllable types at a possibly higher call rate than *Foxp2*$^{+/+}$ mice might indicate that the *Foxp2*^{hum} gene influenced the brain circuits responsible for the complex orofacial and laryngeal movements required during sound production.[9,10]

4.3 | CONCLUSION AND OUTLOOK

Our results suggest that the two human-specific amino acid substitutions in *Foxp2* do not only affect mouse pup vocalizations, but also the USV of adult female and male mice. This complements previous evidence on the influence of the humanized *Foxp2* gene on several neurophysiological and behavioral factors in mice: *Foxp2*$^{hum/hum}$ mice show a higher expression of the Foxp2 protein, neuromorphological differences and changed dopamine levels in the striatum, as well as accelerated learning as opposed to wildtype littermates.[40,42,43] Additionally, *Foxp2*$^{hum/hum}$ pups have been shown to exhibit less exploratory behavior and emit qualitatively different USV.[40] It is still not clear in which way *Foxp2* acts on mouse USV and if it can be at all related to the change in *FoxP2* function during human evolution. It is however interesting to notice that our results confirm the notion that mice carrying the humanized version of the *Foxp2* gene show opposite changes to that of mice carrying disrupted or mutated alleles of this gene. These combined results fit well with the vocal-learning continuum hypothesis, which suggests that the ability for vocal learning is not a dichotomous trait but rather forms a continuum ([33,73,74]). This renders the *Foxp2*$^{hum/hum}$ mouse an important model to study the evolution of human speech and language. Studies to date have shown both an influence of *Foxp2* on innate pup vocalizations[34,40,75] and different types of learning, including auditory-motor association learning.[35,42,76] Adult mouse USV seems to be mainly innate,[31,32] but mice seem to depend on auditory feedback to maintain certain features of their USV.[27,33] Hence, the increased complexity in USV of mice carrying the humanized version of *Foxp2* found in the current study, in combination with the accelerated learning of those mice found in previous studies suggest a possible function of *Foxp2* in vocal learning in mice. This hypothesis is strengthened by the fact that in birds, humans, and some other mammals, *FoxP2* has already been connected to vocal learning.[4,6,12] As several neurophysiological processes involved in vocalization and learning in mice overlap with respective processes in humans,[6,40,42] it might be worthwhile to explore vocal learning patterns in *Foxp2*$^{hum/hum}$ mice.

ACKNOWLEDGMENTS

We would like to thank Wolfgang Enard for providing the mouse strain used in this study. Further, many thanks to Heike Harre for outstanding technical assistance; Cornelia Burghardt, Ellen McConnell, and Michaela Schwarz for assistance in genotyping; the whole Plön mouse team for taking such good care of the mice; Bernhard Haubold for providing the script for double recordings; and Miriam Linnenbrink for discussion. We would further like to thank Wolfgang Enard, Svante Pääbo and Klaus Hammerschmidt for discussion. Additional thanks to three anonymous reviewers for their helpful comments on an earlier version of our manuscript. This study was funded by the Max Planck Society. Additional funding provided by CESAM (UIDP/50017/2020+UIDB/50017/2020) through national funds by FCT/MCTES and an FCT post-doc fellowship to SvM (SFRH/BPD/118053/2016). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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

ORCID

Sophie von Merten https://orcid.org/0000-0001-6010-1127
Diethard Tautz https://orcid.org/0000-0002-0460-5344

