Functions and mechanisms of kin recognition in long-tailed tits, *Aegithalos caudatus*

Amy Elizabeth Leedale

The
University
Of
Sheffield.

Thesis submitted for the degree of PhD

Department of Animal and Plant Sciences
University of Sheffield

September 2018

Contents

Chapter 1

General Introduction 9

Chapter 2

Fine-scale genetic structure and helping decisions in long-tailed tit societies 34

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Acknowledgements

Completing this PhD would not have been possible without the help and support of so many people. Firstly, thank you to Ben Hatchwell for granting me the opportunity to work on this fantastic project and your supervision throughout; our meetings were always motivating, calming and inspiring. I also thank my co-supervisor, Elva Robinson, for help with data extraction, and providing a non-avian perspective.

To Stuart Sharp, thank you for your helpful advice, particularly the sound recording tips in my first field season and the reassurance I needed in the final stages of the project. Thank you to my project supporter, Terry Burke, for several stimulating meetings, and Jon Slate for advice about heterozygosity fitness correlations. Thank you to the two anonymous reviewers whose comments improved Chapter 2.

I am particularly grateful to Robert Lachlan, who wrote the Luscinia software and took the considerable time and effort to show me how to get the most out it - your patience and enthusiasm are greatly appreciated. I also thank Julian Kapoor for bioacoustics advice and recommending Luscinia, and Raimund Specht for answering various questions about Avisoft. I especially thank Selvino de Kort for getting me started with bioacoustics and encouraging me to go for a PhD in the first place.

Thank you to everyone who has contributed in various ways to the long-tailed tit data collection. In particular, thanks to those who worked alongside me during my field seasons: Caitlin Higgott, Jamie Hutchison, Thomas Mondain-Monval and Catherine Hammond. Thank you to everyone in the NERC Biomolecular Analysis Facility for assistance and support; particularly Natalie dos Remedios for all your lab work, and allowing me to follow you around asking questions.

Thank you to the residents of the Rivelin Valley for allowing me to monitor nests on your land and in your gardens; offering tea, chocolate, even wine; and showing such an interest in the project. Thanks as well to Yorkshire Water and Sheffield City Council for granting access to certain areas. A special thanks to Steve and Phil at Sorby Breck, for those amazing mornings ringing and giving me a heads up whenever one of our birds showed up in your nets.

To my fellow B1.204 inhabitants, thank you for making my days in the office so enjoyable. In particular, thanks to René van Dijk for bioacoustic and genetics advice; Mark Adams for introducing me to the long-tailed tit database; Jonathan Green for help with sibship reconstruction and many stimulating discussions; and Nyil Khwaja for help with repeatability and for great chats about science, birds and everything else.

To all the past and present members of LHB, thanks for being such an inspiring, entertaining and supportive lab group. I always looked forward to our meetings, which were dependably productive and engaging. Particularly, thanks to Joel Pick, whose simulations course helped me with randomisation; John Jackson, for taking some lovely fieldwork photos; and Hannah Mumby, for sage advice during the early days.

Outside of LHB, thanks to Ben Jackson and Simon Mills for teaching me how to write better R code, and Lilly Herridge for helping me create some really nice plots. I am also grateful to all the members of the behavioural ecology group at the University of York - you welcomed me to your meetings and always asked interesting questions.

During my time at Sheffield, I have been fortunate enough to make a great many friends, too many to name individually, but all of whom have made the last four years fun. Special thanks to everyone involved in the squash and badminton tournaments; I don’t think I ever won a game, but getting a good rally going was always a great start to the day. To everyone else in department of Animal & Plant Sciences, thanks for making it a fantastic place to work.

Thank you to Lucie Ward for creating a beautiful long-tailed tit illustration. This is a much better start to the thesis than any of my dismal attempts at long-tailed tit photography.

John, this thesis would not have made it onto these pages without you. Thank you for listening with patience to endless natterings, the occasional outburst, and practice rounds of every presentation I have ever had to give. You helped me keep perspective and always found a way to make me laugh.

Finally, I have to thank the long-tailed tits. Their dogged determination to reproduce, despite the unfavourable odds, inspired me to never give up.

This PhD was funded by the Natural Environment Research Council (NERC, UK, 1517208).

Statement of intellectual contribution

The research chapters (2-5) in this thesis are presented as manuscripts, and have benefited from collaboration with or advice from several colleagues. Ben Hatchwell supervised the project and conceived and managed the long-tailed tit study. Elva Robinson co-supervised the project. A large number of field researchers have contributed to the long-term data collection and population monitoring. I collected and analysed all of the bioacoustic data presented in this thesis. DNA extraction and genotyping were conducted at the NERC Biomolecular Analysis Facility at the University of Sheffield, with advice and support from Terry Burke, Deborah Dawson, Natalie dos Remedios and Maria-Elena Mannarelli. I carried out all other molecular genetic analyses. Further, specific contributions made by co-authors are detailed below. All other work is my own.

Chapter 2: I designed the study with Ben Hatchwell, Stuart Sharp & Michelle Simeoni. Ben Hatchwell, Stuart Sharp, Michelle Simeoni & Elva Robinson provided comments that improved the manuscript.

Chapter 3: I designed the study with Ben Hatchwell, Stuart Sharp & Michelle Simeoni. Jonathan Green provided the direct fitness data used to analyse inbreeding costs. Ben Hatchwell, Elva Robinson and Jonathan Green provided comments that improved the manuscript.

Chapter 4: Ben Hatchwell and Elva Robinson provided comments that improved the manuscript.

Chapter 5: Ben Hatchwell provided comments that improved the manuscript.

General Summary

Many cooperative societies are composed of relatives, and kin selection can often explain cooperation in such groups. However, prolonged association with relatives may also increase inbreeding risk. Here, I investigate kin discrimination in the contexts of helping and mate choice in the long-tailed tit *Aegithalos caudatus*, a facultative cooperative breeder in which failed breeders redirect help towards relatives.

In Chapter 2, I quantify the fine-scale genetic structure within breeding populations, generated by the life-history and dispersal patterns associated with this unusual helping system. Kin remain clustered after dispersal, but helping patterns do not reflect kin structure; help is redirected towards close kin more often than expected by indiscriminate helping. This population structure also creates a potential inbreeding risk, and heterozygosity-fitness correlations indicate that inbreeding carries fitness costs (Chapter 3). Remarkably, this risk is alleviated by active avoidance of close kin as partners. In Chapters 4 & 5, I consider the recognition mechanism that permits kin-directed helping and active inbreeding avoidance in this species. I devise a comprehensive method of measuring acoustic variation within and between individuals, and investigate whether vocal similarity may be used to assess relatedness and make adaptive helping and pairing decisions. Failed breeders redirect help towards the nests of males with similar calls, but call similarity within breeding pairs is higher than one would expect from random mating. Possible reasons for this are discussed.

The variety of kin discrimination across social systems, and the circumstances under which certain mechanisms may be adaptive, are discussed. I highlight important considerations for assessing kin recognition mechanisms and the role of familiarity in kin discrimination.

Chapter 1

General Introduction

Cooperation is widespread in nature. From an evolutionary perspective, cooperation includes any behaviour that generates a fitness benefit to another individual, and is selected for because of its beneficial effect on the recipient (West *et al.* 2007a). In many cases, cooperation carries mutual benefits; both actor and recipient gain direct fitness from the actor's behaviour. The evolution of altruistic cooperative behaviour, in which the actor increases the recipient's fitness at a cost to themselves, is more difficult to explain, and at one time presented a major problem for the theory of natural selection, troubling even Darwin himself (Darwin 1859). In what has become one of the most important advances in evolutionary biology, Hamilton (1964) provided a solution to this seeming paradox when he developed kin selection theory and the concept of inclusive fitness (Hamilton 1963; 1964; Axelrod & Hamilton 1981). In a series of seminal papers, Hamilton argued that individuals can increase their genetic contribution to the next generation by helping their relatives, with whom they share a number of genes that are identical by descent, to survive and reproduce. He used the term *inclusive fitness* to describe the direct fitness an individual can gain from independent reproduction, combined with the indirect fitness that can be gained by improving the reproductive success of kin (Hamilton 1964). The process by which indirect benefits accrue was termed *kin selection* by Maynard Smith (1964).

Hamilton's ideas have been applied to various aspects of evolutionary biology. Inclusive fitness theory has helped to explain some of the major evolutionary transitions in the history of life on earth; chromosome evolution, unicellularity to multicellularity, and the evolution of eusociality are characterised by an increase in complexity and all involve cooperation at some level (Bourke 2011). In the case of social evolution, longitudinal studies of cooperative breeders have been particularly insightful, and provide an excellent arena in which to investigate kin selection.

1.1 Cooperative breeding

Cooperative breeding refers to a reproductive system in which more than a pair of individuals collectively raise young in a single brood or litter (Koenig & Dickinson 2016). Although relatively uncommon, this alloparenting has been described in a range of taxa, including fish (Taborsky & Limberger 1981), crustaceans (Duffy 2010), insects and arachnids (Wilson 1971; Choe & Crespi 2010), birds (Koenig & Dickinson 2004) and mammals (Clutton-Brock 2016). Cooperation continues to be a key focus in evolutionary biology; although Hamilton has provided the adaptive framework for social evolution (Hamilton 1964), why we see so much diversity among cooperative breeders, and why this has evolved in some species and not others, is still something of a mystery (Rubenstein & Abbott 2017). As an umbrella term that covers a diverse array of social mating systems, a common evolutionary theme that applies across cooperative breeders has proved challenging to establish.

In the majority of cooperative breeders, adult offspring delay dispersal and remain with their parents on their natal territory, foregoing personal reproduction to help raise their siblings. This has been labelled *helping at the nest* in birds (Brown 1987), but is a common format in non-avian systems as well. However, there is enormous variation on this general theme. In some species, helpers have the opportunity to breed (Martin-

Vivaldi *et al*. 2002), and help is often given by non-kin (Kokko *et al.* 2002; Riehl 2013) or collateral kin (Hatchwell 2016). In most species, cooperative breeding is *facultative*, and breeders can raise young without assistance (e.g. Emlen & Wrege 1988; Russell & Hatchwell 2001), although there are a few *obligate* cooperative breeders, which require help to raise their young (e.g. Russell *et al.* 2010; Wright *et al.* 2010). In cooperatively breeding mammals, dominant breeders are assisted by non-breeding subordinates. Although most are *singular breeders*, there a few *plural cooperative breeders*, in which multiple individuals of both sexes breed in the same social unit (e.g. Cant *et al.* 2013). Cooperative breeding in birds can take various forms of *cooperative polygamy*, e.g. plural or joint-nesting breeders; and *coloniality*, in which singular breeding females nest in extremely close proximity (Cockburn *et al.* 2004). Cooperative breeding has evolved independently across multiple distant lineages and its diversity reflects differences in the strength and nature of selection on helping, mating strategy, dispersal and other aspects of social living across species (Ligon & Burt 2004).

For cooperative breeding, in any form, to be evolutionarily stable, there must be a fitness benefit to helping. The most widely supported driver of cooperation is the indirect fitness one can gain from helping relatives (Emlen 1991), and in most cooperative societies, helping is kin-biased (Hatchwell *et al.* 2009; Riehl 2013). By assisting relatives, helpers can increase the productivity of the current breeding attempt, or reduce the reproductive effort of the parents. This *load lightening* increases the probability that the breeding relative will survive and breed successfully in the future (reviewed in Dickinson & Hatchwell 2004).

There are fewer empirical examples of helpers gaining direct fitness benefits, but these may include: increased social prestige (Zahavi 1995), acquisition of skills that increase future mating success (Komdeur 1996); territory inheritance (Wolfenden 1975); mutualistic benefits associated with group living (*group augmentation*), including delayed reciprocity (Kokko *et al.* 2001, Ligon & Ligon 1978); or access to mates, through extra-pair copulations (Mulder *et. al* 1994) or replacement (Curry & Grant 1990). In some species, cooperation is enforced by aggression, and subordinates may *pay to stay* in the territory, helping dominants in order to increase the likelihood of being tolerated (Kokko *et al.* 2002; Bergmüller *et al.* 2005; Bergmüller & Taborsky 2005).

However, *helpers at the nest* usually pay a net fitness cost by helping instead of breeding themselves; even helping relatives only partly compensates for failing to disperse and breed independently (Ekman *et al.* 2004). So, the question, 'why breed cooperatively?' becomes 'why delay dispersal?' The widely accepted answer is that delayed dispersal has evolved due to ecological constraints on independent breeding (Koenig *et al.* 1992; Emlen 1997). Such constraints include: a shortage of vacant breeding territories; a high mortality risk associated with dispersal; or a low probability of finding a mate or reproducing successfully (Hatchwell & Komdeur 2000). The *ecological constraints hypothesis* proposes that cooperative breeding may evolve in any species under certain environmental conditions that favour the decision to stay and help, rather than disperse and attempt independent reproduction (Emlen 1982). Cooperative breeding appears to have evolved most frequently in harsh, unpredictable environments, with low or variable rainfall (Jetz and Rubenstein 2011; Lukas & Clutton-Brock 2017), which supports the ecological constraints model.

Alternatively, the *benefits of philopatry hypothesis* proposes that the decision to delay dispersal is driven by the benefits of remaining in the natal area, rather than constraints on leaving (Stacey & Ligon 1987). These may include inheritance of a high-quality territory and benefits associated with interactions with kin (Covas & Griesser 2007). These two hypotheses are not mutually exclusive, and are arguably complementary, rather than alternative models (Koenig *et al.* 1992); both assume a cost-benefit analysis of leaving versus remaining, in which remaining returns the higher fitness pay-off (Hatchwell & Komdeur 2000). However, determining the relative importance of the extrinsic constraints and intrinsic benefits of delayed dispersal can provide a more comprehensive understanding of the evolution of cooperative breeding. Recently, the *dual benefits framework* has been put forward to incorporate the direct benefits of group living, such as resource defence, which may also influence dispersal decisions (Shen *et al.* 2017).

In contrast, the *life-history hypothesis* focuses on the role of phylogeny and life-history traits such as longevity, and proposes that cooperative breeding will evolve only in lineages with the appropriate attributes (Arnold & Owens 1998). For example, phylogenetic reconstructions have revealed that monogamy is often a precursor to the evolution of cooperative breeding (Boomsma 2009; Cornwallis *et al.* 2010; Lukas & Clutton-Brock 2012). The effect helpers have on offspring fitness may also need to be high for cooperation to evolve. In birds, the evolution of cooperative breeding is associated with producing altricial young (Ligon & Burt 2004). In mammals, it has been associated with polytocy (Lukas & Clutton-Brock 2017). However, whether ecological traits facilitate the evolution of cooperation in certain species that are predisposed to cooperation because of their life-history, or vice versa, has remained difficult to determine. No hypothesis alone appears adequate in determining a clear evolutionary route to cooperative behaviour. Rather, certain aspects of ecology and life-history may need to align in order for cooperative breeding to be an adaptive strategy (Hatchwell & Komdeur 2000).

The term cooperative breeding encompasses an array of reproductive strategies, and there is great diversity in mating system, life-history trajectory, dispersal patterns and population structure within cooperatively breeding species. These differences drive the evolution of specific cooperative behaviours and decision rules. In species where indirect fitness benefits appear to be driving cooperation, processes that facilitate the differential treatment of conspecifics differing in relatedness are expected to be under strong selection.

1.2 Functions of kin recognition

Integral to kin selection theory is the ability of individuals to discriminate between kin and non-kin (Holmes & Sherman 1983). Consequently, kin discrimination has been the focus of decades of theoretical and empirical research (Hepper 1986; Holmes 2004; Riehl & Stern 2015). Although most researchers agree that kin discrimination is advantageous, the nature of this ability is considerably debated. This has been exacerbated by semantic arguments (Byers & Bekoff 1986; Barnard 1991; Bekoff 1992). Some researchers use the term *kin recognition* simply to describe the differential treatment of conspecifics differing in genetic relatedness (Sherman *et al.* 1997). Others prefer to reserve the term *kin recognition* for specific mechanisms involving an assessment of genetic relatedness based on sensory information acquired from a conspecifics' phenotype (Tang-Martinez 2001). In a further step, Grafen (1990) argued that only when these mechanisms have demonstrably evolved because of their adaptive function, should they be considered true kin recognition (Grafen 1990). While it is important to understand that discriminatory behaviour need not involve complex mechanisms, restrictive definitions can actually hinder our understanding of how these systems work (Komdeur & Hatchwell 1999). Throughout this thesis, I follow Sherman *et al.* (1997) and use *kin recognition* to describe any mechanism by which conspecifics differing in genetic relatedness are differentially treated (which I refer to as *kin discrimination*). Kin recognition in this sense has been identified in a range of taxa from microbes to humans (Sherman *et al.* 1997; Biedrzycki & Bais, 2010) and is adaptive in a wide range of contexts (Hepper 1991).

1.2.1 Kin-directed cooperation

According to Hamilton's rule, cooperation among relatives can provide indirect fitness benefits as long as the co-efficient of relatedness multiplied by the benefits of cooperation to the recipient exceed the costs to the actor (Hamilton 1964). Therefore, an ability to discriminate not only between kin and non-kin, but also between conspecifics of varying degrees of kinship can be selectively advantageous. In addition to the well-documented inclusive fitness benefits of alloparenting, kin recognition can permit kin-biases in reproductive tolerance (Clutton-Brock *et al.* 2010) and cooperative mate attraction (Díaz-Muñoz *et al.* 2014); maintain social cohesion (Queller *et al.* 1993); and facilitate communal investment in defence (Griesser & Ekman 2005) and other public goods (van Dijk *et al.* 2014). Kin recognition can also allow individuals to direct antagonistic or spiteful behaviour towards non-kin (Hamilton 1970; Pfennig *et al.* 1993).

1.2.2 Inbreeding avoidance

The second major function of kin recognition is inbreeding avoidance (Pusey & Wolf 1996). Breeding among relatives can often carry severe fitness costs, primarily because the offspring of related parents are more homozygous than those of unrelated parents. An increase in the proportion of homozygotes in a population can unmask deleterious recessive alleles and lead to inbreeding depression (Charlesworth & Charlesworth 1999). Inbreeding depression, a severe reduction in the fitness of inbred individuals, is well documented in captive animals, but can be exacerbated in the wild because of stochastic processes (Keller & Waller 2002). Yet, inbreeding in nature is extremely rare (Pusey & Wolf 1996), suggesting strong selection for effective avoidance mechanisms. The strength of inbreeding avoidance depends on the relative costs of inbreeding depression, inbreeding risk, and the costs of the avoidance behaviour. For instance, extensive outbreeding can also reduce fitness through the break-up of coadapted gene complexes due to recombination, or the loss or suppression of genes for local adaptations (Bateson 1978; 1983). Selection should favour mechanisms which optimise the balance between inbreeding and outbreeding (Lynch 1991; Bonneaud *et al.* 2006). As a classic example, Bateson (1978) found that male Japanese quails *Coturnix japonica* prefer to mate with unfamiliar females, but that both familiar and unfamiliar individuals are preferred to those with a grossly unfamiliar phenotype.

In many cooperative species, sex-biased dispersal before reproduction alleviates inbreeding risk (Pusey 1987; Koenig & Haydock 2004) without the need for effective kin recognition during mate choice. Indeed, inbreeding avoidance is regarded as a driving force in the evolution of dispersal strategies (Johnson & Gaines 1990). However, dispersal rates are usually low in cooperative breeders (Ekman *et al.* 2004). In such species, incest avoidance is probably an important function of kin recognition, and most evidence of active inbreeding avoidance comes from animals that cooperate with kin (Clutton-Brock *et al.* 2010). To alleviate incest, some social species engage in extra-pair copulations from outside the natal group (Varian-Ramos & Webster 2012); reproductively suppress and evict kin from the breeding group (Clutton-Brock *et al.* 2010); or actively avoid close relatives when choosing mates (Dickinson *et al.* 2016). All of these mechanisms require some form of kin recognition.

1.3 Mechanisms of kin recognition

Studies into kin recognition mechanisms are not new: since this ability was first documented in the late 1970's, researchers have identified kin recognition mechanisms across a range of taxa (Hepper 1991; Sherman *et al.* 1997), including single-celled organisms (Benabentos *et al.* 2009) and plants (Biedrzycki & Bais, 2010). However, much remains unknown. Although kin recognition mechanisms are often categorised based on their apparent complexity (Komdeur & Hatchwell 1999), considering them along a continuum may be more informative. In all recognition systems, a discriminating individual, or *actor*, acquires *cues* to kinship from a *referent* (itself, a subset of kin, or the local environment) and uses these cues to form a *template* (Reeve 1989). This template is compared with the phenotype of an encountered conspecific, or *recipient*, and an assessment about kinship is made based on the perceived similarity between the template and the recipient's phenotype (Lacey & Sherman 1983). A specific action is then taken, based on this assessment.

1.3.1 The acceptance threshold model

Reeve (1989) devised a conceptual framework, known as the *acceptance threshold model*, to explain how an assessment of recognition cues is used to make an adaptive decision (Fig. 1.1). In his model, Reeve (1989) considered the frequency distributions of the level of dissimilarity between the recipient's cue and the actor's template, for two categories of recipients. *Desirable* recipients are those which, following acceptance, provide greater fitness pay-offs to the actor than *undesirable* recipients. The model illustrates how an actor may choose to respond depending on an acceptance threshold: a degree of template-cue dissimilarity, below which it will accept and above which it will reject conspecifics as kin (Reeve 1989). Although selection should favour effective recognition systems, wild populations will always exhibit some amount of error because templates are matched against a finite set of cues (Lacy & Sherman 1983), which overlap in desirable and undesirable recipients because of individual variation (Sherman *et al*. 1997). Therefore, any recognition system will involve a certain amount of *acceptance errors*, where undesirable recipients are accepted, and *rejection errors*, where desirable recipients are rejected (Reeve 1989). Conceptually, the position of the acceptance threshold determines the risk of the two error types and selection will favour a threshold in which they are optimally balanced (Reeve 1989).

The acceptance threshold model theorises how an assessment of template–phenotype dissimilarity allows individuals to recognise kin. But, how do individuals acquire reliable cues to advertise themselves and internal templates to assess the cues of others? In the literature, kin recognition is considered as a three-component system. The mode and development of cues is the *production component* of kin recognition. The referent, template and algorithm for assessing the similarity between the template and the recipient's phenotype is the *perception component*. The nature and determinants of the action taken based on the assessment fall under the *action component* (Reeve 1989). The distinction between simple decision rules and complex cognitive processes merely represents differences in the referent and algorithm used in the assessment. Each component can evolve independently, and when determining the selective pressure acting on kin recognition mechanisms it is useful to consider the three components of the recognition system separately.

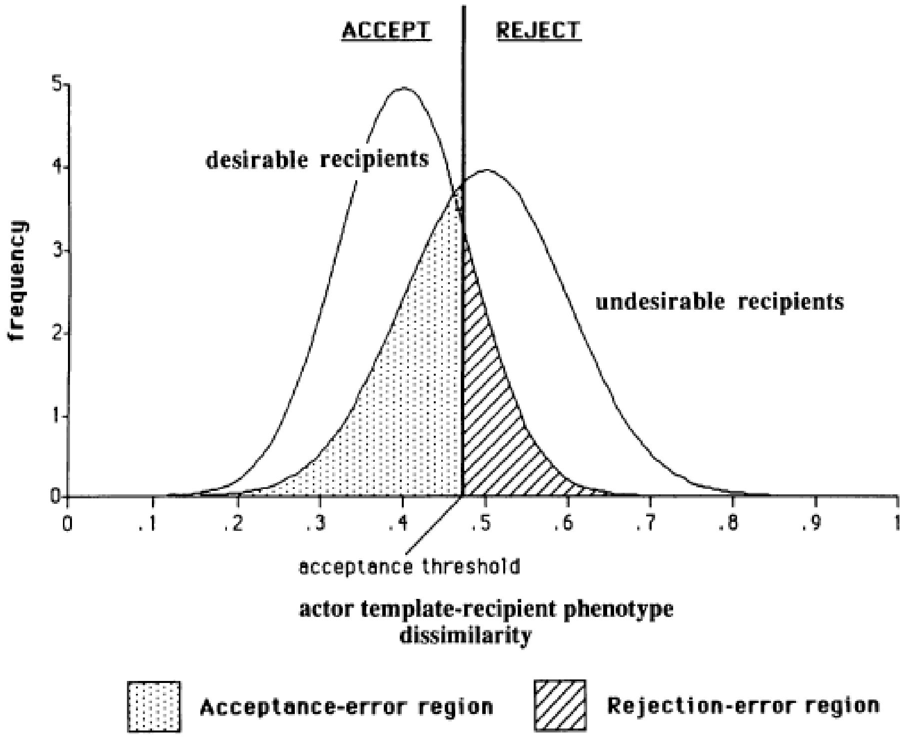

Figure 1.1. Frequency distributions of dissimilarity between the actor template and the recipient phenotype for *desirable* and *undesirable* recipients. Note the position of the acceptance threshold and the acceptance and rejection error regions generated by the threshold. All recognition systems will generate some amount of error depending on the overlap between the recognition cues of desirable and undesirable recipients and the position of the acceptance threshold (from Reeve 1989).

1.3.2 The production component

Any phenotypic trait that reliably indicates kinship may be used for kin recognition. When relatives are predictably distributed in space, location can correlate reliably with genetic relatedness (Komdeur & Hatchwell 1999). More direct phenotypic cues relay sensory information via any of the modalities organisms use to communicate, and include visual, acoustic and chemical stimuli (Halpin 1991).

