



Universiteit
Leiden
The Netherlands

Female blue tits sing frequently: a sex comparison of occurrence, context, and structure of song

Sierro, J.; Kort, S.R. de; Riebel, K.; Hartley, I.R.; Dingemanse Niels

Citation

Sierro, J., Kort, S. R. de, Riebel, K., & Hartley, I. R. (2022). Female blue tits sing frequently: a sex comparison of occurrence, context, and structure of song. *Behavioral Ecology*, 33(5), 912-925. doi:10.1093/beheco/arac044

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3563305>

Note: To cite this publication please use the final published version (if applicable).



Original Article

Female blue tits sing frequently: a sex comparison of occurrence, context, and structure of song

Javier Sierro,^{a,○} Selvino R. de Kort,^b Katharina Riebel,^{c,○} and Ian R. Hartley^a

^aLancaster Environment Centre, Lancaster University, LEC Building, Lancaster LA1 4YQ, UK, ^bEcology and Environment Research Centre, Department of Natural Sciences, Manchester Metropolitan University, E437 John Dalton Building, Manchester Campus, Manchester M15 6BH, UK, and ^cInstitute of Biology (IBL), Leiden University, Room number 7.4.17, Sylvius Building, Sylviusweg 72, Leiden 2333 BE, The Netherlands

Received 1 July 2021; revised 1 April 2022; editorial decision 5 April 2022; accepted 25 April 2022; Advance Access publication 20 June 2022

In species with mutual mate choice, we should expect adaptive signaling in both sexes. However, the role of female sexual signals is generally understudied. A case in point is female birdsong that has received considerably less attention than male song. This holds even for well-studied species such as the blue tit (*Cyanistes caeruleus*), an important model in evolutionary ecology. Although there have been anecdotal reports of female song from three populations, there are no quantitative studies on female song in this species. Here, we report systematic sampling from a population of individually marked blue tits over 3 years, revealing that females sang frequently throughout the sampling period. Notably, daytime singing of females occurred in functionally similar contexts as in males (agonistic, solo song, and alarm contexts) but females had lower song output than males and were not observed singing dawn song, while males showed long singing displays at dawn before copulations take place. Female and male song overlapped substantially in acoustic structure (i.e., same song types or peak frequency) but females had smaller individual song-type repertoires, shorter trills, and lower vocal consistency. Differential selection pressures related to functional differences in male and female song might explain the observed variation in acoustic structure. With the first quantitative study of female song in such a well-studied species, we hope to stimulate further investigations into the functions of female singing, especially in the Northern temperate zones where female song may have been overlooked, not only in this but perhaps in other monomorphic species.

Key words: animal communication, bioacoustics, birdsong, *Cyanistes caeruleus*, female song, sexual characters.

INTRODUCTION

Birdsong plays an important role in the acquisition of breeding resources, in mediating social conflicts and in mate attraction (Marler and Slabbekoorn 2004; Catchpole and Slater 2008) but also in pair coordination and in alarm situations (Cresswell 1994; Halkin 1997). Birdsong is therefore under sexual selection (but see Tobias et al. 2012) and has long been considered a predominantly male trait (Searcy and Andersson 1986; Collins 2004), a sex bias not unique to birdsong (Clutton-Brock 2009). Consequently, the function of birdsong has been studied mostly in males (Langmore et al. 1996; Riebel et al. 2005; Odom et al. 2014; Austin et al. 2021) despite reports of female song from the early days of modern birdsong

research (Nice 1943; Robinson 1949; Nicolai 1959; Hoelzel 1986; Ritchison 1986).

It was not until this century that the first systematic worldwide survey was conducted, showing that female song is common in the basal clades of passerines, making concurrent male and female song the most likely ancestral state (Odom et al. 2014). With this shifting view, new questions arise regarding the function of female song and the selection pressures underlying sexual differences (Riebel et al. 2019). Even though the study of female song represents a very small fraction of the bird song literature (Odom and Benedict 2018), there is growing evidence, from an increasing number of species, that female song serves a variety of functions such as territory advertisement (Cooney and Cockburn 1995; Cain et al. 2015), mate attraction (Langmore et al. 1996), mate guarding (Reichard et al. 2018), or resource defense (Tobias and Seddon

Address correspondence to J. Sierro. E-mail: sierro.2.8@gmail.com.

2009) during inter- or intrasexual interactions (Krieg and Getty 2016; Kirschel et al. 2020). In non-duetting species, one of the most common functions of female song reported to date is related to the competition for breeding resources (and mates) between females (Langmore 1998; Austin et al. 2021). However, the documentation of female song is still too limited to make general statements, and more systematic research is needed to gain a complete picture of shared versus sex-specific functions of song in passerines (Riebel et al. 2019; Austin et al. 2021).

One unresolved question is why female song is much more common in the tropics and subtropics while it seems rare in the Passerida of the temperate zones (Slater and Mann 2004; Price et al. 2009; Odom et al. 2014). A current working hypothesis is that the presence of female song is the ancestral state in passerines, but that short breeding seasons, seasonal territoriality, and migration might be associated with the loss of female song (Benedict 2008; Price 2009; Odom et al. 2014). However, the evolutionary importance of female song seems to be underestimated, perhaps due to sampling biases (Garamszegi et al. 2007), for example by sexing a singing bird as male, especially in unmarked populations of monomorphic species (Eens and Pinxten 1998; Odom and Benedict 2018). To date, there have only been few systematic studies in Northern temperate regions that quantify female song and its functions (discussed in detail in Garamszegi et al. 2007; Riebel et al. 2019, but see for some notable exceptions Beletsky 1982; Baptista and Petrinovich 1986; Johnson and Kermott 1990; Hausberger and Black 1991; Baptista et al. 1993; Langmore et al. 1996; Yamaguchi 1998; Magoolagan and Sharp 2018; Wilkins et al. 2020; and Patchett et al. 2021; for descriptive studies of female song in Northern temperate regions). Importantly, by showing functional female song, these studies highlight the need to increase documentation and quantitative analyses of female song.

The blue tit, a well-studied passerine that breeds in the temperate regions of Europe and western Asia, might be a case in point. This songbird is a model species for studies of birdsong, mating systems, and other aspects of behavioral ecology (Mainwaring and Hartley 2019; Griffith et al. 2021). Even though blue tits show plumage dichromatism in the UV spectra (Andersson et al. 1998), males and females show only minor color and size differences to human observers, with much overlap in color intensity and size between the sexes (Cramp and Perrins 1993; Scott 1993). Female song in blue tits has been only reported anecdotally, but independently, in three different populations across Europe (Hinde 1952; Bijmens and Dhondt 1984; Mahr et al. 2016). Most surprisingly, there are no detailed, quantitative descriptions of female singing behavior, song structure, or context, despite the extensive literature of song research in this species (Stadler 1951; Hinde 1952; Latimer 1977; Doutrelant et al. 1998, 1999, 2000a, 2000b; Poesel and Kempenaers 2000; Poesel et al. 2001, 2004; Gorissen et al. 2002; Gorissen and Eens 2005; Poesel and Dabelsteen 2005).

In line with previous anecdotal reports (Hinde 1952; Bijmens and Dhondt 1984; Mahr et al. 2016), we encountered females singing during the collection of male song recordings and turned this into a systematic recording effort of both sexes in a population of individually marked birds. Based on these recordings, we here present the first quantitative analysis of the context, occurrence, and acoustic structure of female song in blue tits and compare it with the song of their male partners. Although there are some descriptive studies of female song in other bird species, these are predominantly from the tropics and subtropics. For temperate zone species, especially in Europe, studies involving systematic sampling of male and female

singing within the same population and sampling scheme are uncommon, despite being a crucial step to develop testable hypothesis regarding female song functions (Riebel et al. 2019; Austin et al. 2021). Blue tits are of special interest in this context as their breeding biology is well studied in multiple long-term studies across Europe (Mainwaring and Hartley 2019). Blue tits are territorial and generally breed in monogamous pairs but are occasionally socially polygynous and frequently genetically polyandrous (Leech et al. 2001; Schlicht and Kempenaers 2021). Next to analyzing song structure, we will present analyses of the context of female singing, to test for associations with contexts for which female song has been reported, that is, solo singing, female–female competition, or alarm contexts (Langmore 1998; Mahr et al. 2016; Austin et al. 2021). Finally, we will also discuss the sexual similarities and differences in song structure in relation to possible functional differences based on the context in which we find females, and males, singing.

METHODS

Study species

All birds included in this study were part of a long-term monitored population (Mainwaring and Hartley 2009) breeding in 110 nest boxes (67 ± 1.15 broods per year in three breeding seasons), placed in deciduous and mixed woodland at Lancaster University campus, UK (54.01° N, 2.78° W). Nest boxes were made from 25-mm-thick softwood (hole 25 mm diameter, internal dimensions $125 \times 125 \times 200$ mm), placed on tree trunks between 1 and 3 m above the ground and separated on average 25 m from each other (Supplementary Figure S1). Each year, adults were captured at the nest box using mist nets or trap-door traps at the entrance of their nest box, as well as at feeding stations with mist nets during the winter. Feeding stations were removed well before the start of the breeding season, to avoid impact of locally increased resources. Once caught, birds were measured (straightened, flattened, wing length to nearest mm, tarsus length with foot bent down, to nearest 0.1 mm, and head-bill length to nearest 0.1 mm), weighed (to nearest 0.1 g), and ringed with a unique combination of three colored rings and one numbered metal ring (Redfern and Clark 2001). Individuals caught during the breeding season are sexed in the hand based on the presence of a brood patch (females) or cloacal protuberance (males) (Svensson 1992). Birds were aged as first year or older than first year, based on plumage characteristics as described in Svensson (1992). All individuals included in this study were caught at least once during the breeding period.

Sampling scheme and analysis of singing incidence

From the end of January to May 2018–2020, we conducted several walked transects per week to collect song recordings using a Marantz PMD661 recorder (48 kHz, 24-bit) and a Sennheiser ME67 microphone. The study was initially designed to investigate variation in song performance of males, but we regularly encountered singing females and concurrently established a data base of female song. We thus recorded males and females from the same population and season as recommended by Riebel et al. (2019). The linear transects followed the lines of nest boxes placed along strips of woodland (Supplementary Figure S1). When walking along a transect, we waited a maximum of 5 min at each territory (if no bird was detected) or until we had identified all blue tits detected, using binoculars to read the color combination of their leg rings.