REFERENCES

1. Boughman JW, Moss CF. Social sounds: vocal learning and development of mammal and bird calls. *Acoustic communication*. Springer; 2003:138-224.
2. Janik VM. Cetacean vocal learning and communication. *Curr Opin Neurobiol*. 2014;28:60-65.
3. Janik VM, Slater PB. Vocal learning in mammals. *Adv Stud Behav*. 1997;26:59-99.
4. Knörnschild M. Vocal production learning in bats. *Curr Opin Neurobiol*. 2014;28:80-85.
5. Reichmuth C, Casey C. Vocal learning in seals, sea lions, and walruses. *Curr Opin Neurobiol*. 2014;28:66-71.
6. Scharff C, Petri J. Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Phil T Roy Soc B*. 2011;366:2124-2140.
7. Co M, Anderson AG, Konopka G. FOXP transcription factors in vertebrate brain development, function, and disorders. *WIREs Dev Biol*. 2020;9:e375.
8. Takahashi K, Liu F-C, Hirokawa K, Takahashi H. Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *Journal Neurosci Res*. 2003;73:61-72.
9. Fisher SE, Marcus GF. The eloquent ape: genes, brains and the evolution of language. *Nat Rev Genet*. 2006;7:9-20.
10. Scharff C, Haesler S. An evolutionary perspective on FoxP2: strictly for the birds? *Curr Opin Neurobiol*. 2005;15:694-703.
11. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 2001;413:519-523.
12. Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol*. 2007;5:e321.
13. Li G, Wang J, Rossiter SJ, Jones G, Zhang S. Accelerated FoxP2 evolution in echolocating bats. *PLoS ONE*. 2007;2:e900.
14. Yin J-X, Ruan Y-N, Liu J-L, Zhang S-Y, Racey P. FoxP2 expression in an echolocating bat (*Rhinolophus ferrumequinum*): Functional implications. *Mamm Biol*. 2017;6:24-29.
15. Enard W. FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr Opin Neurobiol*. 2011;21:415-424.
16. Enard W. The Molecular Basis of Human Brain Evolution. *Curr Biol*. 2016;26:R1109-R1117.
17. Enard W, Przeworski M, Fisher SE, et al. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*. 2002;418:869-872.
18. Zhang J, Webb DM, Podlaha O. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics*. 2002;162:1825-1835.
19. Atkinson EG, Audesse AJ, Palacios JA, et al. No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. *Cell*. 2018;174:1424-1435.e15.
20. Hammerschmidt K, Radyushkin K, Ehrenreich H, Fischer J. The structure and usage of female and male mouse ultrasonic vocalizations reveal only minor differences. *PLoS ONE*. 2012;7:e41133.
21. Hoier S, Pfeifle C, von Merten S, Linnenbrink M. Communication at the Garden Fence–Context Dependent Vocalization in Female House Mice. *PLoS ONE*. 2016;11:e0152255.
22. Moles A, Costantini F, Garbugino L, Zanettini C, D'Amato FR. Ultrasonic vocalizations emitted during dyadic interactions in female mice: a possible index of sociability? *Behav Brain Res*. 2007;182:223-230.
23. Musolf K, Hoffmann F, Penn DJ. Ultrasonic courtship vocalizations in wild house mice, *Mus musculus musculus*. *Anim Behav*. 2010;79:757-764.
24. Nyby J. Ultrasonic vocalizations during sex behavior of male house mice (*Mus musculus*): A description. *Behav Neural Biol*. 1983;39:128-134.
25. Holy TE, Guo Z. Ultrasonic Songs of Male Mice. *PLoS Biol*. 2005;3:e386.
26. Asaba A, Okabe S, Nagasawa M, et al. Developmental Social Environment Imprints Female Preference for Male Song in Mice. *PLoS ONE*. 2014;9:e87186.
27. Chabout J, Serreau P, Ey E, et al. Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PLoS ONE*. 2012;7:e29401.
28. Grimsley JM, Monaghan JJ, Wenstrup JJ. Development of social vocalizations in mice. *PLoS ONE*. 2011;6:e17460.
29. Hoffmann F, Musolf K, Penn DJ. Spectrographic analyses reveal signals of individuality and kinship in the ultrasonic courtship vocalizations of wild house mice. *Physiol Behav*. 2012;105:766-771.
30. von Merten S, Hoier S, Pfeifle C, Tautz D. A Role for Ultrasonic Vocalisation in Social Communication and Divergence of Natural Populations of the House Mouse (*Mus musculus domesticus*). *PLoS ONE*. 2014;9:e97244.
31. Hammerschmidt K, Reisinger E, Westekemper K, Ehrenreich L, Strenzke N, Fischer J. Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci*. 2012;13:40.
32. Kikusui T, Nakanishi K, Nakagawa R, Nagasawa M, Mogi K, Okanoya K. Cross Fostering Experiments Suggest That Mice Songs Are Innate. *PLoS ONE*. 2011;6:e17721.
33. Arriaga G, Zhou EP, Jarvis ED. Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS ONE*. 2012;7:e46610.
34. Heckman J, McGuinness B, Celikel T, Englitz B. Determinants of the mouse ultrasonic vocal structure and repertoire. *Neurosci Biobehav R*. 2016;65:313-325.
35. Groszer M, Keays DA, Deacon RM, et al. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol*. 2008;18:354-362.
36. Shu W, Cho JY, Jiang Y, et al. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *P Natl Acad Sci USA*. 2005;102:9643-9648.
37. Castellucci GA, McGinley MJ, McCormick DA. Knockout of Foxp2 disrupts vocal development in mice. *Sci Rep*. 2016;6:23305.
38. Chabout J, Sarkar A, Patel SR, et al. A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. *Front Behav Neurosci*. 2016;10:197.
39. Gaub S, Fisher SE, Ehret G. Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. *Genes Brain Behav*. 2016;15:243-259.
40. Enard W, Gehre S, Hammerschmidt K, et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*. 2009;137:961-971.
41. Reimers-Kipping S, Hevers W, Pääbo S, Enard W. Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience*. 2011;175:75-84.