Recognition cues can be genetic (Holmes 1986, Yamazaki *et al.* 1976) or acquired from the environment (Gamboa *et al*. 1986, Sharp *et al.* 2005). Genetic cues have been found in house mice *Mus musculus domesticus*, females of which prefer nest partners with their own major urinary protein (MUP) genotype (Green *et al.* 2015). Environmental cues can be acquired through diet and nest material, or from other individuals (Hepper 1991). Systems which rely entirely on genetic cues may be susceptible to rejection errors because recombination can cause even close kin to be genetically dissimilar at some loci (Gamboa *et al.* 1986). Whereas, systems which rely entirely on environmental cues may be more susceptible to acceptance errors if kin and non-kin are able to acquire them (Sherman *et al.* 1997). Recognition cues that are imprinted or learnt are reliable provided they have been learnt or imprinted from kin (Sharp *et al.* 2005). It may be more common for recognition cues to develop from a combination of genes and the environment. For example, northern paper wasps *Polistes fuscatus* absorb hydrocarbons from nest fibres at eclosion (Gamboa 1996), which combine with genetically determined odours to create colony-specific labels (Gamboa 1986).

The information these cues provide may be individual-, family- or group-specific (Radford 2005; Sharp *et al.* 2006). However, teasing apart these levels of specificity can be challenging. Effective recognition cues must have greater inter-individual than intra-individual variation (Sharp *et al.* 2006), so that they convey information about identity. To this end, they are likely to be made up of several components that vary in their combination from individual to individual. This specific profile of phenotypic traits is referred to as a *signature system* (Beecher 1982).

1.3.3 The perception component

The perception component of kin recognition concerns the referent, the forming of the template, and the algorithm for assessing the similarity between the template and the recipient’s phenotype. The perception component can be categorised into four proximate mechanisms: *spatial recognition*, *recognition alleles*, *associative learning* and *phenotype matching* (Komdeur & Hatchwell 1999; Mateo 2004).

In spatial recognition, individuals encountered in a certain area are treated as kin. Some argue that this is not true kin recognition, as individuals are responding to location, not phenotype (Halpin 1991; Tang-Martinez 2001). However, in many natural populations, non-relatives are rarely encountered in certain areas and simple decision rules such as *treat anything in my nest as kin*, are widely used in birds (Komdeur & Hatchwell 1999; Hatchwell *et al*. 2001b). This is common in parent-offspring interactions (Beecher 1991) and is exploited both intra- and interspecifically, by brood parasites (Soler *et al*. 1995) and extra-pair males (Kempenaers & Sheldon 1996).

Recognition alleles are specific alleles responsible for all three components of the recognition system (Mateo 2004). However, two thought experiments explain why this type of kin recognition is evolutionarily unstable. First, mutant cheats who carry the phenotypic cue, but not the associated behaviour or relatedness may evolve and spread through the population (Hamilton 1964). This is known as the *green-beard effect*, after Dawkins (1976) famously used the example of a green beard to illustrate how a gene or cluster of genes could recognise copies of itself in other carriers. Alternatively, assuming social interactions among kin are beneficial, individuals bearing the recognition alleles will gain higher reproductive success and eventually these alleles will become fixated (Crozier 1986). This is known as *Crozier’s paradox* (Rousset & Roze 2007). In both scenarios, over time, the genetic cue or allele would no longer correlate reliably with kinship or the associated behaviour, rendering it useless for kin recognition (Gardner & West 2007). Empirical evidence of such alleles is scarce (Sherman 1997; Komdeur & Hatchwell 1999), although a greenbeard gene may have been identified in the slime mould, *Dictyostelium discoideum* (Queller *et al.* 2003). Greenbeard effects may be more common in micro-organisms, in which all aspects of the recognition system are orchestrated at the cellular level. Genetic kin recognition could be stabilised by extrinsic processes that maintain allelic diversity, such as an advantage of certain rare alleles (Rousset & Roze 2007) or a secondary function in alleles involved in kin recognition (Holman *et al.* 2013).

It is more likely that recognition templates are acquired through some degree of learning. In the literature, two perceptual processes have been described: associative learning and phenotype-matching. Associative learning is a mechanism by which recognition occurs through prior association (Sherman *et al.* 1997; Mateo 2004). Recognition cues are learned (Sharp *et al.* 2005) or imprinted (Bateson 1978) from parents, siblings or nest-mates (Hepper 1991) at a sensitive phase during development (Komdeur & Hatchwell 1999), and used to form a template with which to compare conspecific phenotypes later in life. Importantly, templates must be acquired at a time when all individuals present are likely to be kin. Kin recognition via associative learning allows individuals to recognise familiar kin only, and could lead to errors if non-kin are encountered during the learning phase, or if kin are not encountered until afterwards (Hatchwell *et al.* 2001b). The frequency of errors will depend on the overlap in phenotypic similarity between kin and non-kin (Reeve 1989). As long as there is a reliable correlation between genetic relatedness and association during the *associative-learning period*, this is a reliable and accurate mechanism.

In phenotype-matching, recognition cues of familiar kin are also learned to form a template (Lacey & Sherman 1983). But, previously learned templates are generalised (Greenberg 1979; Holmes & Sherman 1982), so a period of previous association is not required for kin to recognise one another (Tang-Martinez 2001). Instead, a positive correlation between cue similarity and level of genetic relatedness is required (Mateo 2004), so the recipients with phenotypes that most closely match the actor's general template are its closest kin. Self-referent phenotype matching (Holmes & Sherman 1982), or the *armpit effect* (Dawkins 1982) is an extension of phenotype matching, where individuals learn and use their own phenotype to form a template (Lacey & Sherman 1983). It is possible that individuals use a combination of their own and their familiar kin's phenotype to form recognition templates (Mateo & Johnston 2000).

The categorisation of perceptual kin recognition mechanisms has been widely criticised (e.g. Hepper 1991; Tang-Martinez 2001). Since the production and perception components of kin recognition evolve separately, as long as cues vary with relatedness, any of above processes can be used (Mateo 2004) and they are not necessarily mutually exclusive. Most research in this field has treated phenotype-matching and prior association as alternative processes (Holmes and Sherman 1983), but functionally, in terms of Hamilton's (1964) rule, they provide the same outcome. Furthermore, both involve matching phenotypes to templates; the two mechanisms differ only in the specificity of the template employed (Reeve 1989). However, this makes teasing apart these mechanisms experimentally difficult, and many studies have suggested that both associative learning and phenotype-matching could work together to mediate kin recognition within a species or population (Greenberg 1979; Holmes & Sherman 1982; Komdeur & Hatchwell 1999).

1.3.4 The action component

The degree of template-phenotype similarity required to initiate a particular behaviour depends on ecological and social factors, and the costs and benefits of kin discrimination (Sherman *et al.* 1997). The acceptance threshold model (Reeve 1989) demonstrates how recognition-mediated decision rules vary with recognition context to maximise inclusive fitness. Although evolutionarily stable acceptance thresholds should balance the probability of acceptance and rejection errors, selection should also favour the ability for thresholds to adjust depending on two recognition contexts: the frequency of interactions between actors and recipients, and the fitness consequences of accepting and rejecting recipients (Reeve 1989). When there is a limit to how many desirables can be accepted, the threshold should be restrictive. When the number of desirables is unlimited, the threshold should be permissive. Additionally, as the fitness costs of accepting undesirable recipients increase, the threshold should become more restrictive. Conversely, the threshold should become more permissive as the costs of rejecting desirable recipients increase. These recognition contexts depend on ecology, life-history and the behaviour performed (Reeve 1989).

The first field data to support this model came from honey bees *Apis mellifera* (Downs & Ratnieks 2000). Honey bees often attempt to steal honey from other colonies, and guards recognise and exclude non-nest mates using odour cues (Greenberg 1979; Gamboa *et al*. 1986; 1996). Downs & Ratnieks (2000) found that when nectar availability was low, guards rejected the majority (75%) of non-nest mates and a considerable minority (20%) of nest mates. As nectar availability increased, guards became more permissive, until an accept-all strategy was adopted (Downs & Ratnieks 2000). This shift in acceptance threshold can occur rapidly in response to a sudden increase in conspecific intruders (Couvillon *et al*. 2008).

1.4 Kin recognition in cooperatively breeding birds

The factors driving cooperation have been well-studied in many cooperatively breeding birds, yet relatively few studies have explored how individuals in kin-selected systems recognise kin. Within studied species, there are gaps in our knowledge (Table 1.1).

1.4.1 Kin recognition and helping behaviour

Generally, associative learning has been identified as the most likely mechanism of kin recognition (Curry 1988; Hatchwell *et al.* 2001b; Komdeur 1994). Extended brood care at the nest provides an ideal period during which kin associate and reliable recognition templates can form. When extra-pair paternity (EPP) and brood parasitism is rare, association during the putative period accurately reflects kinship, and a simple rule such as *feed anyone who was present in my nest*, can be selected for (Komdeur & Hatchwell 1999). Galápagos mockingbirds *Nesomimus parvulus* discriminate based on prior association, rather than kinship (Curry & Grant 1990). In complex societies, a more precise rule, such as *feed anyone that fed me as a nestling* may be more reliable (Curry & Grant 1990; Komdeur 1994). In most cooperatively breeding birds, males are the predominant helping sex, but in the Seychelles warbler *Acrocephalus sechellensis* females are more likely to help than males, and choose to help at nests belonging to female breeders who fed them as nestlings, even if they are not the closest genetic relatives (Komdeur 1994; Richardson *et al.* 2003). This may be evolutionarily logical in species with high levels of EPP, such as Seychelles warblers, because helpers are often unrelated to the male that fed them (Richardson *et al.* 2003). Cross-fostering experiments confirm that female subordinates base their helping decisions on associative learning and it is unlikely that young can discriminate between their mother and any other female helper (Komdeur 2004).

Table 1.1. Cooperatively breeding birds in which kin or group recognition mechanisms have been identified.

Species	Recognition Cue	Origin	Signature	Recognition Mechanism	Experimental Method/Analysis	Reference
Bell Miner *Manorina melanophrys*	*Mew* calls	Genetic	Individual-specific	Phenotype matching?	Call similarity analysis	Wright *et al.* 2010; McDonald & Wright 2011
Galápagos Mockingbird *Nesomimus parvulus*	Unknown	Unknown	Unknown	Associative Learning	Observed patterns of helping	Curry 1988
Chestnut-crowned babbler *Pomatostomus rifuceps*	Long-distance contact calls	Unknown	Individual-specific	Unknown	Call similarity analysis	Crane *et al.* 2014
Green Woodhoopoe *Phoeniculus purpureus*	*Rally* calls	Unknown	Group-specific	Unknown	Call similarity analysis	Radford 2005
Long-tailed tit *Aegithalos caudatus*	*Churr* and *Triple* calls	Learned	Family-specific	Associative learning or phenotype-matching?	Call similarity analysis, playback experiments, cross-fostering.	Hatchwell *et al.* 2001b; Sharp *et al.* 2005; Sharp & Hatchwell 2005; 2006.
Mexican jay *Aphelocoma wollweberi*	Primary call	Unknown	Group-specific	Unknown	Playback experiments	Hopp *et al.* 2001
Seychelles Warbler *Acrocephalus sechellensis*	Unknown	Unknown	Unknown	Associative Learning	Cross-fostering	Komdeur 1994; 2004; Richardson 2003
Splendid Fairy-wren *Malurus splendens*	Unknown	Unknown	Unknown	Associative Learning	Playback experiments	Payne *et al.* 1988;
Stripe-backed Wren *Campylorhynchus nuchalis*	*WAY* calls	Learned	Family-specific	Unknown	Playback experiments	Price 1998; 1999
Superb Starling *Lamprotornis superbus*	Flight calls	Unknown	Group-specific	Unknown	Call similarity analysis, playback experiments	Keen *et al.* 2013
Western Bluebird *Sialia mexicana*	*Pew* song notes	Unknown	Group-specific	Unknown	Playback experiments	Açkay *et al.* 2013; 2014

Although associative learning is common, it may not be the only mechanism. Learned individual recognition may be possible in small groups, but in large, complex groups, individuals may not have the cognitive ability to learn individual phenotypes. In this context, phenotype matching of generic templates or a combination of both mechanisms may be more likely (Komdeur & Hatchwell 1999). Phenotype matching using olfactory cues is common in social mammals (Boyse 1991) and acoustic phenotype matching has been found in rhesus macaques *Macaca mulatta* (Pfefferle *et al.* 2013). As yet, there is no conclusive evidence of acoustic phenotype matching in birds (Pfefferle *et al.* 2013).

Playback experiments have demonstrated vocal kin discrimination in several species (Payne *et al*., 1988; Price, 1998; 1999; Hatchwell *et al*., 2001), yet very little is known about how vocalisations develop. A study by Sharp *et al.* (2005) demonstrated that the *churr* calls of long-tailed tits *Aegithalos caudatus*, known to be involved in kin recognition, were learned during development (Sharp *et al.* 2005). In contrast, stripe-backed wrens *Campylorhynchus nuchalis* produced and recognised the *WAY* calls of their genetic fathers, even when they were from different groups, suggesting a genetic component (Price 1998; 1999).

Kin recognition in cooperative breeders has generally been thought of as threshold-based, where helpers decide whether and whom to help based on perceived kinship (Curry 1988; Komdeur 1994; Dickinson *et al*. 1996). In a few species however, helpers have been shown to modify provisioning effort according to degree of relatedness to the recipient brood (Nam *et al*. 2010; Wright *et al*. 2010). For example, Nam *et al*. (2010) found a strong effect of genetic relatedness on helper provisioning rates in long-tailed tits, suggesting a more sophisticated mechanism of kin recognition than associative learning of family-specific cues. If individuals are able to use a call similarity gradient to adjust helping behaviour according to relatedness, call similarity would be expected to positively correlate with: (i) genetic relatedness, and (ii) provisioning effort. So far, this has been investigated in only one species, the bell miner *Manorina melanophrys* (McDonald & Wright 2011). Bell miners show fine-scale facultative adjustment of effort according to brood kinship (Wright *et al.* 2010), putatively assessed using individual-specific *mew* calls (McDonald & Wright 2011). However, whether the relationship is correlative or threshold-based is unclear.

1.4.2 Kin recognition and inbreeding avoidance

There are very few studies on how kin recognition operates in the context of mate choice. However, it is extremely likely that kin recognition is important for inbreeding avoidance in cooperative populations that exhibit kin-structure (Bateson 1983). In such social systems, where the risk of inbreeding could be particularly high (Koenig & Haydock 2004), an already active kin recognition mechanism may be co-opted for inbreeding avoidance. A recent study on western bluebirds *Sialia mexicana* was the first to demonstrate active avoidance of kin during mate choice in a cooperatively breeding bird (Dickinson *et al.* 2016). Playback experiments have previously identified vocalisations as putative cues to kinship in this species (Akçay *et al.* 2013), but precisely how relatives recognise kin through vocalisations is unclear (Akçay *et al.* 2014), and whether vocal similarity influences mate choice remains to be tested.

Kin recognition mechanisms for inbreeding avoidance in non-cooperatively breeding birds have focused on odour cues (Coffin *et al*. 2011). Storm petrels *Hydrobates pelagicus* appear to prefer non-kin odours (Bonadonna & Sanz-Aguilar 2012). Odour has also been suggested as a recognition cue in zebra finches *Taenopygia* *guttata* (Caspers *et al*. 2013), although the evidence is inconclusive (Ihle & Forstmeier 2013). A recent study found that similarity in preen secretion chemicals in black-legged kittiwakes *Rissa tridactyla* is positively correlated with MHC relatedness (Leclaire *et al*. 2014), but although statistically significant, the relationship is weak. As vocal cues appear to be used in cooperative contexts, it is reasonable to assume the same cues may be used in mate choice. One study has shown that breeding splendid fairy-wren females *Malurus splendens* respond more intensely to songs of male helpers from other groups than to those in their own group (Payne *et al*. 1988).

Generally, kin recognition for inbreeding avoidance is thought to involve phenotype-matching (Pusey & Wolf 1996). If this relied on associative learning, there may be acceptance errors because of a lack of prior association early in life (Sherman *et al.* 1997), so associative learning may not be reliable enough. On the other hand, if unfamiliar individuals are consistently sufficiently unrelated then associative learning of kin may be a suitable inbreeding avoidance mechanism, and empirical evidence to support phenotype-matching as a recognition mechanism in cooperatively breeding birds is lacking (Komdeur & Hatchwell 1999).

1.4.3 Kin recognition and population structure

Reeve's (1989) acceptance threshold model predicts that the accuracy of kin recognition should depend on the fitness consequences of accepting and rejecting recipients, and the frequency with which desirable and undesirable recipients are encountered. Previous studies have shown that extent of kin discrimination is positively associated with the risk of helping non-kin (Cornwallis *et al.* 2009) and the benefits of helping, predicted by inclusive fitness theory (Griffin & West 2003; Hatchwell *et al.* 2014). However, there have been no equivalent studies in the context of inbreeding depression, even though the rate at which relatives and non-relatives associate is also crucial for predicting the strength of kin discrimination. In some species, individuals predominantly interact with kin as a by-product of spatial distribution. In others, individuals interact with kin and non-kin, and need to seek out or side-step kin based on one or more of the mechanisms discussed above. This depends on population genetic structure, governed by life-history and dispersal patterns, and varies across cooperatively breeding species.

Most populations of cooperative breeders exhibit strong kin structure, because helpers are usually retained offspring (Ekman *et al.* 2004). Cooperation therefore occurs most often within discrete family units with predictably high relatedness between interacting individuals. In such species, permissive kin recognition thresholds may be adequate (Komdeur & Hatchwell 1999). In a small number of cooperatively breeding birds, individuals associate with kin of variable relatedness across extended networks known as *kin neighbourhoods* (Dickinson & Hatchwell 2004). In such species, more sophisticated kin recognition mechanisms may be advantageous. In a meta-analysis across multiple cooperatively breeding species, Cornwallis *et al.* (2009) showed that kin discrimination was indeed greater in species in which group relatedness was low on average but highly variable. How this variation in relatedness influences inbreeding risk, and the predicted strength of kin discrimination during mate choice, remains relatively unstudied.

Differences in kin discrimination across animal groups can make generalisations difficult, but the variation from sophisticated recognition mechanisms through to indiscriminate helping is unsurprising, given the diversity of cooperative breeding systems. Understanding how demography, life-history and social mating system interact to create selection pressure for kin discrimination within species allows predictions to be made for the evolution of kin recognition systems across species.

1.5 The long-tailed tit, *Aegithalos caudatus*

The long-tailed tit is a small (13-16cm, 7-8g) passerine from the family Aegithalidae, which includes 12 other species, most of which are found in Eurasia (the exceptions are the bushtit of North America *Psaltriparus minimus* and the pygmy tit *Psaltria exilis* of Java). Long-tailed tits are sexually monomorphic, primarily insectivorous birds with a broad distribution stretching from Europe to Japan and from the Middle East to Siberia (Hatchwell 2016). Long-tailed tits are common in the United Kingdom, and spend the non-breeding season (June-March) roosting and foraging in fluid flocks of around 10-20 birds, typically comprising males and females of varying ages and kinship (McGowan *et al.* 2007; Napper & Hatchwell 2016). In early spring, flocks start to break up in to monogamous pairs, all of which attempt to breed independently within overlapping ranges. The male and female build a nest together, usually 2-3m above ground in thorny bushes such as gorse *Ulex europaeus* or bramble *Rubus fruticosus*. Each nest is an elaborate dome of moss and plant fibres bound with spider's silk, covered with lichen flakes and lined with up to 2,500 feathers (Hatchwell 2016). In late March or early April, a clutch of 9-11 eggs is laid and incubated by the females for approximately 15 days (Hatchwell *et al.* 1999). Eggs are laid daily, but females start incubating only when the clutch is complete, so nestlings hatch synchronously and are provisioned by both parents for 16-17 days; fledglings are then fed for a further two or three weeks (Cramp & Perrins 1993).

Nest success rates are low, 72% fail due to corvid or mustelid predation (Beckerman *et al.* 2011). If nest failure occurs early in the breeding season, failed breeders attempt to breed again, otherwise individuals abandon breeding and some become helpers at the nest of another pair, assisting in the provisioning of nestlings and fledglings. Due to the high nest predation rate, around 50% of successful nests receive help, usually by one or two helpers (Hatchwell *et al.* 2004). Helping is costly, but the additional care provided by helpers increases fledgling survival and allows breeders to reduce their reproductive costs, improving their probability of breeding in the future (Meade *et al.* 2010; Hatchwell *et al.* 2014). Helpers do not gain direct fitness benefits (Meade & Hatchwell 2010), but they do accrue indirect fitness by choosing to help at a nest belonging to a close relative (Russell & Hatchwell 2001) and these indirect fitness benefits outweigh the cost of helping (Hatchwell *et al.* 2014). Long-tailed tits live in kin neighbourhoods (Dickinson & Hatchwell 2004). Dispersal of both sexes is limited, and siblings tend to disperse together (Sharp *et al.* 2008a; 2008b) so kin are often spatially clustered when breeding. However, non-kin and kin of varying relatedness breed in close proximity and helpers show strong kin discrimination in their cooperative behaviour (Russell & Hatchwell 2001), recognising kin using learned vocalisations (Sharp *et al.* 2005).

Long-tailed tits are an excellent model species to study the kin discrimination rules involved in helping and pairing behaviour. Their cooperative breeding system, ecology and life-history has been studied extensively (Hatchwell 2016) and kin recognition, in particular, is perhaps better studied in long-tailed tits than in any other cooperative breeder. Secondly, as long-tailed tits are facultative cooperative breeders, individuals make an active choice on whether and who to help (Russell & Hatchwell 2001), and helpers fine-tune their provisioning effort according to the degree of relatedness to the recipient brood (Nam *et al.* 2010). The natal dispersal distances of males and females overlap (Sharp *et al.* 2008a; b) so they run the risk of mating with opposite-sex relatives. It is reasonable to assume that the recognition cues identified for helping behaviour may be used in mate choice as well, in order to avoid inbreeding.

1.6 Thesis aims and outline

The main objectives of this study were to: (i) determine the strength of kin discrimination in long-tailed tits, both in the context of helping behaviour and inbreeding avoidance; and (ii) identity the proximate mechanisms that permit sophisticated kin discrimination in both these contexts.

In this thesis, I first describe the fine-scale genetic structure within long-tailed breeding populations, and investigate how observed patterns of kin-biased helping compare to those expected randomly, based on the distribution of relatives. This allows me to quantify how effectively long-tailed tits discriminate in favour of kin when making helping decisions (Chapter 2). I then investigate whether this population structure generates an inbreeding risk, whether inbreeding is costly in long-tailed tits, and whether breeders are able to actively avoid kin during mate choice. These findings are used to identify a putative discrimination rule for incest avoidance (Chapter 3). In Chapter 4, I develop a comprehensive and accurate method of assessing vocal similarity among long-tailed tits, using various bioacoustic tools, and investigate which acoustic parameters may be particularly important for kin recognition. In Chapter 5, I examine whether an assessment of vocal similarity may be used to adjust helper effort according to precise relatedness and avoid mating with relatives, by (i) correlating vocal similarity with relatedness, and (ii) exploring whether vocal similarity influences helping and pairing decisions. Finally, I summarise the results and assess whether the same recognition mechanism may be co-opted for both indirect fitness and inbreeding avoidance in long-tailed tits, and suggest directions for future work (Chapter 6).

Chapter 2

Fine-scale genetic structure and helping decisions in long-tailed tit societies

This chapter is published as:

Leedale A. E., Sharp S. P., Simeoni M., Robinson E. J. H & Hatchwell B. J. (2018). Fine-scale genetic structure and helping decisions in a cooperatively breeding bird. *Molecular Ecology*, **27**, 1714-1726.

Summary

In animal societies, characteristic demographic and dispersal patterns may lead to genetic structuring of populations, generating the potential for kin selection to operate. However, even in genetically structured populations, social interactions may still require kin discrimination for cooperative behaviour to be directed towards relatives. Here, we use molecular genetics and long-term field data to investigate genetic structure in an adult population of long-tailed tits *Aegithalos caudatus*, a cooperative breeder in which helping occurs within extended kin networks, and relate this to patterns of helping with respect to kinship. Spatial autocorrelation analyses reveal fine-scale genetic structure within our population, such that related adults of either sex are spatially clustered following natal dispersal, with relatedness among nearby males higher than that among nearby females, as predicted by observations of male-biased philopatry. This kin structure creates opportunities for failed breeders to gain indirect fitness benefits via redirected helping, but crucially, most close neighbours of failed breeders are unrelated and help is directed towards relatives more often than expected by indiscriminate helping. These findings are consistent with the effective kin discrimination mechanism known to exist in long-tailed tits, and support models identifying kin selection as the driver of cooperation.

2.1 Introduction

The genetic structure of populations, that is the frequency and distribution of genotypes in space, is a key biological feature that influences diverse behavioural and evolutionary phenomena. Genetic structuring occurs when gene flow is limited, resulting in an increase in genetic differentiation with geographical distance, or *isolation-by-distance* (Wright 1943). Gene flow can be restricted by physical barriers (Frantz *et al.* 2010) or ecological traits (Edelaar *et al*. 2012), such as species' dispersal capacity (Watts *et al.* 2004) and migration patterns (Rolshausen *et al*. 2013). At large spatial scales, genetic structure can lead to local adaptation and speciation (Winker *et al.* 2013), while at finer scales, behavioural traits such as territoriality (Lee *et al.* 2010) or natal philopatry (Woxvold *et al.* 2006) can lead to the spatial clustering of relatives, or kin structure, within populations. This can have important implications for inbreeding (Keller & Waller 2002) and the evolution of sociality (Hamilton 1964; Bourke 2011).

The nature of genetic structuring in wild populations can reveal much about a species' ecology and the strength of selection for phenotypic traits. For instance, since flight increases dispersal capacity, gene flow in avian populations is generally assumed to be high, and genetic structure is expected to be observed at large spatial scales (van Treuran *et al*. 1999; Crochet 2000). However, in cooperatively breeding birds, dispersal patterns characteristic of social living can lead to fine-scale genetic structure, observed at the level of social groups or across territories (Emlen 1997; van Dijk *et al*. 2015). Studies combining molecular and field data can reveal how behavioural and demographic traits shape genetic structure, and predict the evolutionary consequences of such spatial-genetic distributions.