Only a fraction of the population was sampled each day (see last paragraph of this section); therefore, the starting points of daily transects would commence in the end point of the previous day's transect, to cover the entire population equally. During all transects, whenever a blue tit was encountered singing, its behavior and song were recorded simultaneously. The observation ended whenever the focal bird stopped singing for more than 2 min or was out of sight. Once the ring combinations were identified for all birds detected, and after the song observation was recorded for singing birds, the observer continued to the next nest box along the transect. Sampling was predominantly blind with respect to the sex of the bird, as sexes look very similar at a distance in the field (Scott 1993). Thus, sex was determined by cross-checking the ring details with the database after the recording event. We used a dictaphone to take voice notes in a long, continuous recording that lasted the entire day of fieldwork, describing the entire singing observations as well as the ring details of the birds identified. The dictaphone also recorded the ambient sounds, including songs of blue tits that were nearby, and this was useful to record the first songs of a song bout if it was unpredictable, just before we started the high-quality audio recorder (for the operational definition of a song bout and song, see the "Analysis of song-type repertoire of female and male song" section). During the pre-breeding period, from January to the beginning of April, we made song recordings during daytime after sunrise throughout the morning until midday, since singing activity was limited at dawn during this period (Hinde 1952). In April and May, during the egg-laying (fertile) and incubation periods (mean first egg date was the 22nd April over the 3 years), dawn singing behavior was predictable (Hinde 1952), and thus we began sampling 1 h before sunrise until 1 h after sunrise. During the dawn chorus, when birds produce long singing displays starting 90–30 min before sunrise (Poesel et al. 2001; McGregor 2005; Gil and Llusia 2020), we did not conduct transects but waited in a position that was near a few neighboring territories to make song recordings from the local residential individuals each morning. This was because visibility was poor at dawn and, to confirm the identity of a singing bird, we kept following and recording the same bird until light levels were good enough for identification from the leg rings. The change in sampling scheme midway through the season was intended to optimize the collection of high-quality song recordings, based on previously reported seasonal changes in the timing of song in this species (Hinde 1952). Hence, our sampling scheme differed between the pre-breeding and breeding period, and it is thus less optimal to quantify seasonal changes in absolute song output. However, the recording of any singing individual encountered along the transects and around the nest boxes means that males and females were sampled the same way at every time point in the season. The data are thus suitable to compare male and female singing as both sexes were recorded in the same context, same population and with the same sampling schemes. From the data collected over 3 years, we created four data sets from the same group of females: one to compare song output, one to compare behavioral context of song, one to compare song-type repertoire, and one to compare acoustic structure in song between females and males. The same individuals could be part of one or all data sets. For a summary of the entire data base, see [Supplementary Table S1](#).

Analysis of song output

To provide a first description of female singing activity in blue tits, we calculated the probability of females singing as the proportion of females recorded singing per box visited, including all

observations made during daytime before the breeding season (all observations before April), this is the song output data set. We then compared the probabilities of female and male singing per week using a Wilcoxon signed-rank tests, including a total of 29 weeks over 3 years, averaging 3.5 ± 1.8 days of sampling per week, 35.6 ± 22.9 boxes sampled per day, identifying a mean of 6.4 ± 5.6 females per day and 11.8 ± 6.9 males per day. Furthermore, we calculated the number of females and males recorded singing per minute in the field as well as measured the song output per observation, as the number of songs recorded per observation in 216 independent observations (see the sections "Analysis of behavioral context of female and male song" and "Analysis of song-type repertoire of female and male song" for operational definitions of observation and song, respectively). We compared the number of songs per observation between males and females with a Mann-Whitney *U* test, as data were not paired in this case. All measures are presented as mean \pm 1 standard deviation, unless otherwise indicated.

Analysis of behavioral context of female and male song

A posteriori, based on our field notes (dictaphone recordings), observations were categorized into one of four distinct behavioral contexts: alarming behavior, agonistic interactions, solo song, and dawn song (see [Table 1](#) for operational definitions). For all except 10 cases, observations were classified into one of the four behavioral contexts. Occasionally, the behavioral context changed during a continuous focal observation. If this occurred, the observation was split into two separate, shorter, observations each assigned to the respective context (i.e., when switching from solo song to an alarm context). In some cases, two individuals could be recorded simultaneously (i.e., two breeding partners performing a joint alarm display) but only one observation was selected (randomly) in such cases, independently of sex. In some cases, alarm behavior may have been directed towards the human observer, suggesting these observations of male or female might have been observer induced, as has been reported by other authors (Bijens and Dhondt 1984; Hinde 1952). However, assessing which alarm displays were or were not triggered by the observer is difficult based on our sampling scheme. In any case, as we observed both sexes engaging in alarm behavior and both sexes were exposed to a human observer during recordings, this context still allows for a comparison between sexes. To describe singing context in females, we measured the proportion of singing observations for each context, relative to the total number of observations within individual, including all observations along the 3 years of study. Then, for a sex comparison of singing context, we selected a subset of the recorded males in our population that were the breeding partner of, at least, one of the selected females during, at least, one breeding season. For this group of males, we used all observations recorded during the 3 years and, in each case, context was categorized with the same definitions as in females, using the voice notes recorded in the field during the transects. Breeding partners were defined as any two individuals that were observed provisioning the same brood, that is, nestlings in the same nest box. For this, we observed boxes directly to identify provisioning individuals from their leg rings, but also made video recordings of the entrance hole for at least 1 h, when the nestlings were 10–12 days old, with a Sony or Canon HD camcorder placed on a tripod 3–5 m away from the nest box.

Table 1
Operational definitions of behavioral and acoustic terms, by alphabetical order

Term	Definition
Agonistic interaction	Defined as any interaction where two or more blue tits produced conflict calls (Bijnens and Dhondt 1984) or actively chased, displaced, or attacked each other (Irschick et al. 2007) while singing. Observations were classified based on the highest level of aggression observed.
Alarm song context	Characterized by the production of song while 1) displaying mobbing behavior such as approaching closely and circling a potential predator (Cramp and Perrins 1993) or 2) producing song intercalated, or very close in time with high-pitched “tsee” calls or scolding calls. Both call types are typically produced in the presence of a predator (Hinde 1952; Latimer 1977; Bijnens and Dhondt 1984; Mahr et al. 2016)
Dawn song	A long vocal display that starts 90–30 min before sunrise (Poesel et al. 2001; McGregor 2005; Gil and Llusia 2020)
Bandwidth Maximum and minimum frequency	Absolute difference between the maximum and minimum frequency of the note Following Podos (1997), we measured the maximum and minimum frequency of each note in the mean power spectrum (window size: 1024 samples; amplitude threshold; –20 dB)
Peak frequency	The frequency with the maximum amplitude in the mean power spectrum (window size = 1024, window type = “Hanning”)
Solo songs	Characterized by an individual bird singing alone without any active social interactions with conspecific nor alarm responses
Song repertoire	Total number of different song types found across all recordings of a single individual during the entire study
Trill length Trill rate	Duration of the trill in seconds, from the start of the first element until the end of last element. Following Kirschel et al. (2009), trill rate was defined as the number of notes per second excluding the last rendition divided by the time between the start of the first note to the start of the last note in the trill (see “Male 1” in Figure 1). This is because all the notes in the trill include their corresponding internote gap except the last. Including all notes would cause a bias in trill rate due to the number of renditions in the trill.
Vocal consistency	Acoustic similarity between each note to the consecutive rendition within the trill using a spectrogram cross-correlation (SPCC) algorithm (Clark et al. 1987; Coleman et al. 2007; Botero et al. 2009). We calculated the maximum correlation of every pairwise SPCC with a maximum temporal offset of 20 ms and a temporal resolution of 1 ms. The spectrogram matrices were computed using a fast Fourier transform algorithm with a window size of 512 samples and 90% overlap between successive windows, “Hanning” window type
Vocal deviation	Using song of both sexes, we first calculated the upper bound regression between trill bandwidth and trill rate using bins of 100 Hz (Blackburn et al. 1992; Podos 1997). The vocal deviation is the orthogonal distance of each trill to the estimated upper bound limit (Supplementary Figure S2; see Podos 1997). The upper bound regression of vocal deviation including male and female song rendered a significantly negative slope (r (df = 35) = –0.90, $P < 0.001$; Supplementary Figure S2)

After excluding the observations made during the dawn chorus, as we had only observations for males in this context, the behavioral data set included 207 focal observations from 36 females and 29 males, after removing 9 observations with unknown behavioral context. For each individual, we estimated the proportion of observations per context in relation to the total number of observations for that individual, adding zeroes if an individual was never seen in a certain context. This led to a total of 195 data points to be used in the model, three for each of the 65 individuals. We fitted a binomial Generalized Linear Model or GLM (*glm* function from “stats” package; R Development Core Team 2016) using the proportion of observations for each individual per context as the response variable as a function of sex, behavioral context, and the full interaction between these two variables. Note such data structure implies only one data per individual per context (proportion), hence there are no repeated measurements and therefore no random terms were included. In this behavioral data set, we had different numbers of males and females because four male partners were never recorded, and six female–male pairs shared the same three males in same or different years. In many cases, we collected multiple observations of the same individuals on 2 or 3 different years (26/65) and in all except two cases, male–female partners were observed within the same season in at least 1 year. Only 3 out of 65 individuals were

recorded during a year when breeding data were not collected for those 3 individuals.