42. Schreiweis C, Bornschein U, Burguière E, et al. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *P Natl Acad Sci USA*. 2014;111:14253-14258.
43. Schreiweis C, Irinopoulou T, Vieth B, et al. Mice carrying a humanized Foxp2 knock-in allele show region-specific shifts of striatal Foxp2 expression levels. *Cortex*. 2019;118:212-222.
44. Hammerschmidt K, Schreiweis C, Minge C, Pääbo S, Fischer J, Enard W. A humanized version of Foxp2 does not affect ultrasonic vocalization in adult mice. *Genes Brain Behav*. 2015;14:583-590.
45. D'Amato FR, Moles A. Ultrasonic vocalizations as an index of social memory in female mice. *Behav Neurosci*. 2001;115:834.
46. Ey E, de Chaumont F, Bourgeron T. Spontaneous social communication in laboratory mice - placing ultrasonic vocalizations in their behavioral context. *bioRxiv*. 2020. https://doi.org/10.1101/2020.07.09.195362.
47. Moles A, D'Amato FR. Ultrasonic vocalization by female mice in the presence of a conspecific carrying food cues. *Anim Behav*. 2000;60:689-694.
48. Neunuebel JP, Taylor AL, Arthur BJ, Egnor SR. Female mice ultrasonically interact with males during courtship displays. *Elife*. 2015;4:e06203.
49. Venerosi A, Calamandrei G, Ricceri L. A social recognition test for female mice reveals behavioral effects of developmental chlorpyrifos exposure. *Neurotoxicol Teratol*. 2006;28:466-471.
50. Warren MR, Sangiamo DT, Neunuebel JP. High channel count microphone array accurately and precisely localizes ultrasonic signals from freely-moving mice. *J Neurosci Meth*. 2018;297:44-60.
51. Arriaga G, Jarvis ED. Mouse vocal communication system: Are ultrasounds learned or innate? *Brain Lang*. 2013;124:96-116.
52. Hammerschmidt K, Radyushkin K, Ehrenreich H, Fischer J. Female mice respond to male ultrasonic 'songs' with approach behaviour. *Biol Lett*. 2009;5:589-592.
53. Scattoni ML, Ricceri L, Crawley JN. Unusual repertoire of vocalizations in adult BTBR T+ tf/J mice during three types of social encounters. *Genes Brain Behav*. 2011;10:44-56.
54. Keithley EM, Canto C, Zheng QY, Fischel-Ghodsian N, Johnson KR. Age-related hearing loss and the ahl locus in mice. *Hearing Res*. 2004;188:21-28.
55. Someya S, Xu J, Kondo K, et al. Age-related hearing loss in C57BL/6J mice is mediated by Bak-dependent mitochondrial apoptosis. *P Natl Acad Sci USA*. 2009;106:19432-19437.
56. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2017.
57. Levin DA, Peres Y, Wilmer EL. *Markov Chains and Mixing Times*. AMS Bookstore; 2009.
58. Linnenbrink M, von Merten S. No speed dating please! Patterns of social preference in male and female house mice. *Front Zool*. 2017;14:38.
59. Sangiamo DT, Warren MR, Neunuebel JP. Ultrasonic signals associated with different types of social behavior of mice. *Nat Neurosci*. 2020;23:411-422.