Animal societies characteristically exhibit a degree of genetic structure that can provide substantial fitness benefits to individual group members (Cornwallis *et al*. 2009; Hatchwell 2010). In cooperatively breeding animals, limited natal dispersal leads to the spatial clustering of relatives (Ekman *et al*. 2004; Heg *et al*. 2004; Clutton-Brock & Lukas 2012; Rubenstein & Abbott 2017). This genetic structure facilitates kin selection (Hamilton 1964) because to gain indirect fitness benefits from cooperation, relatedness between the donor and recipient of aid must be higher, on average, than between randomly selected individuals within the population (Queller 1994). Essentially, kin-selected helping can evolve only when individuals have the opportunity to interact with kin, so population viscosity is regarded as a necessary precursor to the evolution of kin-selected cooperative breeding (Dickinson & Hatchwell 2004; West *et al*. 2007b). As well as alloparental care, there are other contexts in which long-term kin associations may be beneficial, such as communal investment in public goods (van Dijk *et al*. 2014) or predator defence (Griesser & Ekman 2005).

On the other hand, increased levels of relatedness among neighbouring individuals can also incur fitness costs through kin competition for mates or resources (Taylor 1992; West *et al.* 2002), and an increased likelihood of incestuous matings (Koenig & Haydock 1994). The risk of inbreeding and subsequent inbreeding depression (Keller & Waller 2002) may be alleviated through sex-biased dispersal (Pusey & Wolf 1996), and in many cooperative species, philopatry by breeders is biased strongly towards one sex (Greenwood 1980; Walters *et al*. 2004; Double *et al.* 2005). However, in the absence of effective spatial separation, recognition mechanisms that permit accurate kin discrimination may be necessary to mitigate the costs associated with long-term interactions with relatives, as well as to maximise inclusive fitness benefits via cooperation (Komdeur & Hatchwell 1999). The extent of kin discrimination expected in cooperative societies depends on the benefits of accurate kin recognition, and the costs of recognition errors, which both depend heavily on the genetic structure of the breeding population (Griffin & West 2003; Cornwallis *et al*. 2009). Note that we follow Sherman *et al.* (1997) in using *kin discrimination* to refer to the differential treatment of conspecifics differing in genetic relatedness, and *kin recognition* to describe any mechanism by which this is effected.

In most avian cooperative breeders, individuals within a population are organised into discrete groups, in which relatedness among members is generally high and predictable (Cornwallis *et al*. 2009). Studies of the fine-scale genetic structure of cooperative species has therefore tended to focus on gene flow between such groups, rather than on dyadic genetic differentiation at broader spatial scales or across the population as a whole. However, for a small number of cooperative breeders, such as western bluebirds *Sialia mexicana* (Dickinson *et al*. 1996) and riflemen *Acanthisitta chloris* (Preston *et al*. 2013), cooperative social interactions extend beyond discrete nuclear or extended family units to less defined, connected networks of relatives, known as *kin neighbourhoods* (Dickinson & Hatchwell 2004). In such cases, the relatedness between socially interacting individuals is much more variable and relatively low overall (Cornwallis *et al*. 2009). The underlying genetic structure resulting from this complex social pattern remains largely unexplored, having been measured only in the bell miner *Manorina melanophrys* (Painter *et al.* 2000) and sociable weaver *Philetairus socius* (Covas *et al*. 2006).

Here, we describe the previously undefined population genetic structure underlying the social organisation of long-tailed tits *Aegithalos caudatus* and investigate how this relates to the opportunities and patterns of helping with respect to kinship. By quantifying the distribution of genetically similar individuals within populations, we can gain a deeper understanding of the opportunity for kin selection to operate in this species, and provide insights into the level of kin discrimination required for helpers to maximise inclusive fitness.

Long-tailed tits are facultative cooperative breeders, but unlike most cooperative species, adult offspring do not delay natal dispersal or breeding to help their parents to raise young (Hatchwell 2016). Instead, all adults disperse during their first winter and attempt to breed independently the following spring, and may decide to help at a relative's nest only if their own reproductive attempts fail (Russell & Hatchwell 2001). Local recruitment is male-biased, yet natal dispersal distances of both sexes are short (Sharp *et al*. 2008a) and siblings often disperse together (Sharp *et al*. 2008b). Long-tailed tits also exhibit strong kin associations during winter, which are reflected in their helping decisions the following spring (Napper & Hatchwell 2016). Another aspect of the long-tailed tit's life history that strengthens genetic structure is the pattern of offspring mortality that results in a small effective population size (Lehmann & Balloux 2007). High nest predation rates cause frequent removal of whole broods at the nestling stage, and in the following year result in high recruitment rates of close kin from the small proportion of successful nests (Beckerman *et al.* 2011). Together, these demographic and life-history patterns permit the characteristic flexibility in reproductive strategy from independent breeding to helping within an individual's lifetime. Around 40% of nests are helped, usually by one or two helpers, and helpers gain indirect fitness benefits by increasing the productivity of helped broods (Hatchwell *et al*. 2004; MacColl & Hatchwell 2004). In contrast, no significant direct fitness benefits of helping have been identified (McGowan *et al*. 2003; Meade

& Hatchwell 2010). Helpers are overwhelmingly male (Sharp *et al*. 2011) and move away from their last failed breeding attempt to redirect care (Hatchwell 2016), often skipping the nearest available nest in search of one belonging to a relative (Russell & Hatchwell 2001). Furthermore, helpers adjust their effort according to kinship, provisioning more closely related broods at higher rates (Nam *et al*. 2010). These patterns suggest that kin-biased helping is not solely a result of kin structure, but involves active discrimination among individuals. In the absence of reliable spatial cues to kinship, helpers use vocal cues learned during early development to recognise and preferentially aid close relatives (Hatchwell *et al*. 2001b; Sharp *et al*. 2005). However, the precise relationship between dispersal patterns, genetic structure and helping decisions remains unclear.

Using detailed observational and pedigree data, combined with molecular genetics, we first define the genetic structure among long-tailed tit breeders and relate this to the known patterns of dispersal for this species. We examine spatial genetic clustering both within and between sexes to determine whether this reflects female-biased dispersal and male-biased local recruitment. To assess how population structure impacts social behaviour, we also examine the distribution of close, distant and non-kin through the population to quantify the helping opportunities available to failed breeders in relation to distance. Finally, we compare the expected probability of helping kin based on population genetic structure with the observed patterns of helping to test whether kin-biased helping in long-tailed tits can be explained by this newly described genetic structure, or whether it is a result of active kin discrimination. We calculate kinship using both genetic data and a social pedigree. The measures differ because the population is open, so the social pedigree is inevitably incomplete. However, both measures are informative because although the fitness consequences of helping depend on genetic relatedness, social relatedness is the only information available to birds when making decisions, by way of socially learned recognition cues (Sharp *et al*. 2005).

2.2 Methods

2.2.1 Study area and field methods

A population of 17-72 (mean *c*.50) pairs of long-tailed tits was studied during the breeding season (February-June) between 1994 and 2016 in the Rivelin Valley, Sheffield, UK (53º38'N 1º56W). The site is approximately 2.5km^2 and comprises a variety of suitable breeding habitat; predominantly deciduous woodland (mature oak *Quercus robur*, European beech *Fagus sylvatica*, sycamore *Acer pseudoplatanus*, silver birch *Betula pendula*) and scrub (holly *Ilex aquifolium*, gorse *Ulex europaeus*, bramble *Rubus fruticosus* and hawthorn *Crataegus spp*). This is an open population, with approximately 40% of breeders hatched in the study site (A. E. Leedale & B. J. Hatchwell, unpublished data). The remaining adults are assumed to be first year breeders that emigrated from outside the study site, based on the observation that individuals have high site fidelity following their first breeding year (McGowan *et al*. 2003). Almost all individuals (>95%) were marked with a metal BTO ring and a unique combination of two plastic colour rings for field identification. Native birds were ringed as 11-day old nestlings and immigrant adults were captured in mist nests under BTO licence before or during breeding. When ringed, a sample of 5-30µl of blood was taken by brachial venepuncture under Home Office licence. All breeding attempts were closely monitored and GPS coordinates were taken for each nest (n =

1461); a Cartesian coordinate system (UTM) was used to describe geographic distance between nests.

2.2.2 Molecular analyses

Molecular markers were used to estimate genetic relatedness between individuals and define population genetic structure. Genomic DNA was extracted from blood samples and amplified. All sampled individuals were sexed using the *P2-P8* sex-typing primers (Griffiths *et al.* 1998). Individuals ringed between 1994 and 2006 were genotyped at eight microsatellite loci (Ase18; Ase37; Ase64; Hru2; Hru6; Pca3, PmaD22, Ppi2). Thereafter, individuals were genotyped at an additional nine loci (CAM01; CAM03; CAM15; CAM23; Tgu_01.040; Tgu_04.012; Tgu_05.053; Tgu_13.017; Pca4). For further details on genotyping procedures, see Simeoni *et al*. (2007) and Adams *et al*. (2015). The population allele frequencies used in all analyses were generated using all genotyped individuals (1994-2016, n = 3182) in *CERVUS version 3.0.7* (Kalinowski *et al.* 2007), to maximise accuracy in estimating the frequency of rare alleles and to ensure non-zero frequencies for all alleles in the dataset. The genetic relatedness between pairs of individuals was estimated using Queller and Goodnight's (1989) r_{QG} coefficient of relatedness in *SPAGeDi version 1.1.5* (Hardy & Vekemans 2002). This relatedness estimate has been found to be reliable when tested against our social pedigree (Nam *et al*. 2010).

2.2.3 Social pedigree

The social pedigree was created using 22 years of field observations as described above (n = 2815 birds). For further details on pedigree construction see Appendix 1. To calculate pairwise social relationships (*r*), an additive relationship matrix was generated from the pedigree in *R version 3.3.2* (R Core Team 2016) using the *R* package *nadiv* (Wolak 2012). Six breeding birds in our study population (0.2%) were from cross-fostered broods in 1996-1998, but given that birds raised together treat each other as kin (Hatchwell *et al*. 2001b; Sharp *et al*. 2005), we include them in the social pedigree. Similarly, while there is a low rate of extra-pair paternity in long-tailed tits (Hatchwell *et al.* 2002), we have not corrected for it in the pedigree.

2.2.4 Spatial analysis of genetic relatedness

To describe the overall genetic structure of our breeding population (n = 1022), we use Weir and Cockerham's (1984) inbreeding coefficient, F_{IS}, and the microsatellite allele size-based genetic differentiation estimate, R_{IS} (Slatkin 1995), as calculated in *SPAGeDi,* following the recommendation by Balloux & Lugon-Moulin (2002). To calculate the approximate standard error of genetic relatedness and differentiation estimates, multilocus estimates were jack-knifed over loci and alleles permuted among individuals 20,000 times.

To assess fine-scale genetic structure within our population, we performed spatial autocorrelation analyses of relatedness (r_{QG}) as a function of geographic distance: (i) among all individuals, (ii) among males, (iii) among females, and (iv) between males and females. The distance between breeding birds was based on the locations of their first nests in a given year. We compared the observed r_{QG} values within defined distance bands with the corresponding frequency distributions of r_{QG} when random permutations of the data were performed. The median natal dispersal distance within our study site was 393m for males and 522m for females, and the median distance travelled by established breeders between years was 312m (A. E. Leedale & B. J. Hatchwell, unpublished data). We therefore set distance intervals of 300m as a scale on which to examine population structure, from pairwise comparisons of individuals at the same nest (0m), to those 2100m away; with a final distance band (>2100m) containing pairwise comparisons of individuals from the most distant nests (mean ± SD distance between birds = 1033m ± 596; maximum distance = 3195m). These bands generated enough variation in dyadic genetic relatedness, while maintaining a large enough sample size at each distance interval to ensure meaningful analyses. Long-tailed tits are relatively short-lived, so the genetic structure of the population may fluctuate over time due to demographic processes such as dispersal, mortality and recruitment (Balloux & Lugon-Moulin 2002; Hatchwell *et al.* 2013). Therefore, although all genetic and spatial analyses were performed on the long-term dataset, these were restricted to within-year comparisons among individuals.

Spatial autocorrelation among all individuals, among males and among females were analysed separately using *SPAGeDi*. Individual locations were permuted 20,000 times for tests on each distance band and all tests were two-tailed. *SPAGeDi* cannot be used to perform spatial autocorrelation analyses both within-years and between opposite-sex pairs of individuals only, so to analyse male-female genetic structure, an equivalent randomisation procedure was conducted using *R*. Mean observed r_{QG} estimates were calculated for each defined distance band. The observed r_{QG} values within each band were replaced with an equal-sized sample of r_{QG} values selected at random across all distance bands. From these values, the mean null r_{QG} was calculated, and permuted 20,000 times to simulate the random frequency distribution. In all analyses, the observed r_{QG} in each distance band was considered statistically significant if the mean fell outside the 95% confidence interval of the random distribution.

The genetic structure of males and females was compared using mixed effects models (GLMM). The typical natal dispersal distance was <400m for philopatric male recruits, and <600m for female recruits (Sharp *et al*. 2008a; see above), and neither sex exhibited significant kin structure beyond 600m (see Results). Therefore, we compared relatedness among males with that observed among females within two distance bands (0-300m and 300-600m) only. For each distance band, r_{QG} values were modelled with sex fitted as a fixed effect and the IDs of both birds fitted as random effects.

2.2.5 Helping decisions

The range within which failed breeders search for helping opportunities is likely to interact with population genetic structure to affect the probability of helping kin. Using both genetic data (n = 1022) and the social pedigree (n = 866), we calculated the frequency of first order kin (pedigree $r \geq 0.5$; relatedness coefficient $r_{QG} > 0.25$), second order kin ($0.5 > r \geq 0.25$; $0.25 \geq r_{QG} \geq 0.125$) and non-kin ($r < 0.25$; $r_{QG} < 0.125$) in the breeding population (Appendix 2). Only relationships between breeders that were present in the same year were considered. Helping distance was calculated as the distance between helpers' last failed breeding attempt in a given year and the nest at which they first appeared as a helper the same year. Distance between nests was measured in UTM coordinates and was calculated in the R package, *raster* version 2.5-8 (Hijmans 2016).

To assess the impact of kin structure on helping behaviour, we quantified the proportion of cases in which help was given to broods belonging to at least one first order kin, at least one second order kin, or two unrelated breeders over the three distance bands in which helping is likely to occur: 0-300m, 300-600m and 600-900m.

We used Pearson's Chi-squared tests to determine whether the proportion of helpers assisting kin was affected by distance, and whether the proportion of helpers assisting kin differed between males and females. We also calculated the probability that helpers would provide care to broods belonging to kin if they helped a random brood within 0-300m, 300-600m and 600-900m. For each helping event, a nest was selected at random from the pool of nests the focal helper could have chosen. This pool contained all nests present in the year the helping event occurred within the same distance band as the chosen nest. The proportion of cases in which help was given to nests belonging to at least one first order kin, at least one second order kin or two unrelated breeders was calculated based on this random sample. The procedure was repeated 10,000 times to generate a distribution of expected proportions for each distance band if nests were selected randomly with respect to kinship.

To determine the effect of distance and helper sex on the probability of helping kin, we carried out Pearson's Chi-squared tests using the mean of the randomly generated proportions. Finally, we compared the expected probability of helping at a nest belonging to a relative to the observed proportion of birds helping kin. The observed proportion of helped nests within each distance band in which the helper was a first order, second order or non-relative was considered statistically significant if it fell outside the 95% confidence interval of the randomly generated distribution. This allowed us to determine whether kin structuring alone could explain observed patterns of helping behaviour, and if not, the degree of discrimination required to direct care towards kin. All analyses were carried out on all helpers and separately by sex, using both genetic data and the social pedigree.

2.3 Results

Breeding adults were genotyped at 8 (n = 525) or 17 (n = 497) polymorphic microsatellite loci (multilocus averages used, mean number of alleles per locus = 15, effective alleles = 7.09 (Nielsen *et al.* 2003), allelic richness = 6.88, gene diversity corrected for sample size = 0.761 and individual inbreeding coefficient F_i = 0.007). In total, 264 alleles were detected (for the distribution of alleles among loci, see Appendix 3, Table A3.1). The average genetic variation among breeders was not significantly different from random, based on the population inbreeding coefficient (F_{IS} = -0.002 ± 0.004, p = 0.67) and microsatellite-specific genetic differentiation estimate (R_{IS} = -0.014 ± 0.027, p = 0.36), indicating no significant inbreeding or outbreeding in our study population.

2.3.1 Spatial analysis of genetic relatedness

Mean ± SE population-level relatedness was 0.011 ± 0.003 among all birds (1719 observations of 1022 birds, 73069 within-year comparisons), 0.012 ± 0.004 among males (909 observation of 529 males, 20279 comparisons), 0.012 ± 0.005 among females (810 observation of 493 females, 16041 comparisons) and 0.009 between males and females (1719 observations of 1022 birds, 36749 comparisons). The standard error of relatedness estimates was not quantifiable for male-female genetic structure across years (see Methods), but for within-year estimates of mean relatedness between males and females with standard errors from jack-knifing over loci see Appendix 4, Table A4.1.

Spatial analyses revealed fine-scale genetic structure within our breeding population, with nearby individuals being the most genetically similar (Fig. 2.1). Pairwise relatedness among all individuals was higher than expected (based on permuted pairwise relatedness) within a radius of 300m and between 300m and 600m (both $p < 0.001$, Fig. 2.1a). Within each sex, relatedness was higher than expected within 300m and 300-600m for males (both $p < 0.001$, Fig. 2.1b) and females (both $p < 0.001$, Fig. 2.1c). Although slightly lower than the within-sex comparisons, dyadic relatedness between males and females was also significantly higher than expected by chance within a 300m and 300-600m (both $p < 0.001$, Fig. 2.1d). Within-year comparisons between males and females exhibit a similar spatial pattern (Appendix 4, Fig. A4.1). The distance at which kin structure breaks down in all comparisons is beyond 600m (Fig. 2.1). Within this distance, relatedness among males was significantly higher than relatedness among females at 0-300m (GLMM: $F = 20.63$, $df = 1,780$, $p < 0.001$), but not 300-600m (GLMM: $F = 2.29$, $df = 1,888$, $p = 0.13$).

2.3.2 Helping decisions

The median distance travelled by failed breeders to help another breeding pair was 263m for males ($n = 164$) and 346m for females ($n = 37$), with most helpers travelling within the 0-300m and 300-600m distance bands (Fig. 2.2). Using genetic estimates of relatedness, r_{QG}, mean ± SD relatedness of all helpers to recipient breeding pairs was 0.14 ± 0.16 (95% CI = -0.10-0.36). In the majority of cases ($n = 181$), helpers assisted at least one first or second order relative ($r = 0.5$, 56.9%; $r = 0.25$, 13.3%), but a substantial minority of helpers were unrelated to the recipients ($r = 0$, 29.8%). Estimated helper-recipient relatedness was lower using the social pedigree, although a majority of helped nests were again helped by at least one first or second order relative ($r = 0.5$, 39.4%; $r = 0.25$, 14.2%; $r = 0$, 46.5%; $n = 150$ cases).

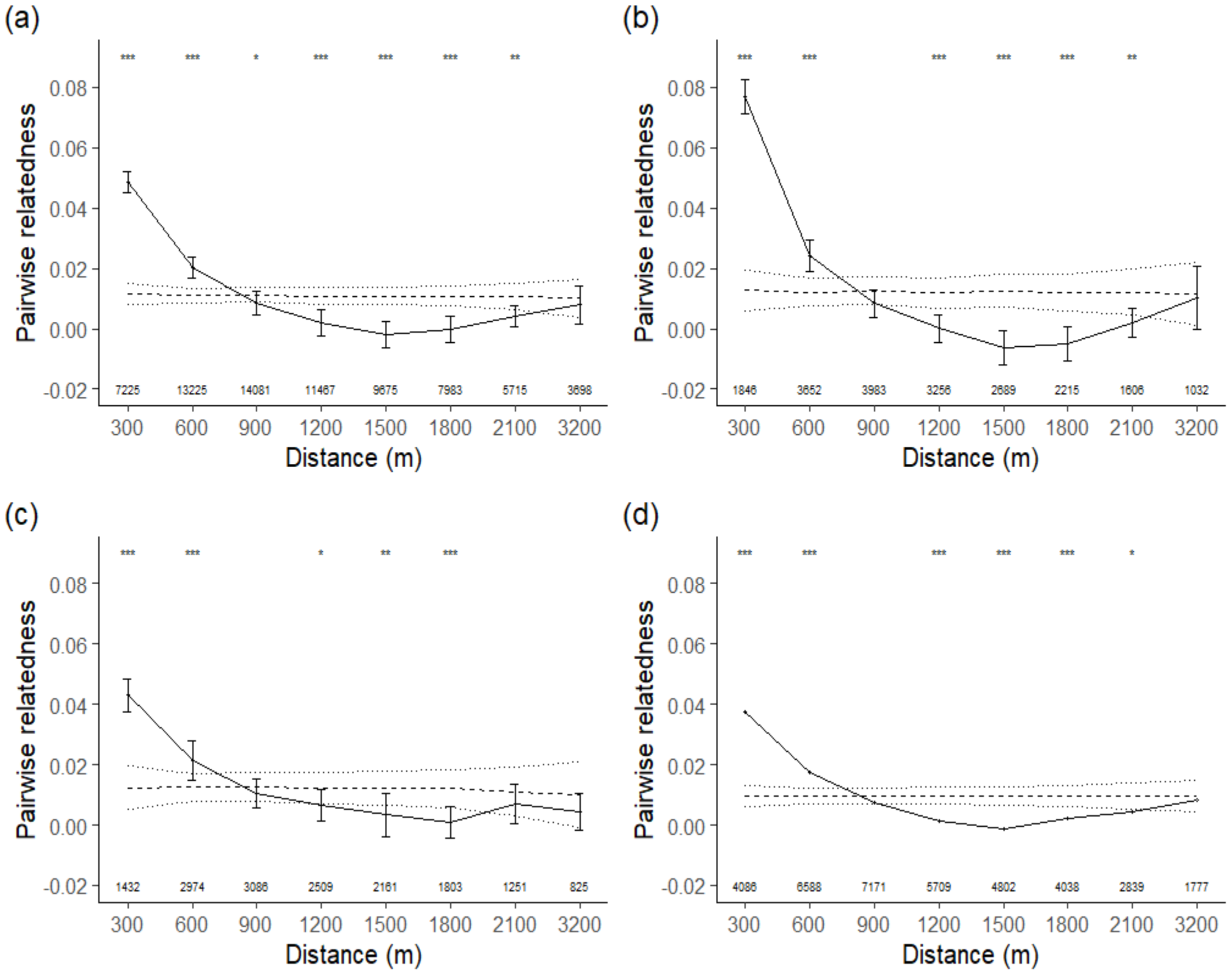

Figure 2.1. Mean pairwise relatedness (r_{QG}) in a long-tailed tit population over eight bands of distance between dyads: (a) among all individuals, (b) among males, (c) among females, (d) between males and females. Dashed lines indicate the simulated null mean r_{QG} and 95% CI in an unstructured population. Error bars approximate SE estimates from jackknifing over loci. Numbers above the x axis represent the number of pairwise comparisons. *$p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

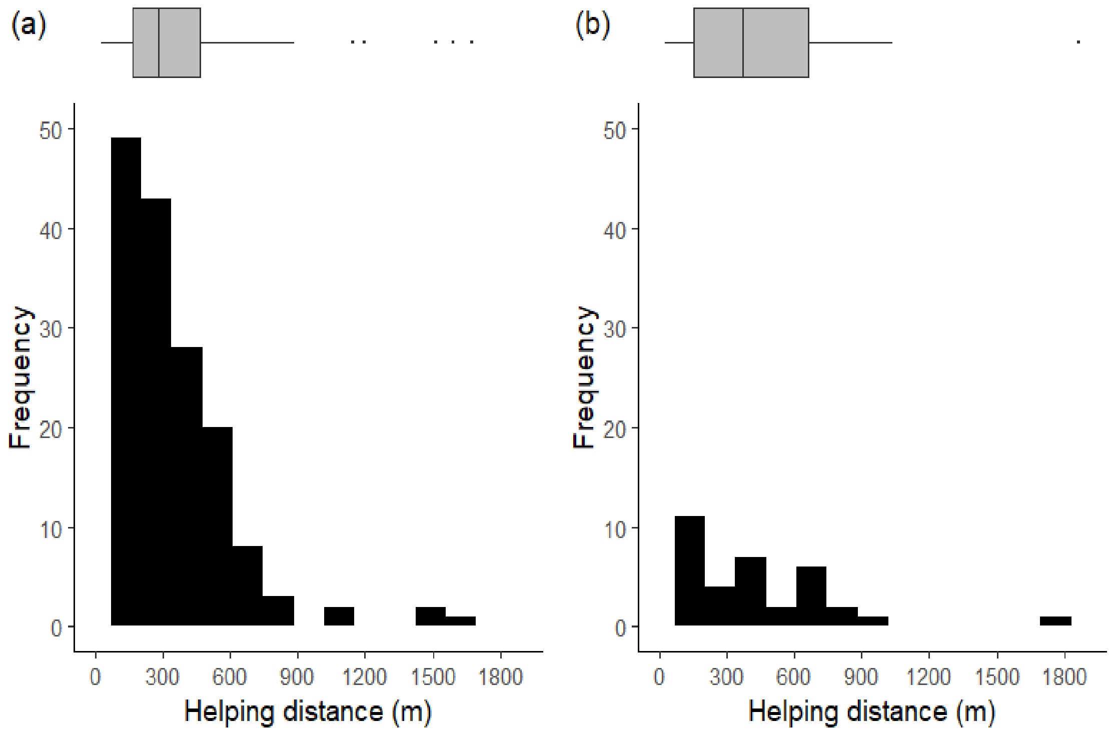

Figure 2.2. Frequency distribution and median (+IQR) helping ranges of (a) male (n = 164) and (b) female (n = 37) long-tailed tit helpers, calculated as the distance between an individual's first helped nest and their last reproductive attempt that year.