Analysis of song-type repertoire of female and male song

In the blue tit, as in many species, functional distinctions between songs and calls are not clear cut (Catchpole and Slater 2008). Often, blue tits integrate long bouts of calls in dawn song, a complex vocal display normally associated with song (Poesel et al. 2001, 2004). On the other hand, song is also used in contexts that are not strictly related with reproduction, such as alarm contexts (Mahr et al. 2016). For these reasons, we used structural criteria to define blue tit song based on spectrograms reported in previous studies (Bijnens and Dhondt 1984; Mahr et al. 2016). Song was defined as a vocalization composed of a few introductory, high-pitched notes followed by a trill, defined as the last part of the song where a note is repeated in succession (Figure 1) (Bijnens and Dhondt 1984; Cramp and Perrins 1993). Blue tit calls, which have a noisy rather than tonal acoustic structure (i.e., scolding or churring calls), were not analyzed. For the purpose of this study, we focused on song, including only those vocalizations that presented tonal structure to ensure that we were indeed describing female song rather than calls. This is a conservative approach since some blue tit song types include other types of sound (see B-syllables in Bijnens and Dhondt

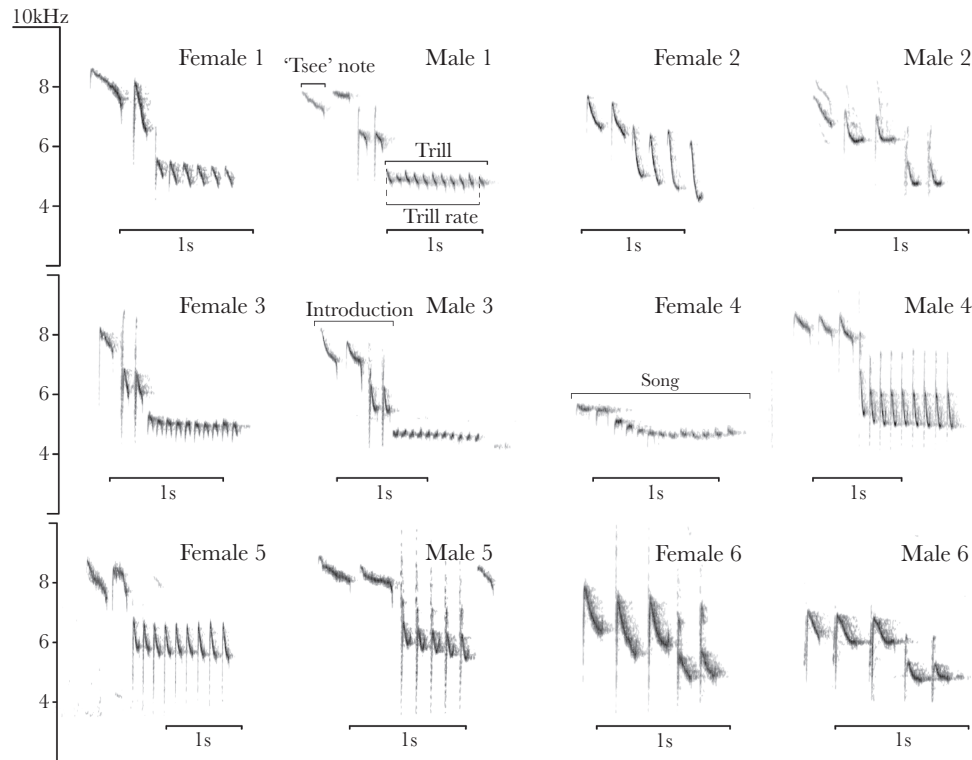


Figure 1

Spectrograms of female and male songs including most of the song types observed in females. For a complete repertoire of all females recorded in this study, please see [Supplementary Figure S3](#). For each female, we selected the same, or similar, song type from her male partner. The basic terminology and structure of blue tit song are indicated in the spectrograms. The songs were selected to provide a good visual representation of song structure, not necessarily an example of statistical differences.

1984). For our description of song and for the song analyses, a note was defined as a continuous trace in the spectrogram, separated from other notes by silent gaps longer than 10 ms. Each recorded individual sang several stereotyped song structures, referred to as song types, and these are comparable between individuals within a population ([Figure 1](#)). During singing, blue tits repeat the same song type many times, alternated with silent pauses (i.e., discontinuous singers), before switching to a different song type which results in so-called song type bouts ([Poesel et al. 2004](#)). Each of these renditions are referred to as a song ([Figure 1](#)) and one song was sufficient to count a bird as singing during one observation.

Blue tits have small individual song-type repertoires that range from three to eight different types ([Bijmens and Dhondt 1984](#)). Since blue tits repeat the same song many times before switching, repertoire size is often estimated on the basis of the long, sustained singing displayed during the dawn chorus ([Doutrelant et al. 2000a; Poesel et al. 2004](#)). We did not record cases of females singing dawn song, and female daytime singing, like that of males, consisted of much shorter singing bouts. Hence, to compare repertoire usage between sexes, we counted the number of distinct song types across all recording days of the same individual made during daytime singing. Blue tits are generally assumed to be close-ended learners ([Bijmens 1988](#); see also [Hansen et al. 2008](#); for imprinted species recognition of song) and based on this assumption, we grouped recordings across years for those individuals recorded in multiple seasons (see [Bijmens and Dhondt 1984](#) for similar methodology). Hence, we created a subset, hereafter the repertoire data set, selecting those females recorded on two or more dates with more than 10 songs recorded overall, including 19 females.

We categorized song types by visual inspection of spectrograms ([Bijmens and Dhondt 1984; Doutrelant et al. 1998](#)) in Audacity ([Mazzoni and Dannenberg 2014](#)) (window type: “Hanning,” window length 1024 samples, 90% overlap and -80 dB range). Based on the song delivery mode of blue tits, that repeat the same song type many times, we focused on the switching points between song types within individuals as the main criteria for identifying song types. The switching point between song types was easily identifiable even if song types were similar. Some of the main features of song that further helped in categorizing song types were trill rate, frequency modulation of trill notes, the trill length, and the structure of the introductory part ([Figure 1](#)). But note that, in all cases, the categorization was based on visual inspection of spectrograms. For each of the 19 females, we selected her male partner to compare song-type repertoires between sexes. In almost all cases, we collected a larger sample of song recordings for the male than for the female (see [Supplementary Table S1](#)). From 38 individuals, 21 were recorded in more than 1 year (10 females and 11 males) and, within pairs, male and female partners were always recorded within the same season in at least 1 year. For each recording, we transformed date into Julian date, taking the first of January of each year as origin resulting in a standardized date across seasons. Within each female–male pair, we used the dates of the female recordings to select the nearest dates of recording for male songs, regardless of the year of recording. By doing this, we reduced potential seasonal variation due to singing modes. After matching female and male recording dates, we selected the same number of songs from each male recording unless there were fewer male songs than female songs for that match. The count of songs started always in the first

Table 3
Output of the full-average model for each song trait, comparing male and female song in five LMMs with a Gaussian family distribution and an identity link function

Parameter	Fixed effects estimates					Random effects estimates			R^2_m	R^2_c
	Variable	Estimate	2.5% CI	97.5% CI	Z value	Relative importance	Variable	Variance		
Peak frequency (kHz) N = 761 songs, 60 individuals	Intercept (female, yearling)	5.02	4.838	5.203	54.01	—	Individual	0.15 (0.08–0.24)	0.12	0.692
	Age (yearling vs. older)	0.152	0.075	0.229	3.854	1	Pair	0.02 (0.00–0.10)		
	Weeks to first egg	-0.255	-0.375	-0.134	4.145	1	Residual	0.09 (0.08–0.10)		
	Sex (female vs. male)	0.025	-0.231	0.28	0.191	1				
	Age: sex (female–yearling vs. older vs. male–yearling vs. older)	-0.232	-0.325	-0.14	4.93	1				
Vocal consistency (spectrogram cross-correlation algorithm score) N = 745 songs, 60 individuals	Weeks to first egg: sex (female–season vs. male–season)	0.073	0.027	0.272	0.843	0.489			0.125	0.435
	Intercept (female, yearling)	0.867	0.84	0.895	60.811	—	Individual	0.002 (0.001–0.003)		
	Weeks to first egg	0.037	0.025	0.049	6.234	1	Pair	0.00 (0.00–0.001)		
	Sex (female vs. male)	0.024	0.016	0.07	1.01	0.556	Residual	0.004 (0.003–0.004)		
	Intercept	0.483	0.402	0.564	11.728	—	Individual	0.06 (0.03–0.08)		
Trill length (s) N = 786 songs, 60 individuals	Sex (female vs. male)	0.196	0.078	0.315	3.284	—	Pair	0.00 (0.00–0.02)	0.105	0.609
	Intercept (female, yearling)	8.582	7.998	9.166	28.82	—	Residual	0.04 (0.04–0.05)		
	Age (yearling vs. older)	0.200	-0.09	0.871	0.763	0.513	Individual	2.9 (1.6–4.7)		
	Sex (female vs. male)	-0.075	-1.089	0.721	0.243	0.405	Pair	0.48 (0.00–2.1)		
	Age: sex (female–yearling vs. older vs. male–yearling vs. older)	-0.127	-0.939	-0.054	0.519	0.256	Residual	1.2 (1.07–1.4)		
Trill rate (notes/s) N = 571 songs, 58 individuals (31 females, 27 males)	Intercept (female, yearling)	5.963	4.948	6.978	11.517	—	Individual	3.2 (0.80–6.5)	0.041	0.744
	Age (yearling vs. older)	0.321	-0.941	1.985	0.503	0.615	Pair	1.1 (0.0–4.2)		
	Sex (female vs. male)	-0.034	-1.179	1.007	0.096	0.39	Residual	9.9 (8.7–11.2)		
	Weeks to first egg	-0.162	-2.345	0.477	0.35	0.174				
	Weeks to first egg: sex (female–season vs. male–season)	0.06	-0.126	0.719	0.389	0.201				
Vocal deviation (orthogonal distance) N = 550 songs, 58 individuals (31 females, 27 males)	Intercept (female, yearling)	5.963	4.948	6.978	11.517	—	Individual	3.2 (0.80–6.5)	0.023	0.317
	Age (yearling vs. older)	0.321	-0.941	1.985	0.503	0.615	Pair	1.1 (0.0–4.2)		
	Sex (female vs. male)	-0.034	-1.179	1.007	0.096	0.39	Residual	9.9 (8.7–11.2)		
	Weeks to first egg	-0.162	-2.345	0.477	0.35	0.174				
	Weeks to first egg: sex (female–season vs. male–season)	0.06	-0.126	0.719	0.389	0.201				

Variance attributed to individual identity (ID) and pair identity (Pair ID), fitted as a random effects, as well as residual variance (Residual) is shown by the standard deviation with the associated 95% CI. For each fixed factor, we present the model estimate, the 95% CI around the estimate, the Z statistic derived from Wald tests, and the relative importance of that factor in the final model. The relative importance reflects the number of models, from the subset selected ($\Delta AIC_c < 2$) that included that specific variable. The last two columns show the marginal R squared represented as R^2_m , as the variance explained only the fixed terms, and the conditional R-squared represented as R^2_c , as the variance explained by both fixed and random terms. For each fixed effect, the specific levels to be compared are shown in parenthesis.

song recorded. In this case, we also excluded male dawn chorus recordings to avoid a contextual bias on the comparison of song repertoires between sexes. While this method means that we might not have included the complete repertoire of each individual, it allowed us to compare repertoire usage between the sexes. We used Wilcoxon signed-rank tests to compare the number of song types of females and males within pairs.