60. Ehret G. Frequency and intensity difference limens and nonlinearities in the ear of the housemouse (*Mus musculus*). *J Comp Physiol*. 1975;102:321-336.
61. Portfors CV, Mayko ZM, Jonson K, Cha GF, Roberts PD. Spatial organization of receptive fields in the auditory midbrain of awake mouse. *Neuroscience*. 2011;193:429-439.
62. Bell AM. Animal Personalities. *Nature*. 2007;447:539-540.
63. Singleton AB. alpha-Synuclein Locus Triplication Causes Parkinson's Disease. *Science*. 2003;302:841-841.
64. Kanduri C, Raijas P, Ahvenainen M, et al. The effect of listening to music on human transcriptome. *PeerJ*. 2015;17:e830.
65. Nair PS, Raijas P, Ahvenainen M, Philips AK, Ukkola-Vuoti L, Järvelä I. Music-listening regulates human microRNA expression. *Epigenetics*. 2021;16(5):554-566. https://doi.org/10.1080/15592294.2020.1809853
66. Kanduri C, Kuusi T, Ahvenainen M, Philips AK, Lähdesmäki H, Järvelä I. The effect of music performance on the transcriptome of professional musicians. *Sci Rep*. 2015;5:9506.
67. Clayton DF, George JM. The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease. *Trends Neurosci*. 1998;21:249-254.
68. Panksepp JB, Jochman KA, Kim JU, et al. Affiliative behavior, ultrasonic communication and social reward are influenced by genetic variation in adolescent mice. *PLoS ONE*. 2007;2:e351.
69. Chabout J, Sarkar A, Dunson DB, Jarvis ED. Male mice song syntax depends on social contexts and influences female preferences. *Front Behav Neurosci*. 2015;9:76.
70. Pasch B, George AS, Campbell P, Phelps SM. Androgen-dependent male vocal performance influences female preference in Neotropical singing mice. *Anim Behav*. 2011;82:177-183.
71. Podos J. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*. 1997;51:537-551.
72. Schmieder DA, Kingston T, Hashim R, Siemers BM. Breaking the trade-off: rainforest bats maximize bandwidth and repetition rate of echolocation calls as they approach prey. *Biol Lett*. 2010;6:604-609.
73. Jarvis ED. Evolution of vocal learning and spoken language. *Science*. 2019;366:50-54.
74. Petkov CI, Jarvis ED. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front Evol Neurosci*. 2012;4:1-24.
75. Gaub S, Groszer M, Fisher SE, Ehret G. The structure of innate vocalizations in Foxp2-deficient mouse pups. *Genes Brain Behav*. 2010;9:390-401.
76. Kurt S, Fisher SE, Ehret G. Foxp2 mutations impair auditory-motor association learning. *PLoS ONE*. 2012;7:e33130.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: von Merten S, Pfeifle C, Künzel S, Hoier S, Tautz D. A humanized version of *Foxp2* affects ultrasonic vocalization in adult female and male mice. *Genes, Brain and Behavior*. 2021;20(7):e12764. https://doi.org/10.1111/gbb.12764