Overall, we found little effect of the distance travelled by helpers on their probability of helping kin. Using genetic estimates of relatedness, helpers were marginally less likely to help relatives when travelling between 300-600m ($\chi^2 = 10.24$, df = 4, $p < 0.05$, n = 177) compared to the shorter and longer distance bands (Fig. 2.3a, Table 2.1). However, this effect was not significant when considering only male helpers, ($\chi^2 = 8.92$, df = 4, p = 0.06, n = 144); there were too few female helpers falling into each category to conduct an equivalent analysis (Table 2.1). Based on the social pedigree, there was no effect of distance on the proportion of first order, second order or non-kin helped when analysing all helpers ($\chi^2 = 5.88$, df = 4, p = 0.22, n = 155 cases; Fig. 2.3b, Table 2.1), or just male helpers ($\chi^2 = 3.49$, df = 4, p = 0.48, n = 129; Table 2.1). Note that again there were too few data to analyse female helpers separately (Table 2.1).

Comparing the relatedness of male and female helpers to the recipients of their care, overall males tended to help kin more often than females (Table 2.1). This was significant using the genetic data (males: $r = 0.5$, 61.8%; $r = 0.25$ 13.9%; $r = 0$, 24.3%; n = 144 cases; females: $r = 0.5$, 42.4%; $r = 0.25$ 12.1%; $r = 0$, 45.5%; n = 33 cases; $\chi^2 = 6.05$, df = 2, $p < 0.05$), but not with the social pedigree (males: $r = 0.5$, 42.6%; $r = 0.25$, 13.2%; $r = 0$, 44.2%; n = 129 cases; females: $r = 0.5$, 23.1%; $r = 0.25$ 19.2%; $r = 0$, 57.7%; n = 26 cases; $\chi^2 = 3.51$, df = 2, p = 0.17). This sex difference in the probability of helping kin was driven by a relatively small number of unrelated female helpers in the 300-600m distance band (Table 2.1).

The randomisation tests that we conducted to determine the random probability of helping a relative showed, not surprisingly given the kin structure of our population, that the random probability of helping first-order kin decreased with distance for all helpers (genetic data: $\chi2 = 13.2$, df = 2, $p < 0.01$; social pedigree: $\chi2 = 28.5$, df = 2, $p < 0.001$; Fig. 2.3, Table 2.1), for male helpers (genetic data: $\chi2 = 16.1$, df = 2, $p < 0.01$; social pedigree: $\chi2 = 28.1$, df = 2 $p < 0.001$; Table 2.1), and female helpers (genetic data: $\chi2 = 7.4$, df = 2, $p < 0.001$; social pedigree: $\chi2 = 32.5$, df = 2, $p < 0.001$; Table 2.1). There was no significant difference in the opportunity to help a relative between male and female helpers travelling within 0-300m (genetic data: $\chi2 = 2.41$, df = 2, $p = 0.29$; social pedigree: $\chi2 = 0.88$, df = 2, $p = 0.69$; Table 2.1), 300-600m (genetic data: $\chi2 = 1.7$, df = 2, $p = 0.45$; social pedigree: $\chi2 = 2.01$, df = 2, $p = 0.42$; Table 2.1) or 600-900m (genetic data: $\chi2 = 0.31$, df = 2, $p = 0.89$; social pedigree: $\chi2 = 1.33$, df = 2, $p = 0.56$; Table 2.1) to provide help.

Crucially, first-order kin were helped significantly more often than expected at random irrespective of distance travelled to provide help (Fig. 2.3; Table 2.1). This kin-bias in helping was also apparent among male helpers within all distance bands, and also among female helpers for all but one comparison (Table 2.1).

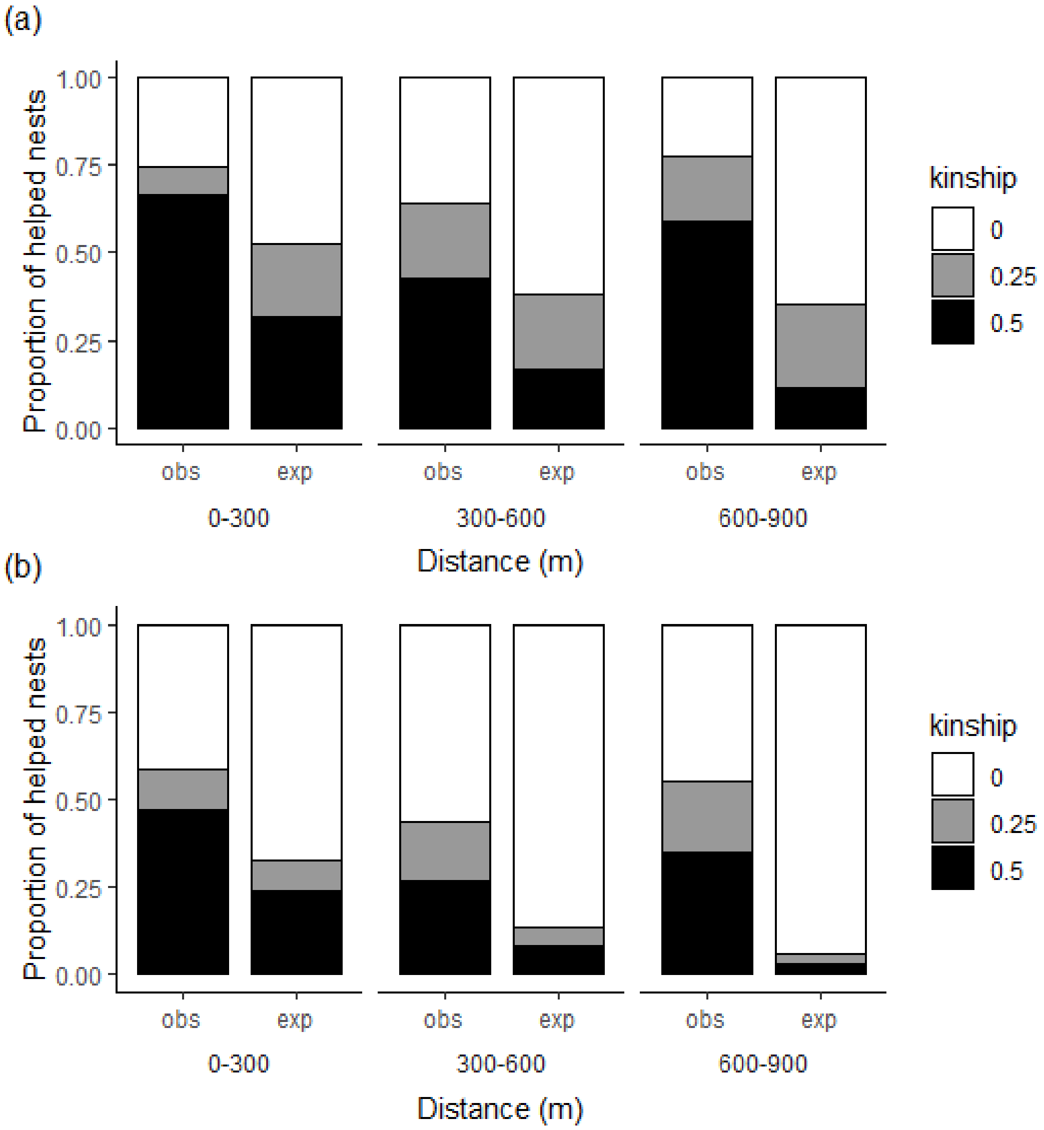

Figure 2.3. Proportion of help given to nests belonging to at least one 1st order kin (black), at least one 2nd order kin (grey), or two non-kin (white) over three bands of distance between helpers and recipients (obs). The respective proportions expected if help was given randomly within that range is also shown (exp). Relatedness between helpers and recipients is estimated using both (a) genetic data and (b) the social pedigree.

Table 2.1. Proportion of help given to nests belonging to at least one 1st order kin, at least one 2nd order kin, or two non-kin over three distance bands between helpers and recipients. Within each band, the proportion of kin helped is compared with that expected if help was given randomly within that range, based on 10000 permutations of potential nests for focal helpers within years (1994-2016). Relatedness between helpers and recipients is estimated using both genetic data and the social pedigree. Analyses were carried out on all helpers, male helpers and female helpers. The observed proportion was considered statistically significant it fell outside the 95% ($p < 0.05$), 99% ($p < 0.01$) or 99.9% ($p < 0.001$) confidence interval of the random distribution.

data	**sex**	**distance (m)**	**kinship**	**helped nests**	**observed proportion**	**potential nests**	**expected proportion (mean ± SD)**	**p**
genetic	both	0-300	0.5	66	0.67	184	0.31 ± 0.03	< 0.001
			0.25	8	0.08	127	0.21 ± 0.02	< 0.001
			0	25	0.25	296	0.47 ± 0.03	< 0.001
		300-600	0.5	24	0.43	182	0.17 ± 0.02	< 0.001
			0.25	12	0.21	281	0.21 ± 0.03	NS
			0	20	0.36	831	0.62 ± 0.03	< 0.001
		600-900	0.5	13	0.59	138	0.12 ± 0.02	< 0.001
			0.25	4	0.18	325	0.24 ± 0.03	NS
			0	5	0.23	921	0.65 ± 0.03	< 0.001
	male	0-300	0.5	58	0.69	162	0.33 ± 0.03	< 0.001
			0.25	6	0.07	105	0.21 ± 0.03	< 0.001
			0	20	0.24	244	0.45 ± 0.03	< 0.001
		300-600	0.5	22	0.49	161	0.18 ± 0.03	< 0.001
			0.25	11	0.24	245	0.22 ± 0.03	NS
			0	12	0.27	700	0.60 ± 0.03	< 0.001
		600-900	0.5	9	0.60	109	0.11 ± 0.02	< 0.001
			0.25	3	0.20	262	0.24 ± 0.03	NS
			0	3	0.20	767	0.64 ± 0.03	< 0.001
	female	0-300	0.5	8	0.53	22	0.24 ± 0.05	< 0.001
			0.25	2	0.13	22	0.21 ± 0.05	NS
			0	5	0.33	52	0.55 ± 0.04	< 0.001
		300-600	0.5	2	0.18	21	0.13 ± 0.04	NS
			0.25	1	0.09	36	0.18 ± 0.06	NS
			0	8	0.73	131	0.68 ± 0.06	NS
		600-900	0.5	4	0.57	29	0.12 ± 0.05	< 0.001
			0.25	1	0.14	63	0.21 ± 0.06	NS
			0	2	0.29	154	0.67 ± 0.07	< 0.001
pedigree	both	0-300	0.5	41	0.47	111	0.24 ± 0.03	< 0.001
			0.25	10	0.11	35	0.09 ± 0.01	NS
			0	36	0.41	316	0.67 ± 0.03	< 0.001
		300-600	0.5	13	0.27	65	0.08 ± 0.02	< 0.001
			0.25	8	0.17	47	0.05 ± 0.01	< 0.001
			0	27	0.56	827	0.86 ± 0.02	< 0.001
		600-900	0.5	7	0.35	23	0.03 ± 0.01	< 0.001

		0.25	4	0.20	27	0.03 ± 0.03	< 0.001
		0	9	0.45	960	0.94 ± 0.02	< 0.001
male	0-300	0.5	36	0.49	101	0.24 ± 0.03	< 0.001
		0.25	8	0.11	31	0.08 ± 0.02	< 0.05
		0	29	0.39	261	0.67 ± 0.03	< 0.001
	300-600	0.5	13	0.32	60	0.09 ± 0.02	< 0.001
		0.25	7	0.17	40	0.06 ± 0.02	< 0.001
		0	21	0.51	717	0.86 ± 0.02	< 0.001
	600-900	0.5	6	0.40	20	0.03 ± 0.01	< 0.001
		0.25	2	0.13	16	0.02 ± 0.01	< 0.001
		0	7	0.47	803	0.95 ± 0.02	< 0.001
female	0-300	0.5	5	0.36	10	0.22 ± 0.05	< 0.01
		0.25	2	0.14	4	0.12 ± 0.03	NS
		0	7	0.50	55	0.66 ± 0.05	< 0.01
	300-600	0.5	0	0.00	5	0.04 ± 0.03	NS
		0.25	1	0.14	7	0.06 ± 0.03	< 0.05
		0	6	0.85	110	0.90 ± 0.05	NS
	600-900	0.5	1	0.20	7	0.03 ± 0.03	< 0.001
		0.25	2	0.40	7	0.05 ± 0.03	< 0.001
		0	2	0.40	157	0.92 ± 0.05	< 0.001

2.4 Discussion

Long-tailed tits breed cooperatively in diffuse family structures, or kin neighbourhoods, such that interactions among non-kin and kin of varying relatedness are frequent. We used a combination of long-term field observations and population genetic analyses to investigate genetic structure and patterns of helping in this atypical social system. Our results showed significant, fine-scale genetic structure in long-tailed tit populations, with positive spatial autocorrelation of dyadic relatedness estimates among breeding birds. Crucially, this pattern exists after natal dispersal, and was strong among males, among females and between the sexes. These findings contrast with most studies of cooperatively breeding birds that have measured spatial-genetic autocorrelation in breeders of both sexes, and show genetic structure in adult males only due to complete female-biased dispersal (Painter *et al*. 2000; Double *et al*. 2005; Temple *et al*. 2006; Woxvold *et al*. 2006), although fine-scale genetic structure among both sexes has also been demonstrated in sociable weaver colonies (van Dijk *et al*. 2015). We found that genetic structure was stronger in males than females, but both males and females remained spatially associated with same-sex relatives once they started breeding. This reflects previously reported patterns of dispersal in this species: although females disperse further than males (Sharp *et al*. 2008a), some adults of each sex disperse only short distances from their natal area to become independent breeders. This underlying genetic structure is also consistent with known patterns of social association during the non-breeding season (Napper & Hatchwell 2016).

This degree of kin structure post-dispersal creates opportunities for failed breeders to gain indirect fitness benefits via redirected helping (Hatchwell *et al*. 2014). Although long-tailed tit societies are not organised into discrete family units of close kin, neither are related individuals distributed randomly in space, but organised into kin neighbourhoods, allowing kin selection to operate. Pairwise relatedness is highest within 300m, and males typically seek helping opportunities within this range. In the rarer instance that females become helpers, they tend to travel slightly further, but still within the range of kin clustering. The higher than expected relatedness among individuals living in close proximity is driven by the tendency of close kin to cluster together, but is low overall because non-relatives or distant kin also breed in the vicinity. In fact, genetic estimates showed that only 12% of dyadic relationships within 300m were between first order kin and as the distance between dyads increased, the proportion of kinships decreased so that only 3% of dyads over 900m were close kin (Appendix 2, Fig. A2.1). Interestingly, using genetic relatedness estimates the proportion of second order kin did not decrease with distance.

According to the social pedigree, however, the proportions of first and second order kinships both decreased over distance and were substantially lower overall than the estimated kinships using genetic data (Appendix 2, Fig. A2.1). Our genetic data may detect more kin relationships than our social pedigree for several reasons. First, our population is open, with over half the breeding adults dispersing into the study site to breed, and their parentage is unknown. Therefore, although we can use genetic data to inform the pedigree for a proportion of immigrants (see Appendix 1), some kin relationships are likely to go undetected. Second, long-tailed tits are relatively short-lived and high nest predation rates generate a low effective population size (Beckerman *et al*. 2011). Therefore, even for birds born in the study site, it is rarely possible to trace their social pedigree further than one generation without using genetic data to fill in the gaps. Third, a small number of individuals may breed just outside the study site in their first year, before moving into the site in subsequent years, and so kin relationships may exist among some immigrants across years. Kinship estimates from the social pedigree may therefore under-estimate the proportion of kin. Finally, the number of individuals in each relatedness category obviously depends on how those categories are defined using either pedigree or genetic data.

Overall, where they differ from the social pedigree, genetic relatedness estimates may be more reliable. However, pedigree data is essential for understanding how accurately individuals are able to recognise kin, particularly when the mechanism depends on socially learned cues (Sharp *et al*. 2005). Kin recognition via associative learning is likely to be effective where there is a reliable correlation between genetic relatedness and association during a sensitive phase of development, a requirement probably satisfied in most species demonstrating parental care (Komdeur & Hatchwell 1999). Associative learning is therefore considered the most widespread mechanism of kin recognition in cooperatively breeding birds, where extended brood care at the nest provides a period of association among relatives, and has been identified in many species, including Galápagos mockingbirds *Nesomimus parvulus* (Curry 1998; Curry & Grant 1990), Seychelles warblers *Acrocephalus sechellensis* (Richardson *et al*. 2003; Komdeur *et al*. 2004) and splendid fairy wrens *Malurus splendens* (Payne *et al.* 1988). On the other hand, in some species, such as stripe-backed wrens *Campylorhynchus nuchalis* (Price 1998; 1999) and green woodhoopoes *Phoeniculus purpureus* (Radford, 2005), kin recognition has been suggested to have a genetic component, and studies on bell miners (Wright *et al*. 2010) and long-tailed tits (Nam *et al*. 2010) found significant effects of genetic relatedness on helper effort, indicating a mechanism more sophisticated than associative learning alone. Further studies of phenotypic cues in relation to genetic similarity and social information are necessary to determine the precise mechanisms of kin recognition in such species.

The social organisation of long-tailed tits gives helpers a choice of whom to aid in a situation where simple decision rules based on spatial cues are insufficient, even at close proximity, for effective kin discrimination. Here, we have shown that helpers assist far more first order kin than expected if help were given randomly with respect to kinship. This is consistent with the finding of Russell and Hatchwell (2001) that helpers exhibit a kin preference in their choice of brood to help. Although most helpers choose to help kin within the range of kin clustering, improving their chances of encountering a relative, 67% helped at a nest belonging to at least one first order kin, compared with a 31% probability of encountering a first order relative within 300m (Table 2.1). Furthermore, the proportion of helpers that assisted kin did not, in general, decrease with distance; the marginally lower probability of helping first order kin at 300-600m was driven by a small number of female helpers within this range that helped non-kin. Therefore, even when fewer kin were available, helpers still identified a similar proportion of first order kin to help. This degree of kin discrimination is unusual in cooperative breeders; helping more often occurs indiscriminately within closely related family groups (Cornwallis *et al*. 2009). This is because cooperation typically precedes dispersal, with offspring staying within natal territories to help their parents raise young (Ligon & Stacey 1991). Such viscous populations exhibit strong kin structure and the probability of helping kin is predictably high within the nuclear family unit. In a minority of cooperative breeders, helping behaviour occurs post-dispersal and extends beyond the confines of a stable group (Ligon & Burt 2004; Hatchwell 2009) within extended social networks such as neighbourhoods of western bluebirds (Dickinson *et al.* 1996) and long-tailed tits (Hatchwell 2016), clans of white-fronted bee-eaters *Merops bullockoides* (Emlen & Wrege 1992) or coteries of bell miners (Clarke & Fitz-Gerald 1994). A crucial characteristic of such social structures is that the proportion of kin available to helpers is relatively low, and relatedness among individuals is too variable to favour indiscriminate cooperation (Cornwallis *et al.* 2009). In such populations, helpers must exercise a degree of discrimination to reliably direct help towards kin, so it is no coincidence that observational and experimental studies of these species have revealed some of the strongest evidence for active kin discrimination in social birds (Emlen & Wrege 1988; Hatchwell *et al*. 2001b; Russell & Hatchwell 2001; Sharp *et al*. 2005; McDonald & Wright 2011; Akçay *et al*. 2013; Dickinson *et al*. 2016).

We found that second order kin were helped much less frequently than first order kin, as reported by Nam *et al*. (2010). It is interesting that the probability of helping second order kin did not differ significantly from that expected by random helping, except within 300m, where it was lower than expected. We suggest two possible reasons for this. First, a kin recognition mechanism that depends on association during early development (Sharp *et al*. 2005) may result in reliable discrimination of first order kin (i.e. siblings and parents), but would offer less opportunity to learn more distant kin, apart from any helpers. Alternatively, birds may be able to recognise second order kin, but prefer to help first order kin to maximise fitness. Previous studies have shown that long-tailed tits modify provisioning effort contingent on kinship (Nam *et al*. 2010), indicative of a capacity to discriminate between kin of varying relatedness. However, a minority of helpers also assist non-kin, and the probability of doing so did not change with distance and did not reflect kin availability. We have detected no direct benefits of helping in long-tailed tits (Meade & Hatchwell 2010), so such help for non-kin may be due to recognition 'errors' caused by a permissive threshold for acceptance of kin (Downs & Ratnieks 2000; Hatchwell *et al*. 2014). In cases where the benefits of helping a relative greatly exceed the costs of helping a non-relative, selection should favour a recognition mechanism that reflects these relative costs despite the potential for recognition errors (Reeve 1989). More studies are required that focus on the precise mechanism of kin recognition to investigate this possibility further.

Within the range that most helpers travelled to help, there was no difference between the sexes in the opportunity to help kin. Thus, the fact that females comprise only 15% of helpers in this population (Hatchwell *et al*. 2004) is not because they have few opportunities to help. This result is consistent with a previous study of another population by Sharp *et al*. (2011), showing that when males and females exhibited similar levels of philopatry, there was still a strong male bias in helping. Therefore, the male bias in helping behaviour is not due to male philopatry, and hence helping opportunities, *per se*. Instead, the decision to help by failed breeders is probably related to individual condition; failed breeders that choose to become helpers tend to be in better condition than those that do not (Meade & Hatchwell 2010). Since egg-laying, incubating and brooding are performed exclusively by females (Hatchwell 2016), the fitness costs associated with parental care may be more pronounced in females than in males, reducing their likelihood of becoming helpers when breeding attempts fail. We also found an intriguing trend for female helpers to assist fewer close kin and more non-kin than male helpers did. However, we treat this result with some caution because the sample of female helpers available for comparison is still relatively small.

The spatial clustering of relatives also has important consequences in terms of inbreeding risk. The population genetic structure revealed by our study shows that the average relatedness between males and females is higher than expected at random within approximately 700m. Therefore, in addition to the cooperative benefits, prolonged kin-associations between males and females at the reproductive stage may carry fitness costs associated with incest and inbreeding depression (Keller & Waller 2002). Whether these kin associations persist over the range that mate choice occurs and whether kin are actively avoided during mate choice, remains to be investigated.

In conclusion, we have revealed fine-scale genetic structure in our long-tailed tit population that is consistent with the kin clustering expected from known demographic and dispersal patterns. This provides an opportunity for kin selection to operate, and, in the absence of any other known benefits of interacting socially with kin (Napper *et al*. 2013; Napper & Hatchwell 2016), is likely to have emerged as a result of selection for kin-directed helping behaviour. However, despite this genetic structure, the random probability of helping at the nest of a relative is still relatively low, demonstrating that in the kin-selected cooperative breeding system of long-tailed tits, active discrimination of kin from non-kin is required when choosing at which nest to help in order to maximise inclusive fitness. The learned vocal cues used by long-tailed tits to recognise kin provides one such mechanism (Sharp *et al*. 2005), although the degree to which they can discriminate among kin of variable relatedness requires further study. Overall, this study demonstrates how population genetic structure, generated by dispersal and life-history traits, determines the opportunity for interactions among relatives, and the selective pressure these exert on kin discrimination rules in social animals.

2.5 Data accessibility

Microsatellite genotypes, breeding locations and the social pedigree (1994–2016) are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.0dm8mv1.

Chapter 3

Inbreeding in long-tailed tits: costs, risk and avoidance

Summary

Sex-biased dispersal can alleviate inbreeding risk in natural populations when inbreeding is costly. But, in social animals, selection for kin associations during adulthood can lead to opposite-sexed relatives remaining spatially clustered. Using a combination of long-term field data and molecular genetics, I investigate the costs, risk and avoidance of inbreeding in the long-tailed tit *Aegithalos caudatus*, a cooperative breeder whose dispersal patterns are non-random with respect to kinship. Heterozygosity-fitness correlations reveal a positive association between heterozygosity and direct fitness, which indicates that inbreeding is costly. Furthermore, opposite-sex kin remain clustered within the range that mate choice occurs post-dispersal, generating a substantial inbreeding risk. However, this population is not inbred and incest is extremely rare. To determine whether kin are actively avoided as partners, I compare relatedness within breeding pairs to that expected under multiple mate choice models and offer a decision rule for pairing behaviour with regard to relatedness. The relatedness of breeding pairs is much lower than that expected if mates were selected randomly with respect to kinship. The results show that long-tailed tits effectively avoid first-order, but not second-order kin as partners during mate choice.

3.1 Introduction

Inbreeding, defined as reproduction among relatives, is generally considered maladaptive in wild populations (Pusey & Wolf 1996). Relatives share alleles that are identical by descent, or, *ibd* (Wright 1949). The probable proportion of *ibd* alleles between two individuals is their coefficient of relatedness, r (Wright 1922), and the inbreeding coefficient, f, is the probability that homologous alleles within individuals are *ibd* (Wright 1965). Both r and f are measured relative to a reference population (Keller & Waller 2002). Inbreeding commonly refers to mating between first ($r = 0.5$) or second-order ($r = 0.25$) kin, although any individual with $f > 0$ may be considered inbred (Nichols 2017).