Analysis of acoustic structure

For detailed analysis of acoustic parameters of female song, we chose a subset of females with song recordings of high signal-to-noise ratio, hereafter referred to as the acoustic data set. We then selected individual males that were breeding partners of these females in at least one breeding season. Similar to the male sample in the analysis of song-type repertoires, we selected the same number of songs for each male partner, recorded on similar dates and excluding song recorded during the dawn chorus. This resulted in a data set of 786 high-quality recording songs, 435 songs from 32 females (14 ± 12 songs per individual) and 351 songs from 28 male partners (12 ± 13 songs per individual). For 2 of 32 females, the male partner was not recorded at all. Four female–male pairs shared the same two males in the same or different years. From 30 females that were matched with a male partner, 7 were recorded in different years than their mate. In 23 pairs, both partners were recorded, at least once in the season they were breeding together. Finally, 31 of 32 females were recorded singing during the same year when they were recorded breeding. From all 60 individuals, we had the exact date of first egg for 90 nests over three seasons to calculate the “weeks to first egg” of each recording. In eight cases across 3 years, we did not have breeding data hence, as an estimate of the weeks in relation of first egg of each recording, we used the mean week of first egg for the entire population in that year. Since all females laid their eggs within 3 weeks of the year (weeks 15–17), the possible error introduced should not impact our analysis.

We conducted all acoustic analysis using Audacity and R software (package “tuneR”: Ligges 2013; package “seewave”: R Development Core Team 2016; Sueur et al. 2006). For each individual, we analyzed a maximum of 10 songs for each song type and each date recorded. Acoustic measurements were made only in the trill since the introductory notes of blue tit song are more variable and may be absent in some song types. Despite having selected high-quality recordings, we still had to exclude some notes that were masked by extraneous sounds but included the rest of the song in the analysis ($7.04 \pm 5.3\%$ of notes in females and $6.4 \pm 5.5\%$ of notes in males). We did this to avoid biasing the sample towards shorter songs, as longer songs were more likely to be partly masked.

Every note was manually labeled in the spectrogram, using the cursor to mark the start and end times in Audacity, and these time marks were exported as a text file. With this file, we used R software to cut out each note from the recording and save it as a single, normalized *wav* file. Following Podos (1997), we measured the peak, maximum, and minimum frequency of each note and from this we derived the bandwidth of the note (Table 1). We also took measurements of song performance including vocal consistency (sensu de Kort et al. 2009), trill length, trill rate (sensu Kirschel et al. 2009), and vocal deviation (sensu Podos 1997). Table 1 shows detailed operational definitions of each acoustic variable. All acoustic variables were measured in each note and then, a mean value per song was calculated for the statistical analysis.

For the acoustic analysis, we built five Linear Mixed-Effects Models (*lmer* function from “lme4” package; Bates 2010) to investigate sex differences for each of the following five parameters: peak frequency, vocal consistency, trill length, trill rate, and vocal deviation (Table 3). Note that sample sizes varied slightly between models. For instance, measuring trill rate unbiased (sensu Kirschel et al. 2009) is only possible in trills of at least three notes, which also affects the measurement of vocal deviation. In some trills, recording quality was insufficient to take spectral measurements but trill rate or trill length were easily measured. In other cases, signal-to-noise ratio was sufficiently high to measure peak frequency, but vocal consistency was not measured due to extraneous sounds in the background outside the spectral range of the note. For each model, the exact sample size is specified in Table 3. To describe the spectral features of song, we selected only the mean peak frequency of each song, because it is a robust measurement that is little affected by recording quality (Linhart et al. 2012) and it was strongly correlated with the maximum frequency ($r = 0.90$, $P < 0.001$, degrees of freedom [df] = 759) and the minimum frequency ($r = 0.89$, $P < 0.001$, df = 759). As explanatory variables we used sex (male and female), age (first year or older than first year), and weeks in relation to first egg date to account for seasonal variation (week of first egg = 0) (sensu Schlicht and Kempnaers 2020). We included sex-specific interactions with age and season effects (in weeks in relation to first egg) to investigate their potential effects on each sex specifically. To model peak frequency, we also included the tarsus length and its interaction with sex, since this song feature could be affected by sexual dimorphism in body size. To account for repeated measurements, observations were nested within individual and within pair, using both variables as random effects.

Statistical analysis

Statistical analyses were carried out in R software (R Development Core Team 2016). To validate all models, we confirmed that the residuals were homoscedastic and showed a normal distribution using diagnostic plots (Zuur et al. 2009; Knief and Forstmeier 2021). We also tested for potential multicollinearity among the explanatory variables of the model by visual inspection of paired correlation plots and by estimating the Variance Inflation Factor (VIF) (*vif* function from “car” package; Fox and Weisberg 2019) for each variable with and without the interactions. Multicollinearity among explanatory variables was assumed if VIF was greater than 3, excluding those variables that were the product (interaction) of simple variables (Zuur et al. 2009). To find which factors were important in explaining variation in song, we used an information theoretic approach, computing all possible model combinations and ranking them using the Akaike Information Criterion for small samples (AICc, *dredge* function in the package “MuMIn”; Barton 2011). This procedure compares the fit of all possible models while penalizing model complexity, in terms of the number of explanatory variables included. We selected all models that had $\Delta\text{AICc} < 2$, in relation to the model with the lowest AICc score (best model), to compute the full-average model as the final model (Burnham and Anderson 2002; Burnham et al. 2011). We used the relative importance of each factor in the final model together with the coefficients and estimated confidence intervals (CIs) with a threshold of 95% (Nakagawa and Cuthill 2007; Burnham et al. 2011), concluding there was a significant effect if the CI did not overlap with zero. In the process of model selection, some predictors were dropped from the final model, indicating that their impact on the response variable

Table 2
Behavioral context of song compared between sexes by fitting a GLM binomial model using a log link function

Fixed effect	Log-transformed model estimates			Back-transformed model estimates			Z value
	Estimate	2.5% CI	97.5% CI	Estimate	2.5% CI	97.5% CI	
Intercept (female, solo song)	-0.351	-0.773	0.060	0.704	0.462	1.062	-1.66
Alarm (female–alarm vs. solo)	0.000	-0.588	0.588	1.000	0.555	1.80	0.00
Agonistic (female–agonistic vs. solo)	-1.207	-1.908	-0.541	0.299	0.148	0.582	-3.476
Sex (female–solo vs. male–solo)	0.867	0.310	1.434	2.38	1.364	4.197	3.028
Sex: alarm (female–alarm vs. solo vs. male–alarm vs. solo)	-2.135	-3.001	-1.29	0.118	0.05	0.275	-4.901
Sex: agonistic (female–agonistic vs. solo vs. male–agonistic vs. solo)	-0.642	-1.538	0.264	0.526	0.215	1.302	-1.399

For each fixed factor, we present the model estimate, the 95% CI around the estimate, and the T statistic, both the log-transformed as well as the back-transformed estimates in the original scale. The R^2 for the full model was 0.327. After model selection, only the full model was selected, as the second-best model had a ΔAICc greater than 2. Therefore, the relative importance is not calculated. The model takes the female category of sex variable as well as the solo song category from the context variable as reference levels (included in the intercept). For each fixed effect, the specific levels to be compared are shown in parenthesis.

was low. Finally, we calculated the R^2 or R^2_{GLMM} ($r.squaredGLMM$ from “MuMIn” package) of the full models to measure the goodness of fit (Nakagawa and Schielzeth 2013). All numerical variables are scaled and centered so model estimates are standardized (Gelman 2008). Other packages used in data management and data visualization during the analysis were “stringr” (Wickham and Wickham 2019), “plyr” (Wickham and Wickham 2020), “data.table” (Dowle and Srinivasan 2021), “ggplot2” (Wickham 2016), “geosphere” (Hijmans 2019), and “lattice” (Deepayan 2008).

RESULTS

In 3 years, we recorded 101 singing observations from 36 different females (Supplementary Table S1). Each year, we identified 49% \pm 18% of breeding females during field sampling hours (39/68 in 2018, 42/69 in 2019, and 10/68 in 2020). Of those identified females, we recorded 47% \pm 14% females singing (14/39 in 2018, 18/42 in 2019, and 12/19 in 2020). Of all 101 observations, 70 observations on 28 individual females were collected from January to April, during daytime transects, the remaining 31 observations (16 individuals) were collected in April and May, when sampling was mostly focused on dawn chorus (until 1 h after sunrise), although females were not observed producing dawn song.

Song output

During the daytime transects from January to April, we recorded a female singing every 110 min in the field, at 8.2% \pm 6.6% of the boxes visited. For the same period and sampling scheme, we recorded a male singing every 33 min of fieldwork, at 28.2 \pm 21.5% of the boxes visited, which is significantly higher than females ($W = 0$, $P < 0.001$, 5% CI = -0.23, 95% CI = -0.12). Per observation, we found that females’ song output (10.2 \pm 10.6 songs per observation) was significantly lower than males’ song output (24.3 \pm 23.5 songs per observation, $W = 3642.5$, $P < 0.001$, 5% CI = -11.0, 95% CI = -5.00). Note that observations after 1st of April were not included in these analyses, as the sampling scheme changed.

Behavioral context of song

For 36 females and 29 males, we conducted the behavioral analysis of song context. In this model, there was no collinearity among the explanatory variables. We found that females sang significantly less often during agonistic than during solo song, while no differences were found between solo song and alarm context (Table 2, Figure 2a).

Males sang proportionally more often during solo song than females (see estimate for “sex” in Table 2, Figure 2a). Females sang more often in the alarm context, relative to solo song, compared to males (see estimate for interactions “sex: alarm” in Table 2, Figure 2a). There was no difference in the use of song during agonistic interactions, compared to solo song, between the sexes (see estimate for interactions “sex: agonistic” in Table 2, Figure 2a).

From 42 agonistic observations with song, 15 were from females. From these, in three cases, a female approached and sang while her partner interacted agonistically with another individual (twice with birds of unknown sex and once with another male). Females showed agonistic behavior (conflict calls, chase, displacement, or attack) in the other 12 female agonistic observations, 3 during a contest between two male–female pairs, 5 during a contest with another individual of unknown sex, and 4 during a contest with another known female. In these four cases of female–female interactions, we observed two cases of direct physical aggression (attack). In two cases of female active agonistic interactions, their male partner approached and sang but did not contribute physically to the interaction. In 2 of 27 male agonistic observations selected for this study, we observed direct physical aggression to an opponent of unknown sex.