Inbreeding increases the frequency of homozygotes in a population (Keller & Waller 2002), leading to inbreeding depression (Pusey & Wolf 1996). Inbreeding unmasks deleterious recessive alleles, which, when expressed, result in traits that reduce fitness (Charlesworth & Charlesworth 1999). The loss of heterozygous advantage (overdominance) may also reduce the fitness of inbred offspring (Bateson 1983), although, recessive deleterious alleles seem to be the most important source of inbreeding depression (Roff 2002; Charlesworth & Willis 2009). Inbreeding depression has multiple proximate causes, including increased parasite load (Coltman *et al*. 1999, Smallbone 2016) and susceptibility to disease (Townsend 2009); reduced birth weight (Walling *et al.* 2011), survival (Kruuk *et al.* 2002) and reproductive success (Slate *et al.* 2000; Szulkin *et al.* 2007); depressed immune system (Reid 2007); and negative effects on secondary sexual traits, such as song repertoire size in song sparrows *Melospiza melodia* (Reid 2005). Inbreeding depression can also have trans-generational effects: offspring of inbred mothers may be smaller (Bérénos *et al*. 2016) or less likely to survive to recruitment (Huisman *et al*. 2016). Consequently, inbred populations are at an increased risk of extinction (Frankham 1995; Saccheri *et al.* 1998; Wright *et al*. 2008). In habitually inbred populations, an increased expression of harmful recessives exposes these alleles to selection, allowing them to be purged from the genome (Keller & Waller 2002), which may explain why some studies have found no evidence of inbreeding depression, despite significant inbreeding (Hoogland 1992; Duarte *et al.* 2003). Yet, purging may only be effective in small, closed populations (Charlesworth & Charlesworth 1999; Kennedy *et al.* 2014). Inbreeding can also lower the genetic variation among offspring, and the probability that at least some will survive in stochastic environments (Bateson 1983). Furthermore, if mating with a relative is at the expense of mating with a non-relative, the transmission advantage of shared alleles is reduced (Waser 1986).

Although the adverse effects of inbreeding are well documented, outbreeding can carry its own fitness costs, as co-adapted gene complexes required for local adaptation may be broken up by recombination (Bateson 1982; 1983). Inbreeding can also be adaptive, through increased relatedness to offspring (Parker 2006), and increased reproductive success of relatives (Kokko & Ots 2006). Early studies on naked mole rats *Heterocephalus glaber* suggested that inbreeding was a major driving force in the evolution of eusociality in this species (Reeve *et al.* 1990; Lacey & Sherman 1991). However, recent work indicates that the low genetic variation observed in the populations studied reflects a founder event, rather than a consequence of their unusual mating system (Ingram *et al.* 2015), and inbreeding is avoided when possible (Clarke & Faulkes 1999). Current theoretical mate choice models predict a preference for distant kin as mates under a range of inbreeding depression strengths, suggesting that individuals can maximise inclusive fitness through so called *optimal* *kinbreeding* (Puurtinen 2011). The selection pressures on the alternative inbreeding strategies of avoidance, preference or tolerance (random mating) depend on the level of inbreeding depression, the fitness benefits of inbreeding, and the costs of inbreeding avoidance (Lehmann & Perrin 2003; Jamieson *et al.* 2009; Szulkin *et al*. 2013; Duthie & Reid 2015). Costs to active mate choice, such as reproductive delays due to time spent mate searching (Kokko & Ots 2006) and missed mating opportunities (Kokko & Ekman 2002) could explain why some species exhibit inbreeding tolerance (Jamieson *et al.* 2009) or even preference (Wang & Lu 2011). Yet, acorn woodpeckers *Melanerpes formicivorus,* for example, exhibit strong inbreeding avoidance despite the reproductive cost of lost breeding opportunities (Koenig *et al.* 1999). So, there is some discrepancy between theoretical predictions and the rarity of inbreeding in nature (Kokko & Ots 2006), possibly because inbreeding depression is difficult to quantify and thus is potentially underestimated in the wild (Duthie & Reid 2016).

In most cooperative breeders, delayed natal dispersal creates kin structured populations (Hatchwell 2009) with a high encounter rate between adult opposite-sex kin. However, the fitness costs associated with inbreeding has selected for various inbreeding avoidance mechanisms (reviewed in Pusey & Wolf 1996; Riehl & Stern 2015), so that incest is extremely rare in animal societies. In most cooperative breeders, one sex is philopatric, while the other disperses away from the natal area to breed (Koenig & Haydock 2004). In birds, females tend to disperse, whereas in mammals, males typically leave the natal group to breed (Greenwood 1980). Phylogenetic analyses show that transitions from female philopatry to female dispersal are associated with inbreeding risk caused by long male reproductive tenure in mammals (Lukas & Clutton-Brock 2011). That dispersal usually occurs before reproduction (Brown 1987) and is influenced by the presence of relatives (Koenig *et* *al.* 1998; Cockburn *et al*. 2003) also suggest that inbreeding avoidance is an important driver of dispersal patterns (Greenwood 1980).

When dispersal is sex-biased, the philopatric sex may still be at risk of inbreeding, particularly in social groups with overlapping generations (Lukas & Clutton-Brock 2012; Blyton *et al.* 2015). The timing and distance over which mate choice occurs drives selection for alternative inbreeding avoidance mechanisms. For example, a risk of inbreeding may arise when a dominant breeder in a social group dies, leaving a parent with opposite-sex offspring as potential partners (Emlen 1995; Koenig *et al.* 1998). Similarly, in species with dominance hierarchies, widowed dominants often abandon their breeding position if the highest ranking opposite-sex subordinate is a relative (Daniels & Walters 2000). However, dispersal may be constrained by an absence of breeding vacancies (Cockburn *et al.* 2003) or if the benefits of philopatry outweigh its costs (Stacey & Ligon 1987). In such situations, breeders may select closely related social partners, but mate with less closely related individuals from outside the family group (Tarvin *et al.* 2005). The most notable example of infidelity as an inbreeding avoidance strategy comes from the fairy-wrens (*Malurus spp*); early studies suggested unprecedented levels of incest (Rowley *et al.* 1986) but molecular genetic techniques have since revealed an extraordinarily high rate of extra-pair paternity (Brooker *et al.* 1990). The effectiveness of this strategy depends on the difference in relatedness between within-group and extra-group partners, and offspring sired by extra-pair males are indeed less inbred than within-pair offspring in red-backed fairy-wrens *M. melanocephalus* (Varian-Ramos & Webster 2012). Alternatively, when mate choice is restricted, and the only option is to mate with close kin, individuals may abstain from breeding altogether. In hierarchical cooperative breeders, subordinates may forego reproduction when an opposite-sex parent is present in the breeding group (Walters *et al.* 1988; Koenig *et al.* 1998), or are reproductively supressed by same-sex dominants (Whittingham *et al.* 1997; Koenig *et al.* 1999; O'Riain *et al.* 2000).

All of these strategies can mitigate inbreeding risk in viscous populations without the need for active kin discrimination. However, some cooperative species do not live in discrete family groups, but rather, cooperation follows local and/or coordinated natal dispersal, resulting in continued association among breeding relatives across extended *kin neighbourhoods* (Dickinson & Hatchwell 2004). Here, interactions between opposite-sex adults of varying degrees of relatedness may be frequent, particularly if there are small dispersal differences between the sexes (Blyton *et al.* 2015). In such species, if inbreeding is costly, effective kin recognition mechanisms may be under strong selection (Cornwallis *et al.* 2009). Indeed, kin-neighbourhood cooperative breeders have shown sophisticated kin discrimination in their cooperative behaviour (Nam *et al.* 2010; McDonald & Wright 2011; Akçay *et al.* 2013). Although empirical evidence of kin recognition during mate choice is scarce, a recent study on Western bluebirds *Sialia mexicana* demonstrated kin avoidance when mates are selected from within the range of kin clustering (Dickinson *et al.* 2016).

Familiarity is often an effective mechanism of kin recognition in social animals, as kinship tends to co-vary with association (Pusey & Wolf 1996). However, in kin neighbourhood species, in which the probability of associating with non-kin, and encountering unfamiliar kin is reasonably high, there may be strong selection for more sophisticated mechanisms, which could include an assessment of relatedness via phenotype matching (Komdeur & Hatchwell 1999). Further, effective recognition of unfamiliar kin may be of particular importance during mate choice, assuming that even moderate inbreeding can lead to inbreeding depression. Direct evidence for effective kin recognition during mate choice is lacking, often hindered by the difficulty of obtaining reliable genetic relatedness estimates or inbreeding coefficients (Pemberton 2004), coupled with detailed knowledge of paring decisions, null mate choice models and a putative mechanism.

Here, I investigate the fitness costs, risk and avoidance of inbreeding in a well-studied, natural population of cooperatively breeding long-tailed tits. Identifying the risk of incest and quantifying inbreeding depression are essential for understanding the selection pressure for alternative inbreeding strategies in wild populations. However, empirical studies that aim to address all of these aspects are rare for several reasons. Firstly, inbreeding depression is difficult to quantify in wild populations because it requires long-term genetic or pedigree data, and enough variation in f to detect any relationship between inbreeding and fitness (Slate *et al.* 2000). Secondly, accurate mate choice models can be hard to construct, as the pool of potential partners is not often known. Finally, the mechanism of kin discrimination that permits incest avoidance remains unidentified. Long-tailed tits are an excellent species in which to investigate inbreeding because they live in kin-structured breeding populations with prolonged association between opposite-sex relatives (Leedale *et al.* 2018), so that both kin and non-kin are often available as partners. Therefore, strong selection for active kin discrimination during mate choice is expected. Furthermore, through long-term field observations, we can build a relatively accurate null model of mate choice that is random with respect to relatedness: long-tailed tits are non-territorial, there is an even adult sex ratio and virtually all birds attempt to breed from their first year (Hatchwell *et al.* 2001a). An effective kin recognition mechanism has also previously been established (Russell & Hatchwell 2001; Sharp *et al.* 2005). Finally, previous studies suggest that although long-tailed tits can effectively identify kin, recognition errors do occur (Hatchwell 2016). I therefore expect to detect some cases of inbreeding, allowing investigation of inbreeding depression.

Our aims were four-fold. First, I used long-term genetic and field data to quantify inbreeding in our long-tailed tit population. Second, I explored the fitness costs of inbreeding, by correlating various fitness traits with heterozygosity. Third, I determined the risk of inbreeding by comparing the mate choice range of males and females to the previously described population genetic structure. Finally, I tested whether or not kin are actively avoided as partners using null mate choice models and identified the likely relatedness threshold for acceptance of individuals as mates.

3.2 Methods

3.2.1 Study population and field methods

A population of 17-72 (mean *c*.50) pairs of long-tailed tits was studied during the breeding season (February-June) between 1994 and 2016 in the Rivelin Valley, Sheffield, UK (53º38'N 1º56W). The site is approximately 2.5km^2 and comprises predominantly deciduous woodland and scrub. The population is open: approximately 40% of breeders hatched in the study site, and are referred to as *native* (A. E. Leedale & B. J. Hatchwell, unpublished data). The remaining *immigrant* adults are assumed to be first year breeders that have dispersed in to the study site, based on the observation that individuals have high site fidelity following their first breeding year (McGowan *et al*. 2003). Each year, almost all individuals (>95%) were marked with a BTO ring and a unique combination of two colour rings. Native birds were ringed as 11-day old nestlings and immigrant adults were captured in mist nests under BTO licence before or during breeding. When ringed, a sample of 5-30µl of blood was taken by brachial venepuncture under Home Office licence. All breeding attempts were closely monitored and GPS coordinates were taken for each nest (n = 1461); a Cartesian coordinate system (UTM) was used to describe geographic distance between nests.

3.2.2 Molecular genetics

Individuals were genotyped at eight (1994-2006) or 17 (2007-2016) microsatellite loci. For further details on genotyping, see Simeoni *et al*. (2007) and Adams *et al*. (2015). The population allele frequencies used in analyses were generated using all genotyped individuals (1994-2016, n = 3182) in *CERVUS version 3.0.7* (Kalinowski *et al*. 2007), to maximise accuracy in estimating the frequency of rare alleles and to ensure non-zero frequencies for all alleles in the dataset (Appendix 3). The genetic relatedness of dyads was estimated using Queller and Goodnight's (1989) r_{QG} coefficient of relatedness in *SPAGeDi version 1.1.5* (Hardy & Vekemans 2002). This estimate is reliable when tested against our social pedigree (Nam *et al*. 2010).

3.2.3 Social pedigree

The social pedigree was created using 22 years of field observations as described above (1994-2016, n = 2815 birds). For further details on pedigree construction, see Appendix 1. To calculate social relatedness (*r*) among dyads, an additive relationship matrix was generated from the pedigree in *R version 3.5.0* (R Core Team 2018) using the *R* package *nadiv* (Wolak 2012). Six breeding birds in our study population (0.2%) were from cross-fostered broods in 1996-1998, but given that birds raised together treat each other as kin (Hatchwell *et al*. 2001b; Sharp *et al*. 2005), I include them in the social pedigree. Similarly, while there is a low rate of extra-pair paternity (<8%)

in long-tailed tits (Hatchwell *et al.* 2002), I have not corrected for it in the social pedigree.

3.2.4 Inbreeding

The average population inbreeding coefficients (F_{IS} = -0.002 ± 0.004, R_{IS} = -0.014 ± 0.027) indicate no significant inbreeding or outbreeding in the study population (Leedale *et al.* 2018). However, that does not mean that incest does not occur. I calculated the genetic relatedness (r_{QG}) and social relatedness (r) of all breeding pairs from 1994-2016 in which both adults were ringed and genotyped. Occasionally, long-tailed tits swap partners within a breeding season. In these cases, I used the first pairing made that year. Individuals often breed in multiple years, either with the same partner or a new partner. The dataset used in this study contained 696 observations of 609 pairs made up of 493 females and 529 males over 22 years.

I identified cases of inbreeding using the social pedigree. Pairs were considered closely inbred if they were formed of first-order (r = 0.5) or second-order (r = 0.25) kin. I also identified breeding between third-order kin (r = 0.125) as distantly inbred. In open populations, the social pedigree is incomplete, and may underestimate the level of inbreeding. Therefore, genetic inbreeding was also quantified by classifying close and distant kin using r_{QG} estimates. The distribution of r_{QG} estimates of known first-order, second-order and non-kin were used to identify two lower genetic similarity thresholds which optimise the inclusion of first or second-order kin and the exclusion of more distant kin. These thresholds were used to determine the proportion of close and distant genetic inbreeding.

3.2.5 Inbreeding costs

Pedigree-derived inbreeding coefficients are reliable only when parentage can be traced back at least two generations. As is inevitable in natural, open populations, our pedigree is incomplete, and both sets of grandparents are known for only a small proportion (5.3%) of native birds. Associations between heterozygosity at microsatellite markers and variation in fitness traits, known as heterozygosity-fitness correlations (HFCs), have been widely used as an indirect method of measuring inbreeding depression (e.g. Coltman *et al.* 1999; Slate *et al.* 2000; Chapman & Sheldon 2011). However, HFCs only reflect variance in inbreeding if certain assumptions are met: (i) genetic diversity at marker loci reflects genetic diversity at loci that affect trait variation; and (ii) the marker and fitness loci are in identity disequilibrium (ID) that is, the increase in fitness with increasing heterozygosity is due to the non-random association of diploid genotypes (Hansson & Westerberg 2002; Chapman *et al*. 2009). Under this *general effect hypothesis* (David *et al.* 1995), the strength of the relationship between heterozygosity and fitness depends on the population variance in f (Szulkin *et al.* 2010). I therefore used equations from Slate *et al.* (2004) to calculate the predicted correlation between heterozygosity and f in our population. Inbreeding coefficients were calculated from the social pedigree using the *R* package *pedigree* (Coster 2013). It was possible to infer reliable f values from the pedigree for 129 birds (native individuals with all grandparents known). I also included f values from an additional 9 birds that were probable offspring of immigrant siblings, based on genetic sibship reconstruction (Appendix 1). The predicted correlation between heterozygosity and f was $R = -0.43$ ($n = 138$, mean $f = 0.03$, variance in $f = 0.004$, number of loci = 17, mean heterozygosity of loci = 0.759), supporting the use of heterozygosity as a proxy for inbreeding coefficient in our population.

Using the *R* package, *inbreedR* (Stoffel *et al.* 2016), I estimated standardised multi-locus heterozygosity (H_S) for 3182 genotypes, defined as the proportion of typed loci for which an individual was heterozygous divided by the mean heterozygosity of those loci at which the individual was typed (Coltman *et al.* 1999). I tested for an association between H_S and fitness using four life-history traits: (i) whether or not an individual recruited to the breeding population; (ii) hatching success in first-year females (number of nestlings on day 11/clutch size); (iii) whether or not a breeder produced recruits (among birds that produced fledglings); and (iv) direct fitness (among birds that produced recruits). Direct fitness was calculated as lifetime reproductive success quantified in terms of genetic offspring equivalents and corrected for extra-pair paternity and the offspring gained by having helpers. The fraction of recruits in a brood that was attributable to helpers was estimated using a mixed effects model of the effect of helper number on recruitment (J. P. Green & B. J. Hatchwell, unpublished data). This fraction was subtracted from the total number of recruits produced over an individual's lifetime. The remaining fraction was halved to obtain the fitness attributable to a single parent, and halved again to reflect the relatedness between parents and offspring.

3.2.6 Inbreeding risk and avoidance

In breeding populations of long-tailed tits, significant levels of relatedness have been identified between males and females within 600m (Leedale *et al*. 2018). For this to generate a potential risk of incest, mates must be selected from within this range. Spatial pairing range was measured as the distance between an individual's natal nest and their first nest as a breeder, and was calculated separately for males (n = 230) and females (n = 109).

The inbreeding risk (IR) was quantified by calculating the mean r_{QG} of breeders from known pairs to all their potential partners, excluding chosen partner, each year, under the pairing constraints of multiple mate choice models. The first null model assumes random mate choice across the study site, and includes all males and females present in the breeding population. However, the limited pairing range of long-tailed tits (see Results) means that spatially constricted null models are more appropriate. Therefore, a series of additional null models was generated restricting the pool of potential partners to opposite-sex birds within concentric rings at 300m, 600m, 900m, and further 300m increments up to 2100m. The distance between adults is based on the location of their first breeding attempt in a given year. In a small number of observations (<0.01%), an individual's first recorded location in a given year was as a helper. This may be because they did not breed independently or they had a failed attempt outside the study area. In these cases, I used the location of the helped nest to calculate the distance between adults, as failed breeders rarely travel beyond 300m to provide help (Leedale *et al*. 2018).

The assumption that all opposite-sex birds within a given distance are available as partners may also be inappropriate because long-tailed tits pair-bonds can persist across years (Hatchwell *et al.* 2000). I therefore generated a more restrictive set of models including only first-year, widowed or divorced opposite-sex breeders as potential partners within the defined spatial ranges. Under each model, IR was compared with the corresponding r_{QG} observed. A lower level of r_{QG} among observed pairs compared to the IR predicted by each model would suggest active kin discrimination during mate choice, no difference would indicate random mate choice, and an r_{QG} that was higher that IR would suggest inbreeding preference. I also tested whether IR predicted pair r_{QG}. Here, a positive relationship between IR and pair r_{QG}

with a slope $\beta \approx 1$ would indicate inbreeding tolerance, whereas a weak or no relationship (slope $\beta < 1$) would indicate inbreeding avoidance.

To identify a plausible discrimination rule for incest avoidance, I compared pair r_{QG} to IR under each model assuming: (i) random choice with respect to kinship (as above), (ii) avoidance of close kin ($r = 0.5$), and (iii) avoidance of close and distant kin ($r = 0.25$ and 0.5), by step-wise removal of close and distant kin from the pool of potential partners, for each focal individual. Close and distant kin were categorised using the genetic relatedness thresholds (see above).

3.2.7 Statistical analysis

Associations between heterozygosity and fitness traits (HFCs) were examined using generalised linear mixed effects models (GLMMs) in the *R* package *lme4* (Bates *et al.* 2015). *Recruitment* was modelled as a binary response variable with a binomial error distribution and logit link. The fixed effects were: H_S, *sex*, to control for male-biased philopatry, *fledge date* (days since March 1st), because offspring fledging earlier in the year have a greater probability of recruitment (MacColl & Hatchwell 2002), and *number of helpers* at natal nest (five factor levels: 0-3+), as helper number has been shown to increase recruitment probability (Hatchwell *et al.* 2014). *Hatching success* was modelled as a continuous response variable with a Gamma error distribution and identity link. The fixed effects were: H_S, *lay date* and *female mass as a nestling. Probability of producing recruits* was modelled as a binary response variable with a binomial error distribution and logit link. The fixed effects were: H_S, *sex*, and *fledgling sex ratio* (proportion of male fledglings produced), to control for male philopatry. *Direct fitness* was modelled as a continuous response variable with a Gamma error distribution and inverse link, and with H_S, *sex*, and *fledgling sex ratio* fitted as fixed effects. In all HFC models, *genetic brood* was fitted as a random effect to avoid pseudoreplication of H_S estimates and control for seasonal differences. In *Recruitment* models, *fledge year* was also fitted as a random effect, and in *hatching success* models, *breeding year* was also fitted as a random effect. For each fitness trait, all appropriate covariates were initially included in a full model and then dropped sequentially unless doing so significantly reduced the fit of the model (best-fit models were selected at the ΔAIC < 2 criterion). A Chi-squared test was used to determine whether the model incorporating H_S plus any other retained covariates explained significantly more variance than a null model without H_S included as a predictor. All effects were considered to be statistically significant if $p < 0.05$. Sample sizes are reported with the model outputs.

Analyses of the mating options available to males and females were conducted in separate mate choice models. As the same allele frequencies are used to calculate r_{QG} across years, the r_{QG} of unique dyads is consistent. However, due to demographic factors such as divorce, migration, birth, death, and dispersal, IR to breeders under each model will vary across years. I compared individual IR under each model to the observed r_{QG} estimates using paired t-tests. When pairs persisted across years, the first year a pair was observed was used in the analysis. To test whether IR predicted pair r_{QG}, I fitted linear mixed effects models (LMMs) with *pair* r_{QG} as the response variable and *IR* under the pairing constraints of each model as the fixed effect. Models were fitted with Restricted Maximum Likelihood (REML) and both the *focal bird ID* and the *ID of the chosen/potential partner* were fitted as random effects. This model had the lowest Akaike information criterion (AIC) value when tested against other models including *dyad ID* and *year* as random effects and accounts for replicates of individuals and dyads.

To identify a putative decision rule for inbreeding avoidance, I fitted LMMs with REML to test for differences between *pair* $r_{QG,}$ and (i) *IR under random mate choice*, (ii) *IR after the removal of close kin*, and (iii) *IR after the removal of close and distant kin*, in males and females. When pairs persisted across years, the first year a pair was observed was used in the analysis. *Year* nested within *focal bird ID* were fitted as random effects. This model was the model with the lowest AIC value when tested against other models including *dyad ID* and *chosen/potential partner ID* as random effects and accounts for replicates of individuals and dyads.

3.3 Results

3.3.1 Relatedness estimates and inbreeding

The r_{QG} estimate of known first-order kin ($r = 0.5$) was 0.454 ± 0.149 (mean ± SD, n = 1211 dyads). For dyads known to be second-order kin ($r = 0.25$), r_{QG} was 0.198 ± 0.154 (mean ± SD, n = 515 dyads). The r_{QG} estimate of all other dyads of known parentage ($r < 0.25$) was 0.002 ± 0.131 (mean ± SD, n = 54521 dyads). The distribution of r_{QG} estimates among known first-order, second-order and non-kin are shown in Fig. 3.1. Based on these distributions, I set a lower r_{QG} threshold of 0.375 to approximate close kin (mean r_{QG} ± SD = 0.503 ± 0.094, n = 1438) and 0.125 to approximate distant kin (mean r_{QG} ± SD = 0.197 ± 0.059, n = 11979). The mean r_{QG} of observed pairs was 0.002 ± 0.123 (mean ± SD, n = 609). Based on the social pedigree, three out of 609 pairs (0.5%) consisted of first or second-order relatives, and a further three pairings were between third-order kin (Appendix 5, Table A5.1). Genetic relatedness estimates revealed a similar frequency of close inbreeding (2/609, 0.3%), but substantially more cases of distant inbreeding (96/609, 15.8%; Appendix 5, Table A5.2).

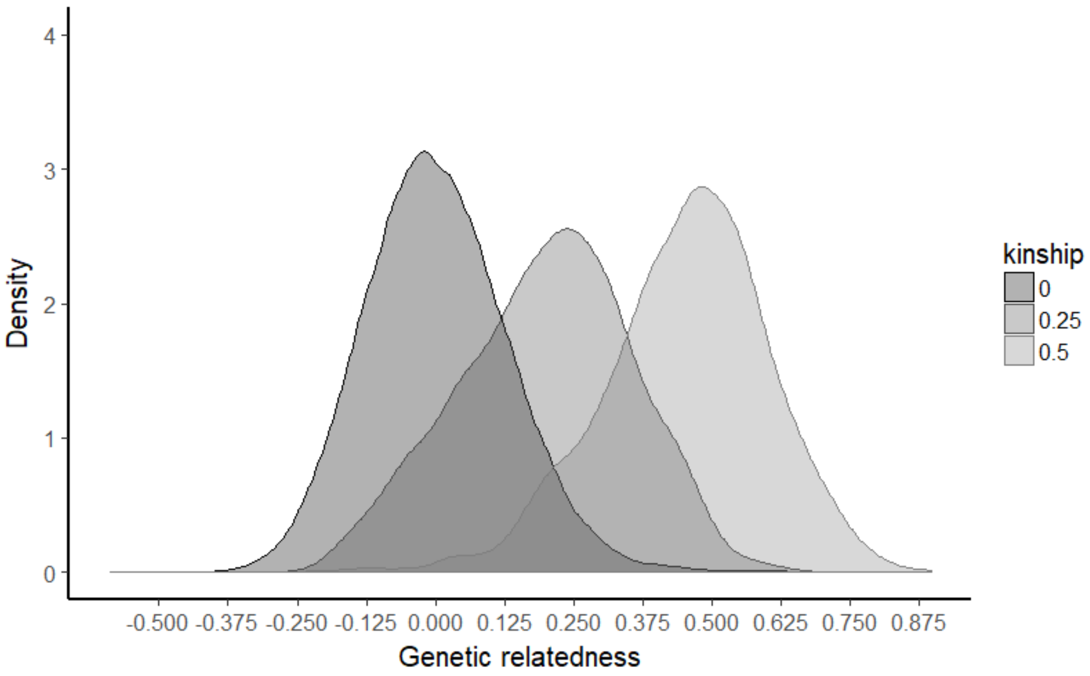

Figure 3.1. The distribution of genetic relatedness estimates (r_{QG}) among known first-order (kinship = 0.5), second-order (kinship = 0.25) and non-kin ((kinship = 0) from the long-tailed tit social pedigree (1994-2016).

3.3.2 Inbreeding costs

Recruitment was greater for males (χ^2 = 51.141, n = 1903, df = 1, p < 0.001) and increased with number of helpers present at the natal nest (χ^2 = 13.002, n = 1903, df = 1, p < 0.001), but did not correlate with individual heterozygosity (χ^2 = 0.093, n = 1903, df = 1, p = 0.761, Fig. 3.2a). There was also no significant correlation between heterozygosity and hatching success among recruits (χ^2 = 2.578, n = 30, df = 1, p = 0.108, Fig. 3.2b). Among breeders that produced fledglings, the probability of producing recruits increased with number of fledglings (χ^2 = 5.876, n = 81, df = 1, p < 0.05), but not heterozygosity (χ^2 = 3.482, n = 81, df = 1, p = 0.062, Fig. 3.2c). However, heterozygosity was positively associated with the direct fitness of individuals that produced at least one recruit (χ^2 = 13.127, n = 38, df = 1, p < 0.001, Fig. 3.2d).

3.3.3 Inbreeding risk and null mate choice models

The median pairing range was 523m for females (n = 109) and 393m for males (n = 230), with a maximum mate searching distance (upper 95% CI) of 1400m for females and 1200m for males (Fig. 3.3). Thus pairing typically occurs within the range of kin clustering (Leedale *et al.* 2018) and could lead to incest if mates are selected at random with respect to kinship. There was very little difference in inbreeding risk (IR) to either sex under null models including all opposite-sex birds as potential partners, and null models including only first-year, widowed and divorced birds (Fig. 3.4). However, this second model is more biologically accurate, and I focus on null models of pairing behaviour that restrict the pool of potential partners to first-year, widowed and divorced birds within 1200m in our analyses.

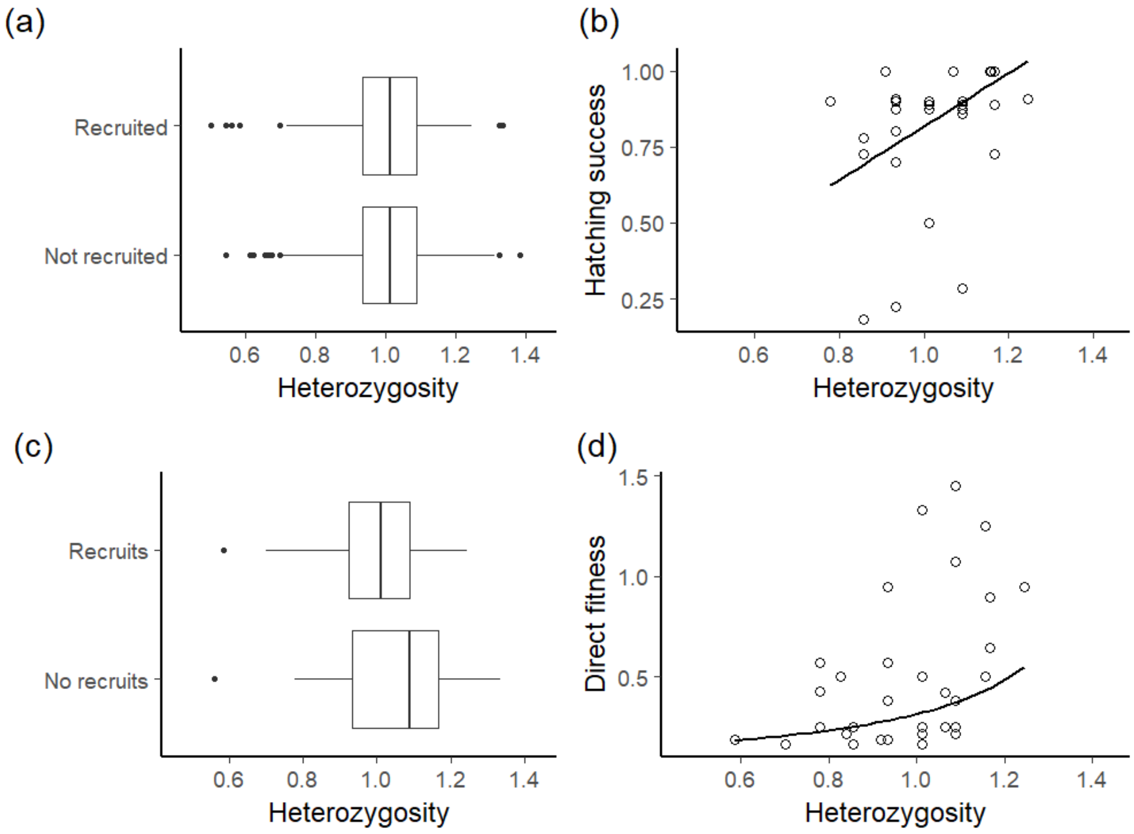

Figure 3.2. Relationship between standardised heterozygosity of microsatellite loci and (a) probability of recruitment (n = 1903), (b) hatching success in first-year females (n = 30), (c) probability of producing recruits in breeders that produced fledglings (n = 81), (d) direct fitness of breeders that produced recruits (n = 38). Solid lines represent model predictions, boxplots represent median (+IQR) heterozygosity of factor levels.

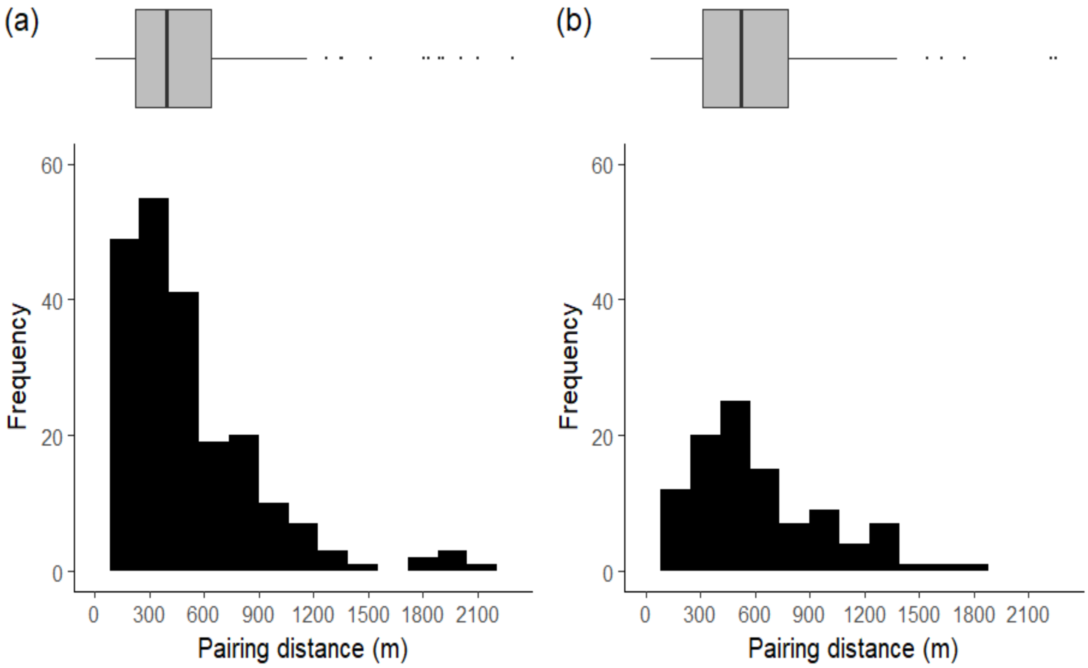

Figure 3.3. Frequency distribution and median (+IQR) pairing ranges of (a) male (n = 230) and (b) female (n = 109) long-tailed tit breeders, calculated as the distance between an individual's natal nest and their first breeding attempt.

3.3.4 Inbreeding avoidance

Combining data across years, mean r_{QG} to a chosen partner was significantly lower than the predicted IR when females selected partners from within 300m (paired t-tests: t = 5.368, df = 520, $p < 0.001$), 600m (t = 2.923, df = 565, $p < 0.01$) and 900m (t = 2.28, df = 571, $p < 0.05$); but not 1200m (t = 1.432, df = 573, p = 0.076; Fig. 3.4a). From the male perspective, mean r_{QG} to their chosen partner was lower than predicted when mates were selected from females within 300m (t = 6.158, df = 506, $p < 0.001$), 600m (t = 3.988, df = 563, $p < 0.001$), 900m (t = 2.888, df = 570, $p < 0.01$) and 1200m (t = 1.875, df = 572, $p < 0.05$; Fig. 3.4b).

For females, I found that IR did not predict the observed level of pair r_{QG} when assuming partners were selected from available birds within 300m (r = 0.031 ± 0.004, df = 545.3, t = 0.778, p = 0.437; Fig. 3.5a) or 600m (r = 0.093 ± 0.068, df = 598.8, t = 1.358, p = 0.175; Fig. 3.5b). There was a positive relationship between IR and pair r_{QG} within 900m (r = 0.229 ± 0.088, df = 597.6, t = 2.603, $p < 0.01$; Fig. 3.5c) and 1200m (r = 0.313 ± 0.096, df = 581.8, t = 3.27, $p < 0.01$; Fig. 3.5d). For males, IR did not predict pair r_{QG} under models assuming mate choice occurs within 300m (r = 0.047 ± 0.038, df = 537.6, t = 1.3, p = 0.194; Fig. 3.5e), 600m (r = 0.126 ± 0.069, df = 507.62, t = 1.825, p = 0.069; Fig. 3.5f) or 900m (r = 0.166 ± 0.095, df = 547.8, t = 1.734, p = 0.083; Fig. 3.5g), although there was a positive relationship between IR and pair r_{QG} within 1200m (r = 0.225 ± 0.011, df = 509.2, t = 2.047, $p < 0.05$; Fig. 3.5h). It should be noted that in those cases where there was a positive relationship between pair r_{QG} and IR, the slope β was substantially < 1 (range: 0.031 – 0.313).

To determine the kin discrimination rule that permits this level of inbreeding avoidance, the observed r_{QG} to a chosen partner was compared with IR after step-wise removal of close ($r_{QG} \geq 0.375$) and distant ($r_{QG} \geq 0.125$) kin from the pool of potential partners at multiple pairing ranges. With the removal of close kin, the r_{QG} of observed pairs did not differ from IR to females within all ranges examined (Table 3.1, Fig. 3.6a). From a male perspective, the removal of close kin resulted in no difference between observed pair r_{QG} and IR within 600-1200m, but pair r_{QG} was lower than IR within 300m (Table 3.1, Fig. 3.6b). In contrast, when close and distant kin were removed, IR was lower than observed pair r_{QG} at all ranges and for both sexes (Table 3.1, Fig. 3.6).

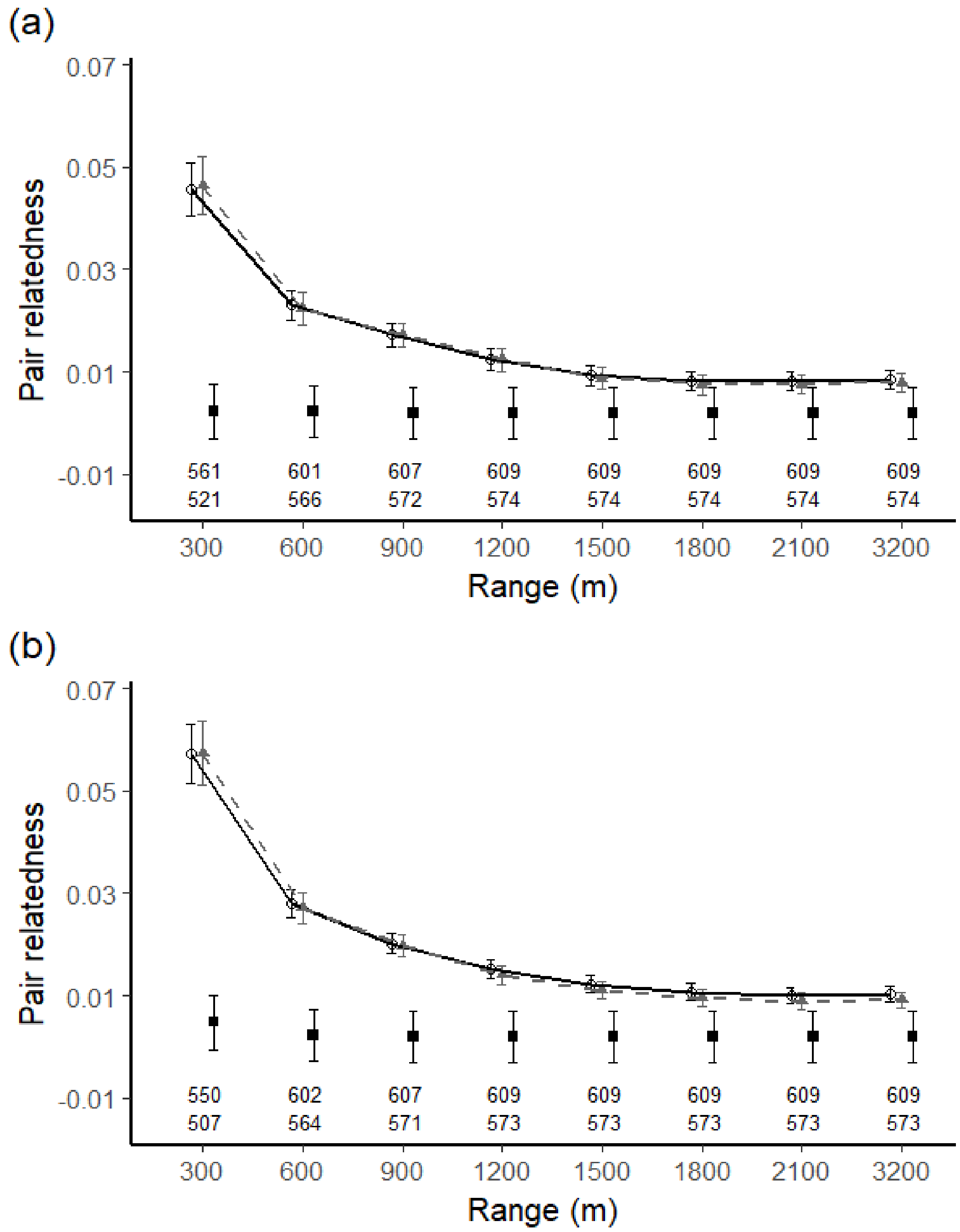

Figure 3.4. Mean genetic relatedness of breeding pairs formed within increasing ranges (closed squares), and the expected relatedness if (a) females or (b) males selected mates at random (grey triangles), or from a restricted pool of first-year, widowed or divorced breeders (open circles) within each range. Expected relatedness is calculated as the mean relatedness of focal birds to all opposite-sex breeders (excluding chosen partner) available under each mate choice model. Error bars represent standard error around the mean. Numbers above the x axis represent the number of observed pairings (top = random mate choice, bottom = restricted pool of first-year, widowed or divorced breeders).

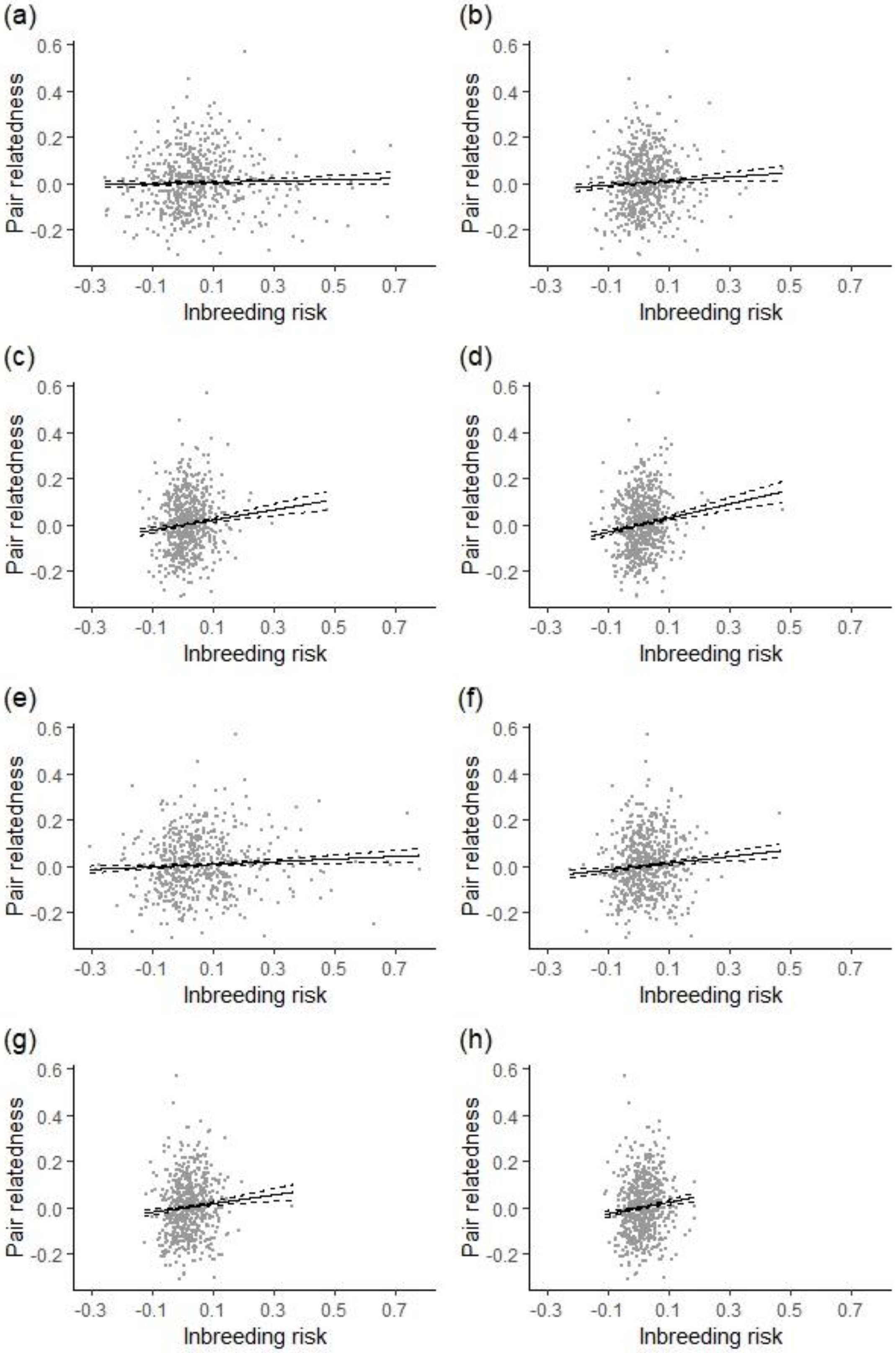

Figure 3.5. Mean genetic relatedness of breeding pairs in response to that predicted by null models of pairing behaviour, in which females select partners from within (a) 300m, (b) 600m, (c) 900m and (d) 1200m; and males select partners from within (e) 300m, (f) 600m, (g) 900m and (h) 1200m. Solid lines represent model predictions with standard error (dashed lines).

Table 3.1. Kin discrimination rules for inbreeding avoidance in female and male long-tailed tits. Genetic relatedness to chosen partner (pair r_{QG}) was compared with inbreeding risk (IR) under a series of mate choice models: random with respect to kinship; avoidance of close kin ($r_{QG} \geq 0.375$); and avoidance of close and distant kin ($r_{QG} \geq 0.125$) at multiple pairing ranges (m).

sex	mate choice model	range	pair r_{QG}	IR	df	t	p
female	random	300	0.005	0.046	1479	7.470	< 0.001
		600	0.005	0.022	1683	4.107	< 0.001
		900	0.005	0.017	1703	3.201	< 0.01
		1200	0.005	0.012	1707	2.010	< 0.05
	avoidance of close kin	300	0.005	0.009	1484	1.037	0.3
		600	0.005	0.001	1684	-1.169	0.242
		900	0.005	0.0004	1703	-1.073	0.284
		1200	0.005	-0.001	1708	-1.400	0.162
	avoidance of close and distant kin	300	0.005	-0.039	1498	-6.478	< 0.001
		600	0.005	-0.039	1684	-10.332	< 0.001
		900	0.005	-0.038	1703	-10.828	< 0.001
		1200	0.005	-0.038	1708	-11.267	< 0.001
male	random	300	0.008	0.057	1400	8.557	< 0.001
		600	0.004	0.027	1683	5.401	< 0.001
		900	0.004	0.02	1705	4.017	< 0.001
		1200	0.004	0.014	1716	2.615	< 0.01
	avoidance of close kin	300	0.008	0.02	1405	2.635	< 0.01
		600	0.004	0.005	1683	0.115	0.909
		900	0.004	0.004	1705	-0.135	0.893
		1200	0.004	0.001	1716	-0.805	0.421
	avoidance of close and distant kin	300	0.008	-0.036	1422	-5.302	< 0.001
		600	0.004	-0.037	1684	-9.639	< 0.001
		900	0.004	-0.036	1707	-10.302	< 0.001
		1200	0.004	-0.037	1716	-10.951	< 0.001

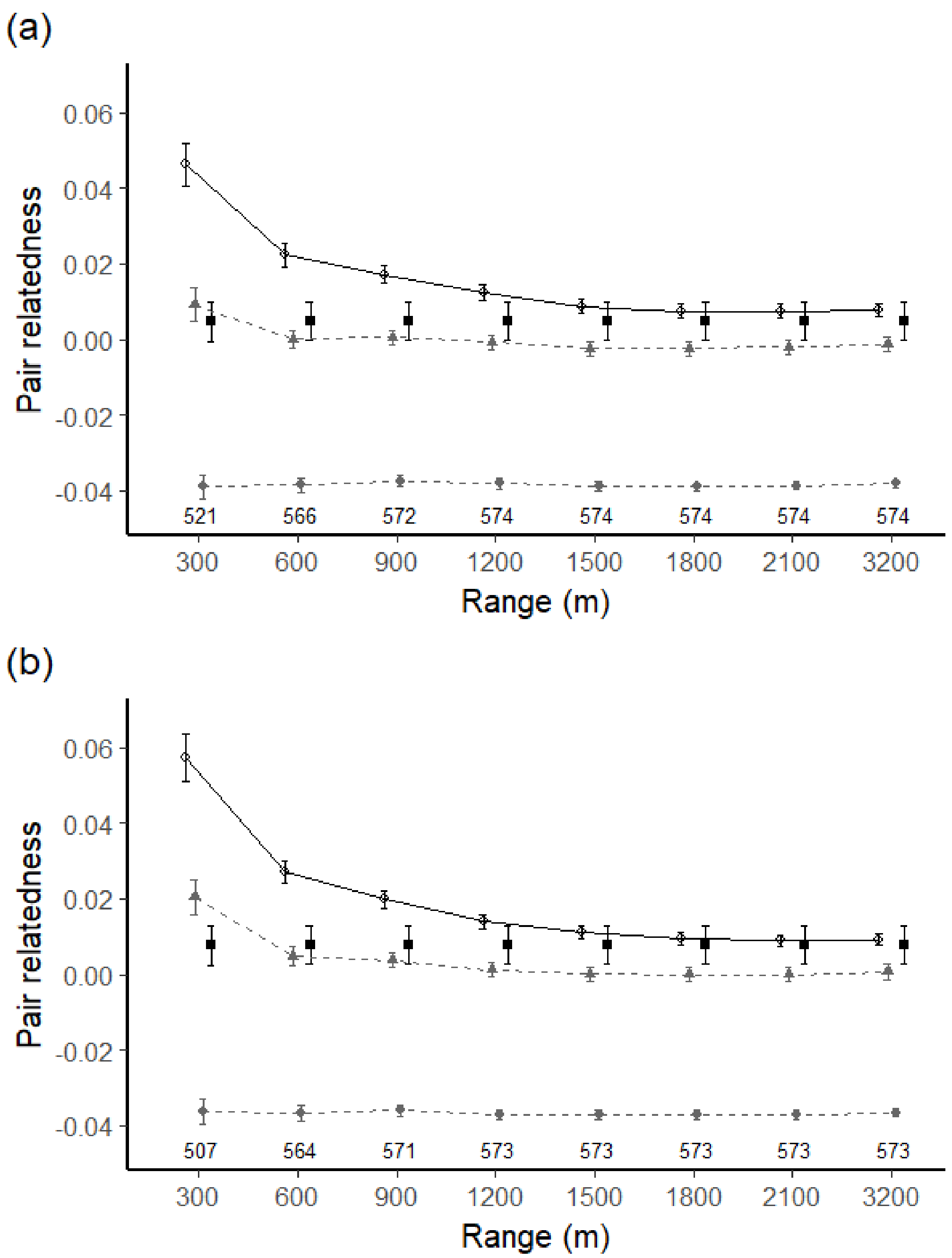

Figure 3.6. Mean genetic relatedness of breeding pairs formed within increasing ranges (black squares), and the expected relatedness if (a) females or (b) males: (i) selected mates at random with respect to kinship (open circles), (ii) avoided kin with $r_{QG} > 0.375$ (grey triangles), and (iii) avoided kin with $r_{QG} > 0.125$ (grey circles). Expected relatedness is calculated as the mean relatedness of focal birds to all opposite-sex breeders (excluding chosen partner) available under each mate choice model. Error bars represent standard error around the mean. Numbers above the x axis represent the number of observed pairings.