Song-type repertoire

Visual inspections of the spectrograms of all songs recorded in the study (1113 female songs and 1087 male songs) revealed that males and females used the same song type categories (printed spectrograms of the full individual song-type repertoire for 36 females can be found in Supplementary Figure S3). Considering the relative dates within season, male and female partners were recorded a mean of 3.0 \pm 21.4 days apart and the difference in the total number of songs recorded within pairs was 1.4 \pm 13.4, from a total of 2200 songs analyzed for both sexes combined. Comparing 19 females (4.1 \pm 1.8 dates and 59 \pm 45 songs) directly with their 19 male partners (3.6 \pm 1.8 dates and 57 \pm 42 songs), we found that females used significantly fewer song types than males (females = 1.0 \pm 1.0 vs. males = 3.0 \pm 1.0 song types, median \pm interquartile range; $N = 19$ pairs, $W = 57.5$, $P < 0.001$, 5% CI = -1.00, 95% CI = -2.00; Figure 2b). For 13 of the 19 females, we only recorded one song type, and this included three of the most recorded females (recorded on at least four different dates and 93 \pm 46 songs sampled; see Supplementary Table S1). In contrast, none of the males in our population sang fewer than two song types.

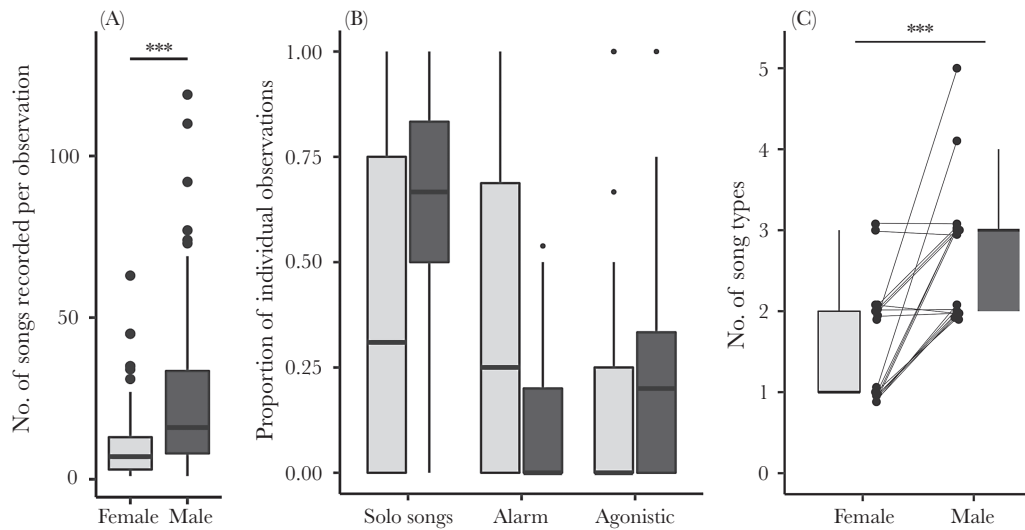


Figure 2

Sex differences in song output (a), singing context (b), and in song-type repertoires (c). Panel (c) shows the comparative analysis of song-type repertoire between sexes using a box and whiskers plot at each side while the points (raw data) are shown in the center linking male–female pairs with lines. Significance code: *** $P < 0.001$ (Mann–Whitney U test in (a) and Wilcoxon signed-rank test in (c)). Box and whiskers plot show median, the 25% and 75% quartiles as lower and upper hinges, respectively, and the whiskers stretch to the maximum (and minimum) value within 1.5 times the interquartile range. Points outside this range are shown individually.

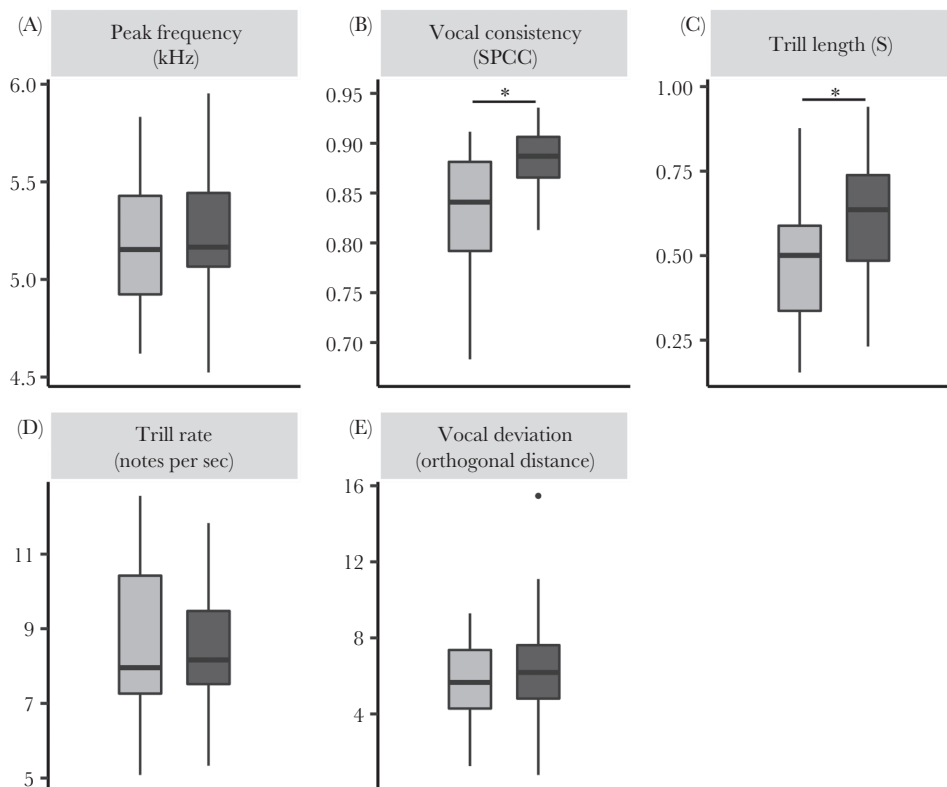


Figure 3

Comparison of female and male acoustic variables (females in gray and males in black) using box and whiskers plots, built as in Figure 2. Females sing with lower vocal consistency (b) and shorter trill length (c) than males. Peak frequency (a), trill rate (d), and vocal deviation (e) were not statistically different between sexes. The asterisk symbol means that the CI for the sex estimate in those models does not overlap with zero.

Acoustic structure of song

We found no collinearity among the explanatory variables in these models. Female blue tits sang with significantly lower vocal

consistency and produced shorter trills than males, but they did not differ in peak frequency, trill rate, or vocal deviation (Figure 3, Table 3). In both males and females, vocal consistency increased,

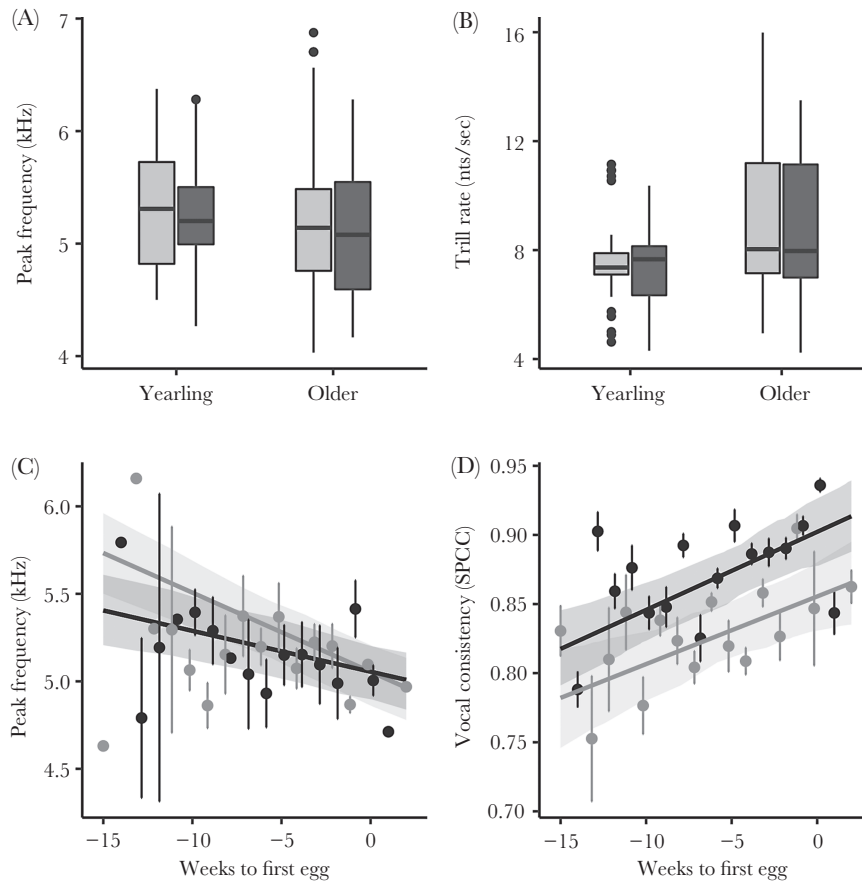


Figure 4

Significant correlations of age with peak frequency (a) and with trill rate (b) as well as season with peak frequency (c) and vocal consistency (d). Points represent raw data and lines represent predicted values from the model (Table 3), with the associated CI in the shaded area around lines. Females are represented in grey and males in black.

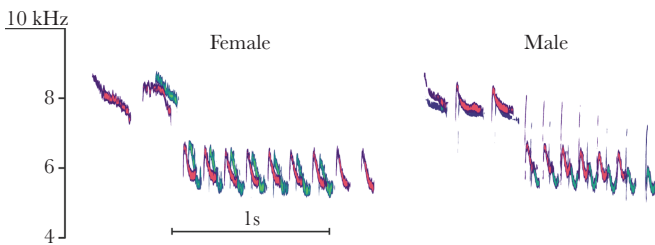


Figure 5

Spectrograms showing examples of spectral flexibility along the season within individuals. On the left, the same song type recorded from the same individual female, 6 weeks before laying the first egg (winter, in red) and 2 weeks before the first egg (spring, in green). Similarly, a male example is shown from 10 weeks before first egg date (winter, in red) and during the week of first egg (spring, in green).

and peak frequency decreased, from winter towards spring (fertile period), although males' seasonal decrease in peak frequency was significantly lower (Figures 4 and 5, Supplementary Figure S4, Table 3). In both males and females, older birds showed significantly lower peak frequency than first-year birds, although such change was significantly smaller in females than in males (Figure 4, Table 3). In the final model, we found a significant increase in trill rate with age in males, but not in females (Table 3), even

though the increase was significant in both sexes in the full model (Supplementary Table S2). Contrary to expectation, we did not find a significant correlation between peak frequency and tarsus length. As all models included the full interaction of sex with the other predictors, all the effects of age, season, and tarsus length were estimated specifically for each sex within the same model, although model estimates for such interactions must be taken with care given our sample size. The estimates and the associated CIs for all parameters in the full models are given in Supplementary Table S2. Supplementary Tables S3 and S4 show model selection tables, including all model combinations with an $\Delta AICc$ lower than 7 with respect to the best model for all song traits, although only the models with $\Delta AICc < 2$ were selected to build the final model.

DISCUSSION

This first season-long, transect-based quantification of female song in blue tits showed that female blue tits sing regularly throughout the season across a variety of behavioral contexts. These include alarming behavior, solo singing, and agonistic contexts, although females show lower song output in both rate of singing and number of songs and were not observed to produce dawn song. The structural analysis of song showed that females used similar song types as males, overlapping substantially in song parameters, but also showed quantitative differences. Females had smaller song-type

repertoires and sang shorter trills of lower vocal consistency. Our results also suggest that female song traits have communicative value as we found that female song 1) changes seasonally, as peak frequency decreased and vocal consistency increased towards the fertile period, and 2) correlates with age class, as older females had lower peak frequency. The same relationships were also found in male song.

Throughout the season, female song was frequent and far from a rare phenomenon in our study population. As birdsong is a complex behavior that involves specialized neural centers, muscle structures, and developmental learning processes (Suthers 2004), it seems unlikely that such behavioral and physiological adaptations are specific to our population. Nevertheless, large-scale social and environmental factors have been shown to influence female song across species (Benedict 2008; Price et al. 2009) and it is possible there are also other environmental factors that may modulate singing rates within species, across populations. For instance, higher habitat fragmentation due to an urbanized environment in our population may increase female–female competition for breeding opportunities (Bain et al. 2014) which may in turn raise singing rates in females (Kluyver 1955; Langmore and Davies 1997). This hypothesis could be tested in the future by investigating and comparing additional populations. Regardless of potential population differences in female singing rates, our observations put previous anecdotal reports of female song from three distinct populations into a new perspective (Hinde 1952; Bijnens and Dhondt 1984; Mahr et al. 2016; see also Gorissen and Eens 2005; for a report of female “subsong” inside the nest). This combined evidence suggests that female song is part of blue tit vocal signaling, a proposition worth investigating across other populations in Europe.

The blue tit can thus be added to a growing list of monomorphic European bird species with prolific female singing (for a review of European species with female song, see Garamszegi et al. 2007), despite the overall lower incidence of female song of Passerida in the Northern temperate zones compared with other biogeographic regions (Garamszegi et al. 2007; Odom et al. 2014; Webb et al. 2016). Moreover, the regular incidence of female song in the blue tit is in line with the results of worldwide surveys showing a higher incidence of female birdsong in non-migratory species (but see Nilsson et al. 2008) with low sexual dichromatism (Webb et al. 2016) that hold year-round territories (Benedict 2008; Price 2009; Logue and Hall 2014; Odom et al. 2014; Riebel et al. 2019). As song is often used to sex singing individuals in unmarked populations, and female song appears more common in monomorphic species, the incidence of female song might be underestimated (Riebel et al. 2005; Odom et al. 2014; Odom and Benedict 2018). It seems possible that there are other European species, like the blue tit, where female song might have been misclassified as incidental or absent.

Another reason that could explain why female song has been overlooked in blue tits is the fact that most studies collect (male) song recordings during the dawn chorus (Doutrelant et al. 1998, 2000a; Poesel and Kempenaers 2000; Poesel et al. 2001, 2004; Gorissen et al. 2002). In our population, we did not observe females producing such long, sustained singing at dawn and, if this pattern holds for other populations, it could explain why female song was so long overlooked in this intensively studied songbird species. In line with this, it is worth noting that the few studies that anecdotally reported female blue tit song in their populations recorded song also during daytime singing throughout the season, not only in the dawn chorus (Hinde 1952; Bijnens and Dhondt 1984; Mahr et al. 2016)

The regular occurrence of female song and its structural similarity to male song raises questions about its potential function(s). Although our study was purely observational, the contexts, in which we observed female song, allow to develop some working hypotheses regarding its function(s). One of the common contexts in which female blue tits sang was alarming behavior. This supports the findings by Mahr et al. (2016) that showed that blue tits sing upon presentation of simulated predator (a taxidermy mount of a sparrowhawk, *Accipiter nisus*). Blue tits are thus among a growing number of species known to produce song in the presence of predators (Langmore and Mulder 1992; Cresswell 1994; Laiolo et al. 2004), suggesting song as a potential antipredator strategy. The other two frequent contexts of female song were agonistic interactions and solo singing, which could indicate a function for female song in territory defense or mate attraction (Langmore 1998; Mikula et al. 2020; Austin et al. 2021). A territorial function of female song has been demonstrated in several species using playback experiments (Hoelzel 1986; Cooney and Cockburn 1995; Krieg and Getty 2016; Magoalagan and Sharp 2018). The observation that both sexes use the same song types could facilitate intersexual interactions in such agonistic context, since song type matching is an aggressive signal in this and related species (Krebs et al. 1981; Langemann et al. 2000; Poesel and Dabelsteen 2005). A shared repertoire of song types also suggests that both sexes learn from the same models in the same locations (Riebel 2003).

More specifically, female song could serve a function during intrasexual competition, as we observed females singing during aggressive interactions with other females. While the sex of the opponent was unknown in some agonistic interactions, we never saw a confirmed agonistic interaction between a single female and a male. A similar role of female song during intrasexual conflicts has been suggested in other species. In dunnocks (*Prunella modularis*) and great reed warblers (*Acrocephalus arundinaceus*), artificially high female–female competition increased the incidence of female song during intrasexual conflicts (Kluyver 1955; Langmore and Davies 1997). Female European starlings (*Sturnus vulgaris*) and dark-eyed juncos (*Junco hyemalis*) produced song during aggression directed towards caged females placed in their territory (Sandell and Smith 1997; Reichard et al. 2018). Female song in duetting Peruvian antbirds (*Hipocnemis peruviana*) could have evolved as a signal jamming strategy, potentially disrupting extra-pair mate attraction by the male (Tobias and Seddon 2009), as has also been suggested for female barn swallows (*Hirundo rustica*) (Wilkins et al. 2020). In eastern whipbirds (*Psophodes olivaceus*), females approached more closely to playback of female solo song than to male or duet songs (Rogers et al. 2007). Like blue tits, many of these species are opportunistically polygynous (Langmore 1998). Female blue tit song could play a role in intrasexual competition over mates and territories and to actively reduce the likelihood of polygyny, which often results in lower reproductive success for females (Schlicht and Kempenaers 2021). This could be in line with early reports of female blue tit song during “reproductive fighting” (Hinde 1952), but this working hypothesis must be tested by future experimental work, measuring whether females sing or react differently to the presentation of simulated male and female intruders.

Next to a potential role in territorial defense, female solo song may also be involved in mate choice and pair formation (Langmore and Davies 1997; Langmore 1998). Female blue tits displayed solo song during daytime singing from winter to spring (as early as 17 of January, the earliest sampling date in our record). Solo singing in winter is of special interest as the social associations during this

time are related to the formation of breeding pairs (Beck et al. 2020). The early pair formation in blue tits also raises questions as to why males, but not females, sing prolifically at dawn during the fertile period. In blue tits, dawn song is tightly associated with seeking within- and extra-pair copulations by males (Welling et al. 1995; Kempenaers et al. 1997; Poesel et al. 2004; Parker et al. 2006), and with territory defense (Poesel et al. 2004). Typically, male dawn song ends when the female exits the nest and partners copulate (Sierro J, personal observation; Poesel et al. 2004; Parker et al. 2006). Normally, whether a copulation takes place or not is under female control (Kempenaers et al. 1995) and this, with the presence of frequent extra-pair copulations (Leech et al. 2001), implies a particular selection pressure on male song that appears absent in female song. Given such a contextual difference in singing behavior around the time of copulation, we can infer that selection pressures shaping song are not the same between the sexes (Austin et al. 2021), even though we need to know more about the function of female song in blue tits. Theoretically, such differential pressures could imply differences in song parameters that play a role in female choice during extra-pair copulations (Kempenaers et al. 1997; Austin et al. 2021), which is consistent with our results.

Indeed, we found that certain song traits differed between sexes, but we also found age and seasonal correlates with song that were similar in both sexes. Peak frequency was negatively correlated with age in both females and males. Female song, like male song, also varied seasonally in acoustic structure, decreasing in peak frequency, and increasing in vocal consistency towards the fertile period. These results provide female song with communicative value in many contexts since age is an important factor associated with contest success or reproductive capacity (Bradbury and Vehrencamp 1998). Furthermore, we show circumstantial evidence of spectral flexibility within individuals and song types, indicating that individuals may shift their songs to lower frequencies during the fertile period. This contrasts with previous reports of lower song frequencies during the fertile period in a related species, the great tit, that were attributed to changes in the use of song types rather than a frequency shift of the same song type (Halfwerk et al. 2011). To our knowledge, this is the first study to document seasonal changes in acoustic structure of female song, which could be of relevance in communication during different breeding stages. In males, such seasonal variation has been associated with specific functions of song during reproduction (Ballentine et al. 2003; Halfwerk et al. 2011; Vehrencamp et al. 2013). Future studies that shed light into the functional role of female song may help us to understand our observed variation in song traits.

From this quantitative assessment of female song in blue tits, we can conclude that female blue tits sing throughout the breeding season albeit with a lower total song output than males. The behavioral contexts of female song overlap with those in males, but females do not produce dawn song. Female song traits correlate with age and show significant seasonal variation indicating that variation in female song could have communicative value and may play a role in inter- or intrasexual interactions. Future experimental work should address which sexual differences and similarities in song are meaningful in communication and which selection pressures shaped the observed differences. Finally, increasing documentation of female secondary sexual traits is crucial for our understanding of the evolution of sexual signals in animal communication and should inspire large-scale comparative studies of female sexual signaling.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at <http://www.behco.oxfordjournals.org/>

We want to thank Dr. Thomas Mondain-Monval for his helpful advice on various aspects of the statistical analysis. Moreover, we are grateful to Dr. Stuart Sharp and Dr. Sarah Collins for their comments on an earlier version of the manuscript. Finally, we also thank the referees that revised this manuscript during the peer-reviewed process for their contributions.