3.4 Discussion

This study aimed to investigate the costs, risk and avoidance of inbreeding in cooperatively breeding long-tailed tits. My results show that reduced heterozygosity is associated with lower direct fitness (Fig. 3.2a), providing evidence of inbreeding depression, and a selective pressure for inbreeding avoidance to evolve. I also reveal that both males and females typically choose partners from within 600m of their natal nest (Fig. 3.3). As earlier studies on this species have demonstrated that opposite-sex kin remained clustered within this range during breeding (Leedale *et al.* 2018), this generates a potential inbreeding risk if mate choice is random with respect to kinship. However, over the 22 years that this population was monitored, I identified just six cases of pedigree inbreeding ($f \geq 0.125$), only one of which was between familiar, first-order kin. Moreover, the relatedness of breeding pairs was much lower than that expected if mates were selected randomly (Fig. 3.4 & Fig. 3.5). When close kin ($r_{QG} \geq 0.375$) were removed from the pool of potential partners, the observed pair relatedness matched that expected under most models of random mate choice (Fig. 3.6). Previous studies have also shown that long-tailed tits have an effective mechanism of kin recognition (Sharp *et al.* 2005), and my results indicate that long-tailed tits avoid close kin during mate choice, thereby minimising the fitness costs associated with inbreeding.

3.4.1 Inbreeding costs

I found a significant correlation between heterozygosity and direct fitness, demonstrating a fitness cost to inbreeding in long-tailed tits. Heterozygosity did not correlate significantly with any other trait measured. This suggests that inbreeding in long-tailed tits has long-term fitness consequences. Inbreeding depression may be masked in the short-term because external factors such as weather and the stage in the breeding season that reproductive events occur (e.g. onset of egg-laying) have disproportionately large impacts on offspring fitness in the early stages of life-history. As the influence of these external factors lessens later in life, individual quality becomes the most important determinant of fitness. This may explain why inbreeding depression in long-tailed tits appears delayed.

Inbreeding depression can be masked in natural populations because a large proportion of offspring often die before being sampled, particularly if they are inbred (Taylor *et al.* 2010). The effects of heterozygosity on hatching success could be obscured by using the number of nestlings present in the nest on day 11 as a proxy for number of eggs hatched. Although nestling mortality is infrequent (Hatchwell *et al.* 2004) some may have died before 11 days old, which could lead to minor discrepancies. Inbreeding depression is also likely to have an environmental component. For example, maternal care has been shown to buffer inbreeding depression in burying beetles *Nicrophorus vespilloides* (Pilakouta *et al.* 2015). Sociality and associated cooperative behaviours, such as the effects of helpers in long-tailed tits, can increase offspring fitness and may mitigate some of the fitness consequences of inbreeding depression.

Measuring inbreeding depression directly is challenging in populations with low levels of incest, but is often inferred using HFCs. HFCs are an effective proxy for inbreeding depression when the marker and fitness loci are in identity disequilibrium, (David *et al.* 1995), but two other processes can lead to non-random associations of alleles, that are not associated with inbreeding: linkage disequilibrium, and functional overdominance, whereby markers directly affect fitness when expressed (Chapman *et* *al.* 2009). It is therefore crucial to assess the relationship between heterozygosity at microsatellite loci and inbreeding coefficients across the population before implementing HFCs as a measure of inbreeding depression. Here, the relationship between marker heterozygosity and *f* was -0.43. This confirms marker heterozygosity as a suitable proxy for inbreeding coefficient, and evidence of inbreeding depression has been detected in studies with similar correlations (Slate *et al.* 2004).

3.4.2 Inbreeding risk

Long-tailed tits do not exhibit delayed dispersal, but breed within kin neighbourhoods in which philopatric offspring are reproductive (Hatchwell *et al.* 2000). Despite the fitness costs associated with inbreeding, mates are selected from within a range that incurs an inbreeding risk. This suggests that there are compensating benefits of breeding close to relatives. Nest failure rate in long-tailed tits is high, yet failed breeders can acquire indirect fitness by increasing production of non-descendant kin (Hatchwell *et al.* 2014), so there are inclusive fitness benefits as well as costs from limited dispersal. Outside of the breeding season, failed breeders join nonbreeding flocks containing relatives, and although some flock switching occurs, dispersers often move with opposite-sex kin (Napper & Hatchwell 2016). Previous studies have suggested that although localised, sex-biased dispersal does not reduce the encounter rate between opposite-sexed kin enough to eliminate incest, this may facilitate the evolution of kin discrimination during mate choice, by ensuring that some non-kin are available as potential partners (Blyton *et al.* 2015).

3.4.3 Inbreeding avoidance

This study demonstrates very effective avoidance of inbreeding in long-tailed tits, despite the substantial risk. Although relatives tend to be clustered, the mean relatedness between male and female breeders within 300m is just $r_{QG} = 0.038 \pm 0.003$ (mean ± SE, n = 3947 dyads). This means that even if mates were selected randomly, the probability of incest would be fairly low. Yet, I found strong evidence of kin discrimination during mate choice, consistent with my finding that inbreeding carries fitness costs. But, the magnitude of inbreeding depression required to drive this mechanism also depends on the costs of inbreeding avoidance. Active mate choice is considered costly because the time taken searching for a suitable partner delays reproduction (Kokko & Ots 2006). This is important in species with temporal constraints on mate choice and/or a reliance on ephemeral resources. In long-tailed tits, breeding phenology is synchronised with caterpillar abundance, and offspring that fledge earlier in the season are more likely to survive (MacColl & Hatchwell 2002). Therefore, the length of the breeding season is strictly constrained (Hatchwell *et al.* 2013). However, long-tailed tits spend the non-breeding season in winter flocks containing kin and non-kin (Hatchwell *et al.* 2001a) and immigrants continually move into these flocks throughout the non-breeding season (Napper & Hatchwell 2016), so the opportunity to find partners is less constrained and avoiding kin as mates need not delay reproduction.

The second hypothesised cost of kin discrimination during mate choice is the evolution of a suitable mechanism. In the absence of spatial cues, kin recognition requires learning phenotypic cues to kinship by association, or an assessment of genetic similarity based on matching external phenotypic cues to an internal template (Reeve 1989; Sherman *et al.* 1997). Long-tailed tits have an effective means of kin recognition in cooperative contexts, permitting kin-biased helping when spatial cues are unreliable (Russell & Hatchwell 2001; Sharp *et al.* 2005). The same mechanism may be co-opted to enable inbreeding avoidance. Together, these behavioural traits could drive the evolution of kin recognition during mate choice even when the risk of incest is relatively low.

Our analyses show that the avoidance of close kin is sufficient to minimise inbreeding in long-tailed tits. This is consistent with previous studies demonstrating recognition of first-order kin (Leedale *et al.* 2018; Russell & Hatchwell 2001; Sharp *et al.* 2005), based on associative learning (Komdeur & Hatchwell 1999; Sharp & Hatchwell 2006). However, other studies suggest long-tailed tits can discriminate between kin of varying relatedness (Nam *et al.* 2010), which would require phenotype matching. Although phenotype matching is an attractive potential mechanism of inbreeding avoidance, because it allows individuals to recognise unfamiliar kin, association is the most likely mechanism of kin recognition in cooperatively breeding birds (Curry 1988; Hatchwell *et al*. 2001b; Komdeur 2004) and support for phenotype matching remains inconclusive, having been shown only in bell miners *Manorina melanophrys* (McDonald & Wright 2011). In our mate choice models, removing second-order kin as potential partners reduces the expected level of pair relatedness below that observed, suggesting long-tailed tits do not avoid second-order kin as mates. This is supported by previous work that shows that the proportion of help given to second-order kin is no different to the random expectation (Leedale *et al.* 2018).

On the other hand, just two cases of breeding between second-order kin were identified from the social pedigree. However, identifying kin with $r < 0.5$ from social pedigrees can be problematic because there are a variety of relationships that can generate this level of kinship and detecting all possibilities requires more detailed and extensive family histories than are available for most individuals. In contrast, I identified 96 breeding pairs with relatedness estimates with r_{QG} values ranging from

0.125-0.375, suggesting that pairing with distant kin does occur regularly. In cooperative species, the fitness costs associated with low rates of moderate inbreeding may be outweighed by the benefits of associating with kin post-dispersal (Jamieson *et al.* 2009; Puurtinen 2011).

The observed relatedness of breeding pairs closely matches the pattern expected by females avoiding first-order kin, while males travelling within 300m to find a partner appear to do slightly better than expected in terms of inbreeding avoidance. Philopatry is male-biased, and males are more likely to encounter opposite-sex kin than females travelling the same distance to find a mate. As their risk is slightly greater, and because males are much more likely to become helpers than females are (Hatchwell *et al.* 2004), it is possible that selection for kin recognition is stronger in males, particularly if this mechanism initially evolved for the kin-selected benefits of helping. However, observed relatedness is only slightly lower than that expected by avoidance of first-order kin and not low enough to suggest they are able to recognise more distant kin (Fig. 3.5a). In most animals, females invest more in a single reproductive event than males (Trivers 1972). Even in species with egalitarian parental care, females still invest more in egg production and incubation or gestation. Consequently, females are more selective than males when choosing partners (Bateman 1948). This may explain why the observed pattern of inbreeding closely matches female-focused mate choice models, and the apparent superior kin avoidance by males may in fact be female-driven. Indeed, theory predicts that females should avoid inbreeding more than males because of unequal reproductive investment (Parker 1979; 2006). A study on red junglefowl *Gallus gallus* (Pizzari *et al.* 2004) found that females retained fewer sperm after mating with siblings, whereas males invested more sperm in incestuous copulations when unrelated females were unavailable, supporting sex-specific inbreeding strategies that reflect asymmetry in parental investment.

3.4.4 Summary

In summary, dispersal patterns, demography and life-history traits interact to influence inbreeding strategy in natural populations. This study demonstrates that in long-tailed tits, inbreeding has deleterious fitness consequences and is alleviated by the avoidance of close kin as partners. Active mechanisms of kin discrimination that arise through kin-selection may be co-opted to effectively avoid inbreeding in systems where kin associations persist after dispersal. Further research on the mechanisms of kin recognition is required, but the pattern of avoidance indicates one similar to that identified previously (Sharp *et al.* 2005) and is likely to involve associative learning based on vocal cues.

Chapter 4

Bioacoustic analysis of call variation in long-tailed tits

Summary

Vocalisations have been identified as the most common recognition cues in cooperatively breeding birds. In order for vocal recognition cues to be reliable, they must have greater inter-individual than intra-individual variation. In this chapter, bioacoustic techniques are used to investigate acoustic variation in the contact calls of long-tailed tits *Aegithalos caudatus*, a cooperative breeder known to use contact calls to recognise kin. I devise and implement suitable bioacoustic methods to accurately measure call similarity at the population level, in order to further investigate the role of call similarity in kin recognition. I assess two methods commonly used to measure acoustic similarity; spectrographic cross-correlation (SPCC) and dynamic time-warping (DTW). DTW and SPCC showed similar effectiveness in describing acoustic variation within and between individuals. However, DTW is better able to cope with background noise and is computationally faster than SPCC. Repeatability tests based on specific acoustic parameters identified frequency characteristics as the most important source of individuality in long-tailed tit calls. These bioacoustic analyses create a baseline with which to investigate how call similarity correlates with relatedness, and facilitate the selection of meaningful sound parameters to further investigate kin recognition mechanisms in this species.

4.1 Introduction

In most social animals, cooperation occurs before natal dispersal when the relatedness between cooperating individuals is predictably high (Hamilton 1964). In a small number of cooperative species, cooperation occurs post-dispersal, across extended networks of conspecifics varying in their relatedness (Dickinson & Hatchwell 2004). In such groups, the ability to recognise close relatives is crucial for maximising fitness (Komdeur & Hatchwell 1999; Griffin & West 2003; Cornwallis *et al.* 2009). Prolonged associations between opposite-sex relatives during adulthood can also increase inbreeding risk (Koenig & Haydock 2004), necessitating active kin discrimination during mate choice. Effective recognition mechanisms that allow individuals to assess the relatedness of conspecifics are therefore likely to be under strong selection in such species.

Recognition mechanisms involve the production of external cues, the perception and comparison of these cues to an internal template, and the performance of discriminative behaviour depending on the perceived similarity between the cue and the template (Reeve 1989; Gamboa *et al.* 1991; Sherman *et al.* 1997). Vocalisations are the most common recognition cues in cooperatively breeding birds (Sharp & Hatchwell 2005; 2006; Açkay *et al.* 2013). Auditory signals can penetrate through or around obstacles and can be broadcast over long-distances (Catchpole & Slater 2008). Vocalisations are also metabolically relatively cheap to produce (Franz & Goller 2003; Ward *et al.* 2004) and large amounts of information can be transferred very quickly. These properties make vocalisations a highly efficient mechanism of communication in most bird species.

Bird vocalisations are traditionally grouped into songs and calls. Songs are generally long, complex sounds, which are, at least in part, sexually-selected traits involved in mate attraction and territory defence (Collins 2004). Calls tend to be shorter and simpler, and may serve a various functions, including maintaining group cohesion (Radford 2005), or alerting conspecifics to the presence of specific predators (Evans *et al.* 1993). There is evidence of both songs and calls also being used as recognition cues (Payne *et al*. 1988; Price 1999; Sharp *et al*. 2005). Although there is huge variation in complexity, all vocalisations can be defined as a series of acoustic units occurring together in a specific pattern. The structure is hierarchical, with the smallest units, elements, joining to form syllables, which themselves join to form phrases. Although vocalisations are species-specific (Marler 2004), most birds have more than one version of the same song or call. The number of different versions, or types, an individual can produce is called a repertoire. The hearing capabilities of most birds is between 1-5kHz and they are most sensitive at 2-3kHz (Catchpole & Slater 2008). Although the frequency range of sound production varies greatly, it is predicted to closely match the audibility range (Catchpole & Slater 2008).

Bioacoustics is the field of biology dedicated to analysing animal sounds. Analogue sound waves are digitised by an A/D converter, which converts a continuous signal into a numeric sequence by taking measurements at discrete time intervals. These continuous measurements are then converted to discrete integers (quantization). The rate at which these measurements are taken is called the sampling frequency (Clements 1998). The number of different integers that can be assigned to each continuous measurement is called the sample size (Clements 1998), and is measured in binary digits (bits). Therefore, the precision with which the original acoustic signal is digitally represented is determined by the sampling frequency and the sample size

(Bradbury & Vehrencamp 1998). Digitised sounds are graphically represented in a spectrogram; a plot of frequency against time with the darkness of the plot representing the amplitude (Catchpole & Slater 2008). Spectrograms are produced using fast Fourier transformation (FFT); the resolution of the image depends on the length of the FFT window. Thus, FFT length also affects the precision with which the signal is represented. Spectrograms are the standard way to illustrate animal sounds and an understanding of these principal techniques is important in all bioacoustic analyses.

In this chapter, bioacoustic techniques will be used to investigate how vocalisations may be used as kin recognition cues in cooperatively breeding long-tailed tits *Aegithalos caudatus*. Bird vocalisations have been studied in a variety of recognition contexts, including parent-offspring interactions (Beecher 1991), species recognition (Wolfenden *et al*. 2015), mate recognition (Berg *et al*. 2011), territorial contests (Fisher 1954) and kin or group-member recognition (Radford 2005; McDonald & Wright 2011). In long-tailed tits, vocalisations play a major role in in the coordination of cooperative behaviour (Hatchwell *et al.* 2001b; Sharp *et al.* 2005). In order to understand how vocalisations may be used as kin recognition cues, it is important to understand how calls vary within and between individuals. Effective recognition cues must have greater inter-individual than intra-individual variation (Sharp *et al*. 2006). That is, they must be highly stereotyped within individuals and variable between individuals, so that they convey information about identity. To achieve this, they are likely to be made up of several different trait components that vary in their pattern or combination profile from individual to individual. This specific complex of phenotypic traits is referred to as a *signature system* (Beecher 1982). Long-tailed tits have a small vocal repertoire of five main call types, but no territorial song (Cramp & Perrins 1993). Of these calls, previous studies have demonstrated individuality in two long-tailed tit contact calls: the *churr* and the *triple*. Furthermore, Sharp *et al.* (2005) demonstrated that individuals can recognise kin using the churr alone, and that these calls are learned during development. To investigate call similarity as a cue to kinship, this project will involve comparisons of the churr and the triple across individuals of varying relatedness. It is therefore important to first determine the consistency of these calls within individuals over time. These bioacoustic analyses will create a baseline of variation in long-tailed tit contact calls with which to investigate how call similarity correlates with relatedness.

I aim to devise and implement suitable bioacoustic methods to investigate call variation in long-tailed tits, focusing on churr and triple calls. The churr call is a short range contact call, often used by breeding adults and helpers at the nest. The triple call is a longer range contact call, and is used more often during group foraging, and by females during incubation to alert males that she has left, or is returning to the nest. Both calls are single-phrased; the churr call is made up of an initial syllable of one or two unique elements, followed by a second syllable comprising a single element that may be repeated several times. The triple call is a monosyllabic, comprising a single element that may be repeated several times. This distinction between elements and syllables is somewhat arbitrary and is used here purely to capture any differences in variation in the two distinct sections of the churr call specifically. Temporal variation relative to the rest of the call is often a good indicator of whether to split call units into elements or syllables (Catchpole & Slater 2008). Therefore, the separate syllables in the churr call can be viewed spectrographically as separated by a larger gap than the gap separating the two elements in the first syllable (Fig. 4.1).

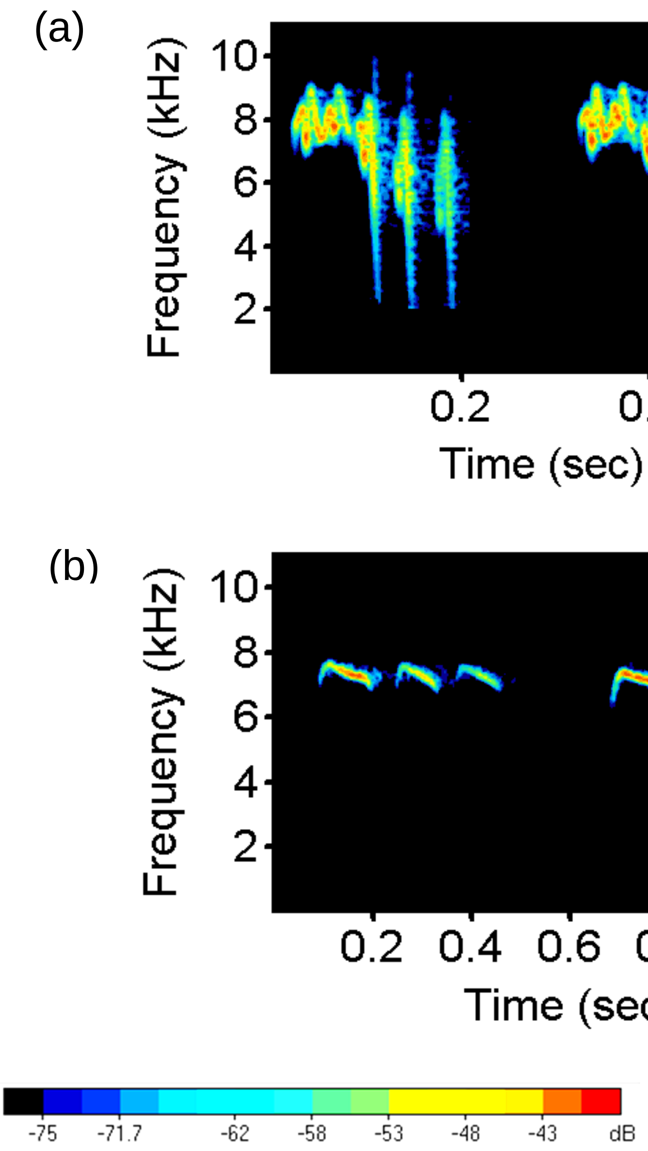

Figure 4.1. Spectrogram of the long-tailed tit (a) *churr* call and (b) *triple* call. The churr call presented here is made up of two syllables, each comprising two elements. The triple call presented here is monosyllabic, comprised of three elements. The colour of the signal represents the relative amplitude, or energy across the call.

Initially, I assess two methods which describe how the overall acoustic structure of calls vary within and between individuals, by comparing the pairwise similarity coefficients between calls made by the same individual, and calls made by different individuals. After selecting the best method to compare call structure, I then quantify the amount of variation in churr and triple structure within and between adult long-tailed tits using a larger dataset. I also aim to determine which sound parameters explain this variation, by analysing the individual repeatability of calls based on multiple acoustic measurements. Certain call characteristics may differ in the extent to which they are individually repeatable and characteristics with particularly high repeatability may be particularly important for recognition. This bioacoustic analysis facilitates the selection of meaningful sound parameters for subsequent analysis of call similarity among relatives, in order to further investigate kin recognition mechanisms.

4.2 Methods

4.2.1 Study site and field methods

Fieldwork was carried out on a population of 31-46 breeding pairs of long-tailed tits in a section of the Rivelin Valley, Sheffield, UK (53º38'N 1º56W) from 2015-2017. The site is approximately 2.5km^2 and comprises a variety of suitable breeding habitat; predominantly deciduous woodland and scrub. The site also encompasses areas of open pasture, gardens and a golf course. The Rivelin Valley long-tailed tit population has been studied extensively, with an approximate breeding population of 25-72 pairs since data collection began in 1994. Almost all individuals (>95%) are marked with a metal BTO ring and a unique combination of two colour rings for field identification.

During each breeding season, which runs from March-June, all nests were located and breeding pairs were monitored closely. Any unringed adults immigrating into the study site were caught using mist nets and ringed. At this time, individuals were weighed and morphometric measurements were recorded. A sample of 5-30µl of blood was taken by brachial venepuncture under Home Office licence. Blood samples were genotyped at 17 polymorphic microsatellite loci to determine relatedness, as well as two sex markers. All nesting attempts were monitored every two days until fledging or nest failure. The onset of egg-laying, incubation period, hatching date and fledging or failure date were recorded. In the case of nest failure, the cause of failure and predation agent was recorded, and the study site was searched intensively for re-nesting attempts. The majority of long-tailed tit nests are built 1-2m above ground in gorse, bramble or similar shrubs, but a small proportion of nests (~10%) are built high in deciduous and coniferous trees and were inaccessible. For accessible nests, clutch size and the timing of breeding events were recorded directly by checking nests. For inaccessible nests, data was collected indirectly by behavioural observation of breeding activity. Nestlings were removed from accessible nests at 11 days old, ringed and processed in the same way as unringed adults before being replaced. All nests were watched for at least one hour every two days during the nestling period to record the identity of parents and helpers. For all visible nests, provisioning rates were also recorded. Nest locations were recorded using GPS with an accuracy of 8m.

4.2.2 Sound recordings

Adult vocalisations were recorded using a *Sennheiser ME67/K6* shotgun microphone fitted with a Rycote standard windjammer. Recordings were made onto a *Roland R-05 version 1.03* WAV/MP3 recorder with a 6GB *SanDisk* memory card, set to a sample rate of 48kHz with WAV-16bit accuracy. The input settings were kept as thus: mono, required when using one external microphone; limiter off, to ensure high volume sounds were not cut out; high mic gain with an input level of 60db, to optimise the sensitivity of the microphone; and a low-cut frequency of 400Hz, to cut out extreme low frequency noise. Each year, sound recording began once the majority of first-attempt nests had been located. All recordings were made in approximately similar conditions between 06:00 and 18:00 BST. Birds were recorded at a distance of approximately 3-15m, to minimise the effects of sound degradation and reverberation. Birds were recorded at the nest and identified by their unique colour ring combinations. If more than one bird was present, vocalisations could be assigned to individuals by observing movements of the bill and throat feathers. All recordings were made during the breeding season because outside the breeding season, long-tailed tits form large, noisy flocks, making it difficult to identify the caller with certainty. At the start of each recording, date, time, nest number and recording number were spoken into the microphone. Bird ID was dictated after each call. Bird ID is a three or four letter acronym based on the colour ring combination and orientation. For example, an individual with a white ring above a pink ring on the left tarsus has the ID: *WPL*. Only when the calling individual could be identified with absolute certainty was a recording of a call used. Most churr recordings were made during the nest-building and provisioning stages of breeding, when breeding adults are the most vocal. Triple calls are more often made by females during incubation. In total, 213 recordings were made, containing 1116 churr calls from 98 individuals (mean ± SD = 11.39 ± 10.24 per bird; range 1 – 42) and 400 triple calls from 57 individuals (mean ± SD = 7.02 ± 5.99 per bird; range 1 – 23).

4.2.3 Bioacoustic analysis

All recordings were digitized with 16-bit accuracy at a sampling rate of 48kHz, matching the recorder settings. Spectrograms were produced in *Avisoft SAS-Lab Pro version 4.52* (Avisoft Bioacoustics, Raimund Specht, Berlin, Germany) using a 256-point FFT length with a Hamming window, 100% frame size and 50% window overlap. This gave a frequency resolution of 188Hz and a time resolution of 2.7ms. The sampling frequency was converted to 22.05 KHz to make the recordings more manageable and the calls easier to visualise spectrographically. If the sampling frequency is too low, distortion of frequency components, or aliasing, can occur. The highest frequency that can be digitised without aliasing is called the Nyquist frequency, and is equal to half the sampling frequency (Bradbury & Vehrencamp 1998). Therefore, although higher sampling frequencies can provide more information, the optimum sampling frequency is usually considered to be twice that of the maximum frequency of the signal. As long-tailed tit calls range from 2-9kHz, re-sampling the recordings to 22.05kHz does not affect the signal. All recordings were analysed spectrographically to assess call quality. Only calls which were followed immediately by bird ID were considered. Some background noise was removed by setting a high-pass filter of 1.5 kHz. Recordings with high levels of background noise or calls masked by other vocalisations were excluded from the analysis.