Data availability: Analyses reported in this article can be reproduced using the data provided by Sierro et al. (2022).

Handling Editor: Niels Dingemans

REFERENCES

- Andersson S, Örnborg J, Andersson M. 1998. Ultraviolet sexual dimorphism and assortative mating in blue tits. *Proc Biol Sci.* 265:445–450.
- Austin VI, Dalziel AH, Langmore NE, Welbergen JA. 2021. Avian vocalisations: the female perspective. *Biol Rev.* 96:1484–1503.
- Bain GC, Hall ML, Mulder RA. 2014. Territory configuration moderates the frequency of extra-group mating in superb fairy-wrens. *Mol Ecol.* 23:5619.
- Ballentine B, Badyaev A, Hill GE. 2003. Changes in song complexity correspond to periods of female fertility in blue grosbeaks (*Guiraca caerulea*). *Ethology.* 109:55–66.
- Baptista LF, Petrinovich L. 1986. Song development in the white-crowned sparrow: social factors and sex differences. *Anim Behav.* 34:1359.
- Baptista LF, Trail PW, DeWolfe BB, Morton ML. 1993. Singing and its functions in female white-crowned sparrows. *Anim Behav.* 46:511–524.
- Barton K. 2011. MuMIn: multi-model inference. R package version 1.0.0. Vienna (Austria): R Foundation for Statistical Computing. Available from: <https://cran.r-project.org/web/packages/MuMIn/index.html>.
- Bates DM. 2010. lme4: mixed-effects modeling with R. Available from: <http://lme4.r-forge.r-project.org/book>.
- Beck KB, Farine DR, Kempenaers B. 2020. Winter associations predict social and extra-pair mating patterns in a wild songbird. *Proc Biol Sci.* 287:20192606.
- Beletsky LD. 1982. Vocalizations of female northern orioles. *Condor.* 84:445–447.
- Benedict L. 2008. Occurrence and life history correlates of vocal duetting in North American passerines. *J Avian Biol.* 39:57–65.
- Bijnens L. 1988. Blue tit *Parus caeruleus* song in relation to survival, reproduction and biometry. *Bird Study.* 35:61–67.
- Bijnens L, Dhondt A. 1984. Vocalizations in a Belgian blue tit, *Parus c. caeruleus*, population. *Le Gerfaut.* 74:243–269.
- Blackburn TM, Lawton JH, Pery JN. 1992. A method of estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. *Oikos.* 65:107–112.
- Botero CA, Rossman RJ, Caro LM, Stenzler LM, Lovette IJ, de Kort SR, Vehrencamp SL. 2009. Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Anim Behav.* 77:701–706.
- Bradbury JW, Vehrencamp SL. 1998. Principles of animal communication. Sunderland (MA): Sinauer Associates.
- Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.
- Burnham KP, Anderson DR, Huyvaert KP. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol.* 65:23–35.
- Cain KE, Cockburn A, Langmore NE. 2015. Female song rates in response to simulated intruder are positively related to reproductive success. *Front Ecol Evol.* 3:119.
- Catchpole CK, Slater PJB. 2008. Bird song: biological themes and variations. Cambridge (UK): Cambridge University Press.
- Clark CW, Marler P, Beeman K. 1987. Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology.* 76:101–115.
- Clutton-Brock T. 2009. Sexual selection in females. *Anim Behav.* 77:3–11.
- Coleman SW, Patricelli GL, Coyle B, Siani J, Borgia G. 2007. Female preferences drive the evolution of mimetic accuracy in male sexual displays. *Biol Lett.* 3:463–466.

- Collins S. 2004. Vocal fighting and flirting: the functions of birdsong. In: Slabbekoorn HM, Marler P, editors. *Nature's music: the science of birdsong*. San Diego (CA): Elsevier Academic Press. p. 39–79.
- Cooney R, Cockburn A. 1995. Territorial defence is the major function of female song in the superb fairy-wren, *Malurus cyaneus*. *Anim Behav*. 49:1635–1647.
- Cramp S, Perrins C. 1993. *Handbook of the birds of Europe, the Middle East and North Africa. The birds of the western Palearctic*. Vol. VII. Oxford: Oxford University Press.
- Cresswell W. 1994. Song as a pursuit-deterrent signal, and its occurrence relative to other anti-predation behaviours of skylark (*Alauda arvensis*) on attack by merlins (*Falco columbarius*). *Behav Ecol Sociobiol*. 34:217–223.
- Deepayan S. 2008. Package “lattice.” Available from: <http://lmdvr-r-forge.r-project.org>.
- de Kort SR, Eldermire ERB, Valderrama S, Botero CA, Vehrencamp SL. 2009. Trill consistency is an age-related assessment signal in banded wrens. *Proc Biol Sci*. 276:2315.
- Doutrelant C, Aubin T, Hitier S, Lambrechts MM. 1998. Two distinct song populations of blue tit (*Parus caeruleus*) in the French Mediterranean. *Bioacoustics*. 9:1–16.
- Doutrelant C, Blondel J, Perret P, Lambrechts MM. 2000a. Blue tit song repertoire size, male quality and interspecific competition. *J Avian Biol*. 31(3):360–366.
- Doutrelant C, Leitaó A, Giorgi H, Lambrechts MM. 1999. Geographical variation in blue tit song, the result of an adjustment to vegetation type? *Behaviour*. 136:481–493.
- Doutrelant C, Leitaó A, Otter K, Lambrechts MM. 2000b. Effect of blue tit song syntax on great tit territorial responsiveness—an experimental test of the character shift hypothesis. *Behav Ecol Sociobiol*. 48:119–124.
- Dowle M, Srinivasan A. 2021. Package “data.table.” R package version 1.14.0. Available from: <https://CRAN.R-project.org/package=data.table>.
- Eens M, Pinxten R. 1998. Female song for mate attraction: an overlooked phenomenon? *Trends Ecol Evol*. 8:322–323.
- Fox J, Weisberg S. 2019. *Companion to applied regression*. 3rd ed. Thousand Oaks (CA): Sage.
- Garamszegi LZ, Pavlova DZ, Eens M, Møller AP. 2007. The evolution of song in female birds in Europe. *Behav Ecol*. 18:86–96.
- Gelman A. 2008. Scaling regression inputs by dividing by two standard deviations. *Stat Med*. 27:2865–2873.
- Gil D, Llusia D. 2020. The bird dawn chorus revisited. In: Aubin T, Mathévon N, editors. *Coding strategies in vertebrate acoustic communication*. Cham (Switzerland): Springer. p. 45–90.
- Gorissen L, Eens M. 2005. Complex female vocal behaviour of great and blue tits inside the nesting cavity. *Behaviour*. 142:489–506.
- Gorissen L, Janssens E, Pinxten R, Eens M. 2002. Differences in song repertoire size and composition between two populations of blue tits (*Parus caeruleus*). *Avian Sci*. 2:39–48.
- Griffith SC, Ton R, Hurley LL, McDiarmid CS, Pacheco-Fuentes H. 2021. The ecology of the zebra finch makes it a great laboratory model but an outlier amongst passerine birds. *Birds*. 2:60–76.
- Halfwerk W, Bot S, Buikx J, van der Velde M, Komdeur J, ten Cate C, Slabbekoorn H. 2011. Low-frequency songs lose their potency in noisy urban conditions. *Proc Natl Acad Sci USA*. 108:14549–14541.
- Halkin SL. 1997. Nest-vicinity song exchanges may coordinate biparental care of northern cardinals. *Anim Behav*. 54:189–198.
- Hansen BT, Johannessen LE, Slagsvold T. 2008. Imprinted species recognition lasts for life in free-living great tits and blue tits. *Anim Behav*. 75:921–927.
- Hausberger M, Black JM. 1991. Female song in European starlings: the case of non-competitive song-matching. *Ethol Ecol Evol*. 3:337–344.
- Hijmans RJ. 2019. Package “geosphere.” R package version 1.5-10. Available from: <https://CRAN.R-project.org/package=geosphere>.
- Hinde RA. 1952. The behaviour of the great tit (*Parus major*) and some other related species. *Behav Suppl*. 2:III, V–X, 1–III, V–X201.
- Hoelzel AR. 1986. Song characteristics and response to playback of male and female robins (*Eritacus rubecula*). *Ibis*. 128:115–127.
- Irschick DJ, Herrel A, Vanhooydonck B, van Damme R. 2007. A functional approach to sexual selection. *Funct Ecol*. 21:621–626.
- Johnson LS, Kermott LH. 1990. Structure and context of female song in a north-temperate population of house wrens. *J Field Ornithol*. 61:273–284.
- Kempnaers B, Verheyen GR, Dhondt AA. 1995. Mate guarding and copulation behaviour in monogamous and polygynous blue tits: do males follow a best-of-a-bad-job strategy? *Behav Ecol Sociobiol*. 36:33–42.
- Kempnaers B, Verheyen GR, Dhondt AA. 1997. Extrapair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. *Behav Ecol*. 8:481–492.
- Kirschel AN, Blumstein DT, Smith TB. 2009. Character displacement of song and morphology in African tinkerbirds. *Proc Natl Acad Sci USA*. 106:8256.
- Kirschel AN, Zanti Z, Harlow ZT, Vallejo EE, Cody ML, Taylor CE. 2020. Females don't always sing in response to male song, but when they do, they sing to males with higher-pitched songs. *Anim Behav*. 166:129–138.
- Kluyver H. 1955. The behaviour of the great reed warbler, *Acrocephalus arundinaceus*, with special reference to nest building and territorial defence. *Ardea*. 43:1–50.
- Knief U, Forstmeier W. 2021. Violating the normality assumption may be the lesser of two evils. *Behav Res Methods*. 53:2576–2590.
- Krebs JR, Ashcroft R, Van Orsdol K. 1981. Song matching in the great tit, *Parus major* L. *Anim Behav*. 29:918–923.
- Krieg CA, Getty T. 2016. Not just for males: females use song against male and female rivals in a temperate zone songbird. *Anim Behav*. 113:39–47.
- Laiolo P, Tella JL, Carrete M, Serrano D, López G. 2004. Distress calls may honestly signal bird quality to predators. *Proc Biol Sci*. 271:S513–S515.
- Langemann U, Tavares J, Peake TM, McGregor PK. 2000. Response of great tits to escalating patterns of playback. *Behaviour*. 137:451–472.
- Langmore NE. 1998. Functions of duet and solo songs of female birds. *Trends Ecol Evol*. 13:136–140.
- Langmore N, Davies N. 1997. Female dunlocks use vocalizations to compete for males. *Anim Behav*. 53:881–890.
- Langmore NE, Davies N, Hatchwell B, Hartley IR. 1996. Female song attracts males in the alpine accentor (*Prunella collaris*). *Proc R Soc B*. 263:141–146.
- Langmore NE, Mulder RA. 1992. A novel context for bird song: predator calls prompt male singing in the kleptogamous superb fairy-wren, *Malurus cyaneus*. *Ethology*. 90:143–153.
- Latimer W. 1977. A comparative study of the songs and alarm calls of some *Parus* species. *Z Tierpsychol*. 45:414–433.
- Leech DI, Hartley IR, Stewart IR, Griffith SC, Burke T. 2001. No effect of parental quality or extrapair paternity on brood sex ratio in the blue tit (*Parus caeruleus*). *Behav Ecol*. 12:674–680.
- Ligges U. 2013. tuneR—analysis of music. Available from: <https://cran.r-project.org/web/packages/tuneR/tuneR.pdf>.
- Linhart P, Slabbekoorn H, Fuchs R. 2012. The communicative significance of song frequency and song length in territorial chiffchaffs. *Behav Ecol*. 23:1338–1347.
- Logue DM, Hall ML. 2014. Migration and the evolution of duetting in songbirds. *Proc Biol Sci*. 281:20140103.
- Magoollagan L, Sharp SP. 2018. Song function and territoriality in male and female white-throated dippers (*Cinclus cinclus*). *Bird Study*. 65:396–403.
- Mahr K, Seifert CL, Hoi H. 2016. Female and male blue tits (*Cyanistes caeruleus*) sing in response to experimental predator exposition. *J Ornithol*. 157:907–911.
- Mainwaring MC, Hartley IR. 2009. Experimental evidence for state-dependent nest weight in the blue tit, *Cyanistes caeruleus*. *Behav Process*. 81:144–146.
- Mainwaring MC, Hartley IR. 2019. Blue tits. In: Choe J, editor. *Encyclopedia of animal behavior*. San Diego (CA): Elsevier. p. 11–22.
- Marler PR, Slabbekoorn H. 2004. *Nature's music: the science of birdsong*. San Diego (CA): Elsevier Academic Press.
- Mazzoni D, Dannenberg R. 2014. Audacity. Available from: <https://www.audacityteam.org/>.
- McGregor PK. 2005. *Animal communication networks*. New York: Cambridge University Press.
- Mikula P, Tószögová A, Hořák D, Petrusková T, Storch D, Albrecht T. 2020. Female solo song and duetting are associated with different territoriality in songbirds. *Behav Ecol*. 31:322–329.
- Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev*. 82:591–605.
- Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol*. 4:133–142.
- Nice MM. 1943. Studies in the life history of the song sparrow II. The behavior of the song sparrow and other passerines. *Trans Linn Soc N Y*. 6:1–328.
- Nicolai J. 1959. Familientradition in der Gesangsentwicklung des Gimpels (*Pyrrhula pyrrhula* L.). *J Ornithol*. 100:39–46.
- Nilsson AL, Alerstam T, Nilsson J-A. 2008. Diffuse, short and slow migration among blue tits. *J Ornithol*. 149:365–373.

- Odom KJ, Benedict L. 2018. A call to document female bird songs: applications for diverse fields. *Auk*. 135:314–325.
- Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE. 2014. Female song is widespread and ancestral in songbirds. *Nat Commun*. 5:1–6.
- Parker TH, Barr IR, Griffith SC. 2006. The blue tit's song is an inconsistent signal of male condition. *Behav Ecol*. 17:1029–1040.
- Patchett R, Kirschel AN, Robins King J, Styles P, Cresswell W. 2021. Female song in the Cyprus Wheatear *Oenanthe cyprica*. *J Ornithol*. 162:1199.
- Podos J. 1997. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*. 51:537–551.
- Poesel A, Dabelsteen T. 2005. Territorial responses of male blue tits to simulated dynamic intrusions: effects of song overlap and intruder location. *Anim Behav*. 70:1419.
- Poesel A, Dabelsteen T, Pedersen SB. 2004. Dawn song of male blue tits as a predictor of competitiveness in midmorning singing interactions. *Acta Ethol*. 6:65–71.
- Poesel A, Foerster K, Kempenaers B. 2001. The dawn song of the blue tit (*Parus caeruleus*) and its role in sexual selection. *Ethology*. 107:521–531.
- Poesel A, Kempenaers B. 2000. When a bird is tired from singing: a study of drift during the dawn chorus. *Etologia*. 8:1–7.
- Price JJ. 2009. Evolution and life-history correlates of female song in the New World blackbirds. *Behav Ecol*. 20:967–977.
- Price JJ, Lanyon SM, Omland KE. 2009. Losses of female song with changes from tropical to temperate breeding in the New World blackbirds. *Proc Biol Sci*. 276:1971.
- R Development Core Team. 2016. R: a language and environment for statistical computing. Vienna (Austria). Available from: <https://www.r-project.org/>.
- Redfern CPF, Clark JA. 2001. Ringers' manual. Thetford (UK): BTO.
- Reichard DG, Brothers DE, George SE, Atwell JW, Ketterson ED. 2018. Female dark-eyed juncos (*Junco hyemalis thurberi*) produce male-like song in a territorial context during the early breeding season. *J Avian Biol*. 49:jav-01566.
- Riebel K. 2003. The “mute” sex revisited: vocal production and perception learning in female songbirds. In: Slater PJB, Rosenblatt S, Snowdon CT, Roper TJ, editors. *Advances in the study of behavior*. San Diego (CA): Elsevier Academic Press. p. 49–86.
- Riebel K, Hall ML, Langmore NE. 2005. Female songbirds still struggling to be heard. *Trends Ecol Evol*. 20:419–420.
- Riebel K, Odom KJ, Langmore NE, Hall ML. 2019. New insights from female bird song: towards an integrated approach to studying male and female communication roles. *Biol Lett*. 15:2019.
- Ritchison G. 1986. The singing behavior of female northern cardinals. *Condor*. 88:156–159.
- Robinson A. 1949. The biological significance of bird song in Australia. *Emu*. 48:291–315.
- Rogers AC, Langmore NE, Mulder RA. 2007. Function of pair duets in the eastern whipbird: cooperative defense or sexual conflict? *Behav Ecol*. 18:182–188.
- Sandell MI, Smith HG. 1997. Female aggression in the European starling during the breeding season. *Anim Behav*. 53:13–23.
- Schlicht E, Kempenaers B. 2021. Origin and outcome of social polygyny in the blue tit. *Ardea*. 109:91–118.
- Schlicht L, Kempenaers B. 2020. The effects of season, sex, age and weather on population-level variation in the timing of activity in Eurasian blue tits, *Cyanistes caeruleus*. *Ibis*. 162:1146–1162.
- Scott G. 1993. Sexing members of a Scottish blue tit (*Parus caeruleus*) population in the hand during the winter months. *Ring Migr*. 14:124–128.
- Searcy WA, Andersson M. 1986. Sexual selection and the evolution of song. *Annu Rev Ecol Syst*. 17:507–533.
- Sierro J, Selvino SR, Hartley IR. 2022. Data from: Female blue tits sing frequently: a sex comparison of occurrence, context and structure of song. *Behav Ecol*. doi: 10.5061/dryad.98sf7m0m1.
- Slater PJ, Mann NI. 2004. Why do the females of many bird species sing in the tropics? *J Avian Biol*. 35:289–294.
- Stadler H. 1951. Die stimmen der Blaumeise. *Anz Ornithol Ges Bayern*. 4:155–170.
- Sueur J, Aubin T, Simonis-Sueur C. 2006. Seewave. Available from: <https://rug.mnhn.fr/seewave/>.
- Suthers RA. 2004. How birds sing and why it matters. In: Slabbekoorn H, Marler P, editors. *Nature's music: the science of birdsong*. San Diego (CA): Elsevier Academic Press. p. 272–295.
- Svensson L. 1992. Identification guide to European passerines. Thetford, UK: British Trust for Ornithology.
- Tobias JA, Montgomerie R, Lyon BE. 2012. The evolution of female ornaments and weaponry: social selection, sexual selection and ecological competition. *Philos Trans R Soc Lond B Biol Sci*. 367:2274–2293.
- Tobias JA, Seddon N. 2009. Signal jamming mediates sexual conflict in a duetting bird. *Curr Biol*. 19:577–582.
- Vehrencamp SL, Yantachka J, Hall ML, de Kort SR. 2013. Trill performance components vary with age, season, and motivation in the banded wren. *Behav Ecol Sociobiol*. 67:409–419.
- Webb WH, Brunton DH, Aguirre JD, Thomas DB, Valcu M, Dale J. 2016. Female song occurs in songbirds with more elaborate female coloration and reduced sexual dichromatism. *Front Ecol Evol*. 4:22.
- Welling P, Koivula K, Lahti K. 1995. The dawn chorus is linked with female fertility in the willow tit *Parus montanus*. *J Avian Biol*. 26:241–246.
- Wickham H. 2016. Package “ggplot2.” Available from: <https://ggplot2.tidyverse.org>.
- Wickham H, Wickham MH. 2019. Package “stringr.” R package version 1.4.0. Available from: <https://CRAN.R-project.org/package=stringr>.
- Wickham H, Wickham MH. 2020. Package “plyr.” Available from: <http://www.jstatsoft.org/v40/i01/>.
- Wilkins MR, Odom KJ, Benedict L, Safran RJ. 2020. Analysis of female song provides insight into the evolution of sex differences in a widely studied songbird. *Anim Behav*. 168:69–82.
- Yamaguchi A. 1998. A sexually dimorphic learned birdsong in the Northern cardinal. *Condor*. 100:504–511.
- Zuur A, Ieno E, Walker N, Saveliev A, Smith G. 2009. In: *Mixed effects models and extensions in ecology with R*. Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editors. Chapter 21, p. 471–490. New York: Springer Science and Business Media.