In order to maximise the amount of useable data whilst still ensuring robust results, a subset of these data were tested for individual repeatability. If long-tailed tit calls are repeatable, all calls can be used in subsequent analysis. To determine the minimum number of calls required to capture individual variation, the cumulative repertoire size (number of distinct calls, based on number of syllables) was plotted against the number of calls considered to that point, for 100 churr calls, 10 from each of 10 birds recorded on at least two days in 2015. The resulting plots generally levelled off before the number of calls reached six (mean ± SD calls needed to reach asymptote = 5.5 ± 2.89, range 2-10). Therefore, repeatability tests were carried out on all calls from individuals with recordings of at least six calls from at least two days between 2015-2017: 943 churr calls from 54 individuals (mean ± SD = 17.46 ± 10.02 per bird; range 6 – 42) and 263 triple calls from 22 individuals (mean ± SD = 11.95 ± 5.38 per bird; range 6 – 23). Within-individual repeatability was tested using two approaches. The first approach compared within and between-individual variation in overall call structure using spectrographic cross-correlation (SPCC) and dynamic time-warping (DTW). The second tested the repeatability of specific vocal characteristics, to determine which are responsible for call variation. Those with the highest repeatability value may be used for recognition and will be considered in future analyses.

4.2.4 Variation in acoustic structure

Spectrographic cross-correlation (SPCC) was devised by Clark *et al*. (1987) as an objective method to quantitatively measure the degree of similarity between spectrograms. Rather than assessing sounds based on acoustic parameters selected *a priori,* SPCC compares the overall sound structure, by integrating all parameters and weighing them equally (Janik 1999). SPCC cross correlates spectrograms frame-by-frame, calculating a correlation coefficient at each point of overlap. The peak coefficient for every pairwise comparison of spectrograms is stored in a matrix. The peak value of a spectrogram correlation provides a measure of one type of similarity between two signals, with a higher peak correlation equivalent to greater similarity. Single calls were converted from WAV files in to ASCII coded text files using *Avisoft SAS-Lab Pro* and cross-correlations were carried out using *Avisoft Correlator version 3.1 (*Avisoft Bioacoustics, Raimund Specht, Berlin, Germany*).*

Dynamic time warping (DTW) is a distance-based programming technique used to search for an optimal alignment of two time-series, or signals. The algorithm calculates a distance score between two signals based on certain acoustic features (Lachlan 2007), with greater distance meaning lower similarity. This has been implemented for use in bioacoustics in the program *Luscinia version 2.16.10.29.01* (Robert Lachlan, Queen Mary University of London, UK). All calls were uploaded in to a library in Luscinia in WAV format and the signal of interest was highlighted from the sonogram and multiple acoustic measurements taken (Fig. 4.2). The acoustic features used to calculate DTW distance score were selected and weighted as follows: Time = 1, fundamental frequency = 2, change in fundamental frequency = 1, change in peak frequency = 1 (Fig. 4.2, Table 4.1). These settings generated a DTW algorithm which correctly matched visually similar calls, assessed in Luscinia using a dendrogram and multidimensional scaling plot.

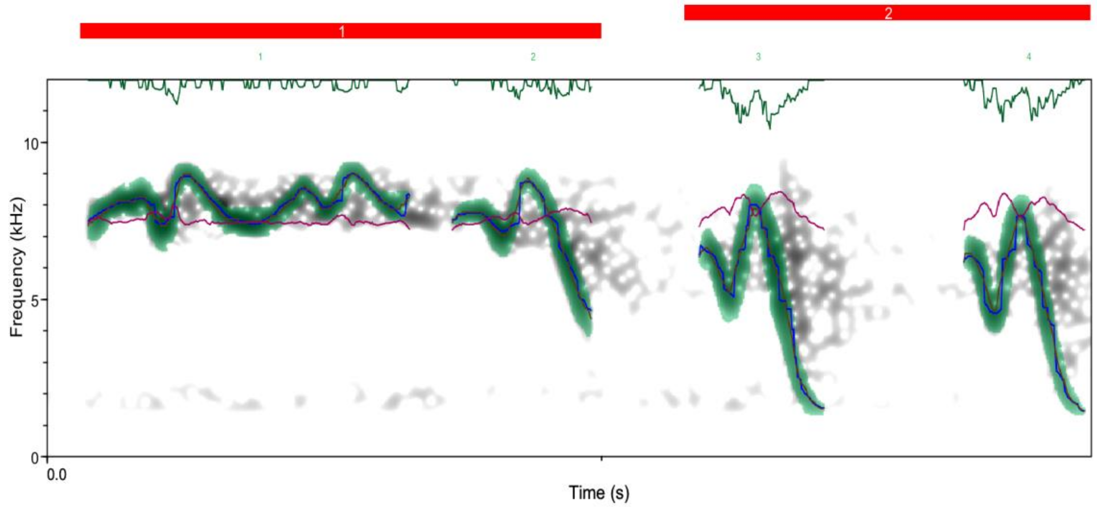

Figure 4.2. Acoustic parameters of the long-tailed tit *churr* call, measured using Luscinia (Lachlan, 2007). The churr call comprises two syllables (red bars) of multiple elements (green signal). Signals are traced from the sonogram semi-automatically and parameters such as: fundamental frequency (brown), peak frequency (blue), Weiner entropy (purple) and bandwidth (dark green), are measured automatically and saved into the database for analysis.

Table 4.1. Definitions of the acoustic parameters used to analyse long-tailed tit calls. Frequency parameters are the mean of multiple measurements taken at five time points across the signal.

Call	**Acoustic Parameter**	**Definition**
Churr	Total call duration (s)	Time between start and end of the signal.
	Element gap/length	Ratio of between-element gap to total length of elements in the signal.
	Fundamental frequency (Hz)	Frequency of pure-tone signals and common denominator frequency of harmonic signals.
	Peak frequency (Hz)	Frequency of maximum intensity.
	Max. fundamental frequency (Hz)	Maximum fundamental frequency across the signal.
	Max. peak frequency (Hz)	Maximum peak frequency across the signal.
	Change in peak frequency (Hz)	The change in peak frequency across the signal.
	Change in fundamental frequency (Hz)	The change in fundamental frequency across the signal.
	Weiner Entropy (pure number)	Uniformity of the signal (noisiness).
	Bandwidth (Hz)	Frequency difference between the first and final maximum intensity of the signal.
	S1 duration (s)	Time between start and end of syllable 1.
	S1 fundamental frequency (Hz)	Fundamental frequency of syllable 1.
	S1 peak frequency (Hz)	Peak frequency of syllable 1.
	S1 max. fundamental frequency (Hz)	Maximum fundamental frequency across syllable 1.
	S1 max. peak frequency (Hz)	Maximum peak frequency across syllable 1.
	S2 duration (s)	Time between start and end of syllable 2.
	S2 fundamental frequency (Hz)	Fundamental frequency of syllable 2.
	S2 peak frequency (Hz)	Peak frequency of syllable 2.
	S2 max. fundamental frequency (Hz)	Maximum fundamental frequency across syllable 2.
	S2 max. peak frequency (Hz)	Maximum peak frequency across syllable 2.
	S2 repeats	Number of element repetitions in syllable 2.
	S2/S1 duration (s)	Ratio of syllable 2 length to syllable 1 length.
Triple	Total call duration (s)	Time between start and end of signal.
	Element gap/length	Ratio of between-element gap to total length of elements.
	Element repeats	Number of elements in signal
	Fundamental frequency (Hz)	Frequency of pure-tone signals and common denominator frequency of harmonic signals.
	Peak frequency (Hz)	Frequency of maximum intensity.
	Max. fundamental frequency (Hz)	Maximum fundamental frequency across signal.
	Max. peak frequency (Hz)	Maximum peak frequency across signal.
	Weiner Entropy (pure number)	Uniformity of the power spectrum (noisiness).
	Bandwidth (Hz)	Frequency difference between the first and final maximum intensity of the signal.
	Change in peak frequency (Hz)	The change in peak frequency across the signal.
	Change in fundamental frequency (Hz)	The change in fundamental frequency across the signal.

SPCC and DTW analyses were first carried out on a subset of calls from 2015; churr: n = 13 birds, 143 calls (mean ± SD = 11 ± 2.48 per bird; range 6 – 15), triple: n = 3 birds, 34 calls (mean ± SD = 11.33 ± 5.13 per bird; range 7 – 17). Pairwise comparisons of individual calls generated a matrix of correlation coefficients and a matrix of distance scores for each pair of calls in the subset. To compare call similarity within and between individuals, the matrices were converted into a pairwise comparison table and each row was assigned a value according to whether the comparison was made within individuals (0) or between individuals (1). The correlation coefficients and distance scores were aggregated and mean call similarity within and between individuals was compared. The most suitable method was then used to determine the degree of call variation within and between individuals, with a larger subset of data from 2015-2017 (churr: n = 54 birds, 907 calls, triple: n = 23 birds, 246 calls). As this analysis contains calls from across years, the measures of call similarity were also compared within and between years.

4.2.5 Repeatability of call parameters

In order to assess which sound parameters explain variation in call structure, further analysis compared the subset of data from 2015-2017 (churr: n = 907 calls from 54 birds, triple: n = 246 calls from 23 birds) based on multiple acoustic measurements. For every call, a range of acoustic parameters were measured in Luscinia (Fig. 4.2, Table 4.1). Amplitude parameters were excluded from the analysis, as amplitude is strongly influenced by external factors such as background noise, elevation and distance (Klump 1996). As well as being difficult to control for in the field, the influence of external factors on amplitude may also make call amplitude unreliable for recognition (Catchpole & Slater 2008). Each call was measured hierarchically: by element, by syllable and by call. Not all levels of resolution may be appropriate for analysing call similarity in the context of recognition; it is unlikely that long-tailed tits assess calls element by element. However, for the churr call, it is possible that one syllable is more individually distinctive, and another more context dependent. Therefore, certain characteristics were tested for repeatability within churr call syllables (Table 4.1).

4.2.6 Statistical analysis

Statistical analysis was carried out using *R version 3.5.0* (R Core Team 2018). Overall similarity in call structure within and between individuals was compared using generalised linear mixed models (GLMM), with either the *SPCC correlation coefficient* or the *DTW distance score* as the dependent variable, comparison types (*within or between individuals* and *within or between years*) as fixed effects, *bird ID 1* and *bird ID 2* as nested random effects, and *year 1* and *year 2* as nested random effects. Having ID as a random effect controls for pseudoreplication in interrelated data, where values are derived from comparing all possible pairs of a set of individuals. The effect of year can also be quantified. GLMMs allow for large differences in sample size. This is necessary in such analyses where there are many more between individual than within individual comparisons. Correlation coefficients and distance scores were analysed using GLMMs fitted by maximum likelihood (Laplace approximation) and the Gamma error family with a log link function in the *R* package, *lme4* (Bates *et al.* 2008). Repeatability tests were carried out to test for individual consistency and to identify specific call characteristics with particularly high repeatability. Individual repeatability based on call parameters was carried out using multiple GLMMs in the *R* package, *rptR* (Nakagawa & Shielzeth 2010; Shielzeth *et* *al*. 2011). The repeatability models were built with each sound parameter as the dependent variable, *sex* as a fixed effect and *bird ID* and *year* as random effects. *Year* and *ID* were set as grouping variables, allowing for effects of year and ID to be tested. Gaussian models were used to test the repeatability of continuous variables and Poisson models were used to test the repeatability of count variables. To test for sex differences in call characteristics, GLMMs were built with each sound parameter as the dependent variable, *sex* as a fixed effect and *bird ID* and year as random effects, with the *lmerTest* function (Kuznetsova *et al.* 2017) to report the significance of variation attributed to fixed effects.

4.3 Results

4.3.1 Variation in acoustic structure

To determine the best method to analyse within-individual repeatability of long-tailed tits calls, SPCC and DTW were carried out on a subset of calls from individuals with recordings of at least 6 calls from at least two days in 2015 (churr: n = 13 birds, 143 calls; triple: n = 3 birds, 34 calls). Visual inspection of spectrograms suggested that churr and triple calls from the same individual were more similar in acoustic structure than those of different individuals (Fig. 4.3). The mean correlation coefficient from SPCC for within-individual comparisons was significantly higher than that for between-individual comparisons of churr calls (GLMM, estimate = 0.27 ± 0.04, residual df = 10148, t = 6.3, $p < 0.001$, Fig. 4.4a) and triple calls (GLMM, estimate = 0.4 ± 0.19, residual df = 556, t = 2.12, $p < 0.05$, Fig. 4.4b). The mean distance score from DTW for within-individual comparisons was lower than that for between-individual comparisons of both churr calls (GLMM, estimate = -0.33 ± 0.08, residual df = 10726, t = -3.98, p < 0.001, Fig. 4.5a,) and triple calls (GLMM, estimate = -0.68 ± 0.28, residual df = 556, t = -2.46, p < 0.05, Fig. 4.5b).

The DTW and SPCC analyses showed similar effectiveness for describing variation in the acoustic structure of churr and triple calls. However, DTW may be more suitable than SPCC for measuring call similarity for several reasons. Firstly, as the DTW analysis is carried out on signals selected from the sonogram, rather than ASC11 codes converted from raw WAV files, background noise is not incorporated into the analysis. Secondly, the sonogram library, signal measurements and DTW are stored and computed within the Luscinia program, providing an organised method of bioacoustic analysis which can be visualised and updated easily. Furthermore, DTW is computationally fast. To analyse a very large number of calls using SPCC may take several weeks, but may take less than an hour using DTW. Finally, the DTW is flexible, allowing the user to modify the DTW algorithm by manual selection and weighting of certain acoustic parameters based on previous assumptions or exploration of the data using dendrograms and multi-dimensional scaling plots.

Dynamic time warping was therefore selected as the most appropriate method for measuring variation in the acoustic structure of churr and triple calls within and between individual long-tailed tits. Within-individual repeatability of calls was analysed further using a larger subset of calls from individuals with recordings of at least 6 calls from at least two days in 2015-2017 (churr: n = 54 birds, 907 calls; triple: n = 23 birds, 246 calls). The mean distance score for within-individual comparisons was lower than that for between-individual comparisons of both churr calls (GLMM, estimate = -0.4 ± 0.04, residual df = 411770, t = -9.76, p < 0.001; Fig. 4.6a,) and triple calls (GLMM, estimate = -0.55 ± 0.09, residual df = 30620, t = -5.65, p < 0.001; Fig.

4.6b). Although an interaction between the years in which the calls being compared were recorded accounted for a small amount of the variation among calls (variance ± standard deviation = churr: 0.002 ± 0.05; triple: 0.015 ± 0.01), whether comparisons were made between calls recorded in the same or different years did not affect DTW distance score (churr: GLMM, estimate = -0.069 ± 0.05, df = 411770, t = -1.21, p = 0.23; triple: GLMM, estimate = -0.33 ± 0.2, df = 30620, t = -1.67, p = 0.1).

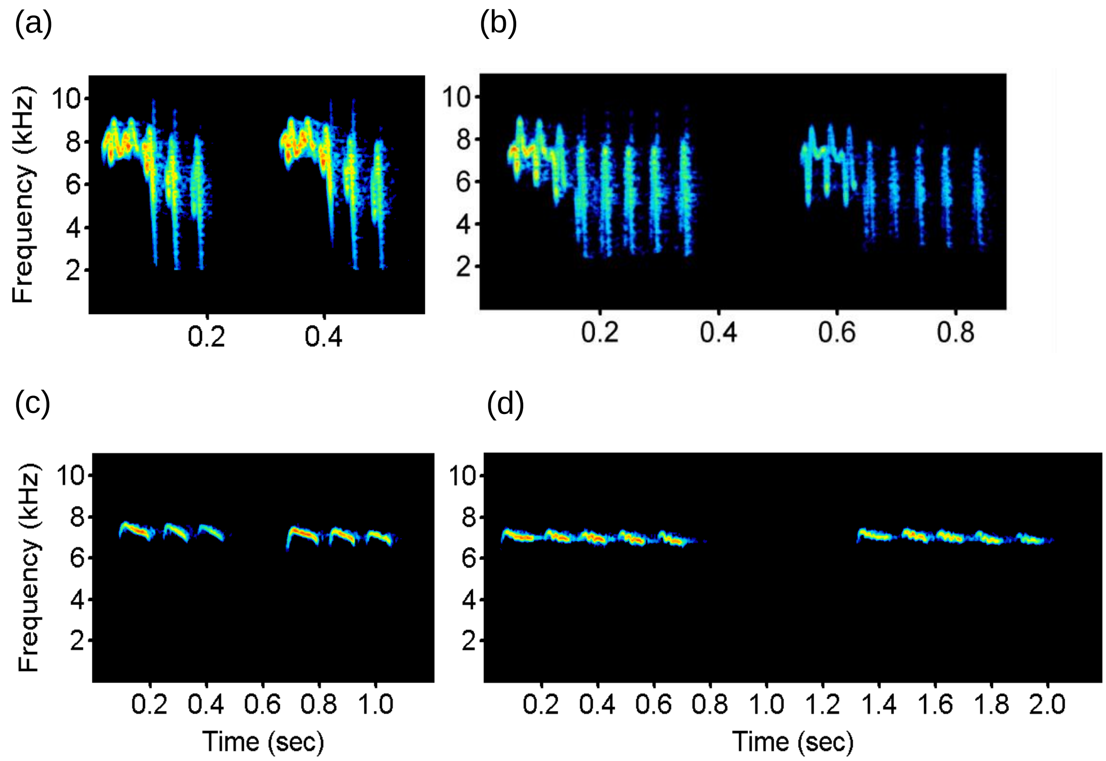

Figure 4.3. Spectrograms of two *churr* calls from each of two adult long-tailed tits: (a) RwdL and (b) NGR; and two *triple* calls from each of two adult long-tailed tits: (c) DDL and (d) YGR.

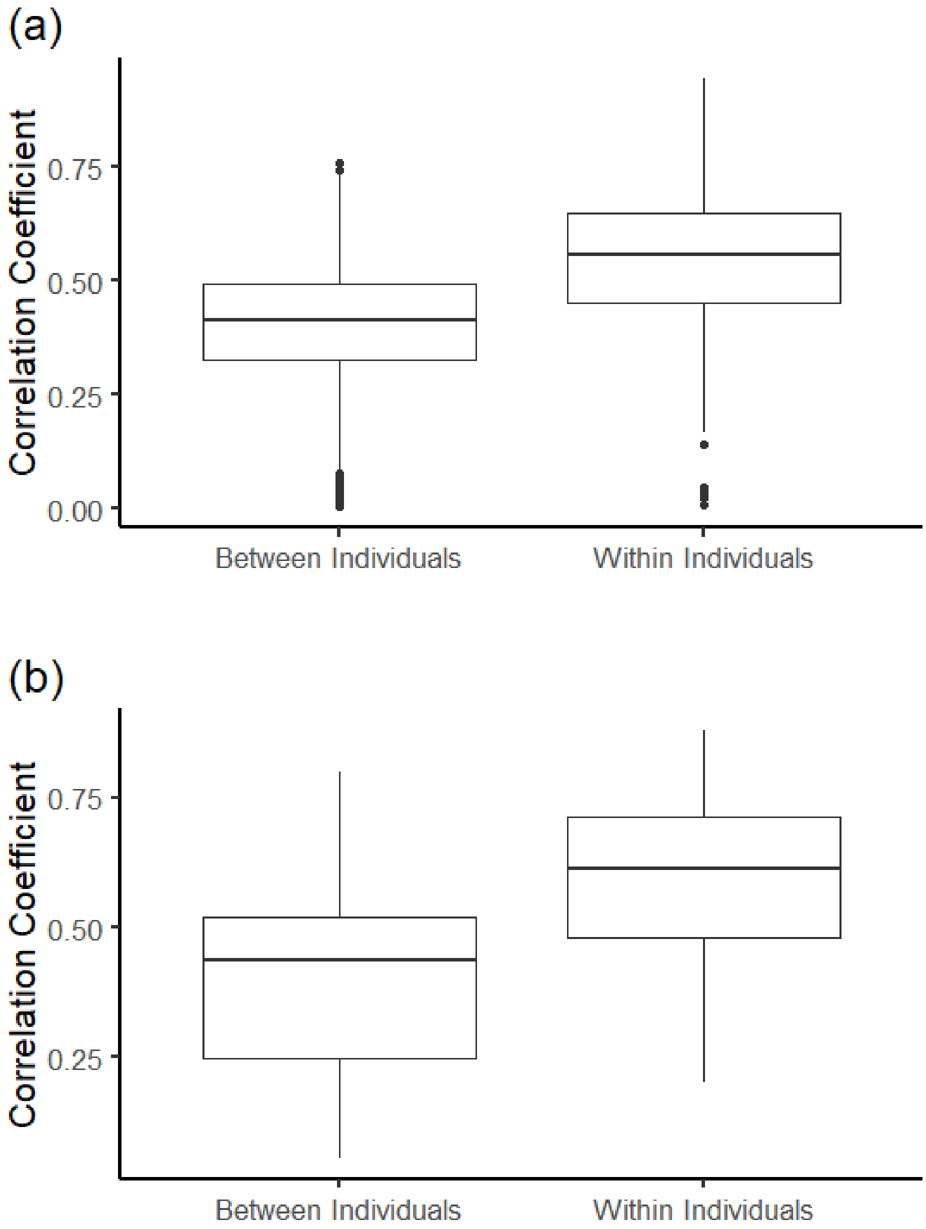

Figure 4.4. Similarity of (a) the *churr* call (n = 143 calls from 34 birds) and (b) the *triple* call (n = 34 calls from 3 birds) within and between individual long-tailed tits, measured in correlation coefficients generated by SPCC.

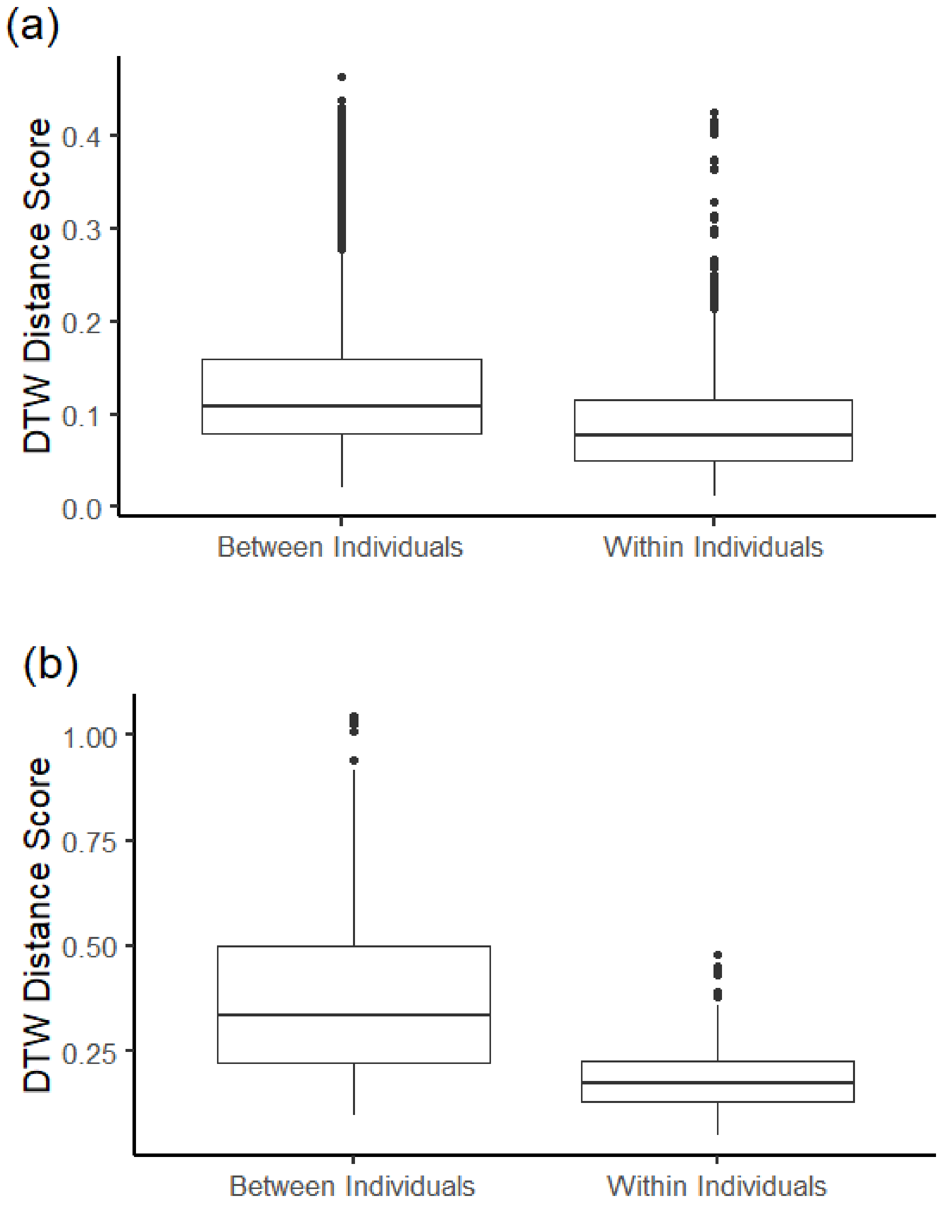

Figure 4.5. Dissimilarity of (a) the *churr* call (n = 143 calls from 34 birds) and (b) the *triple* call (n = 34 calls from 3 birds) within and between individual long-tailed tits, measured in distance score generated by DTW.

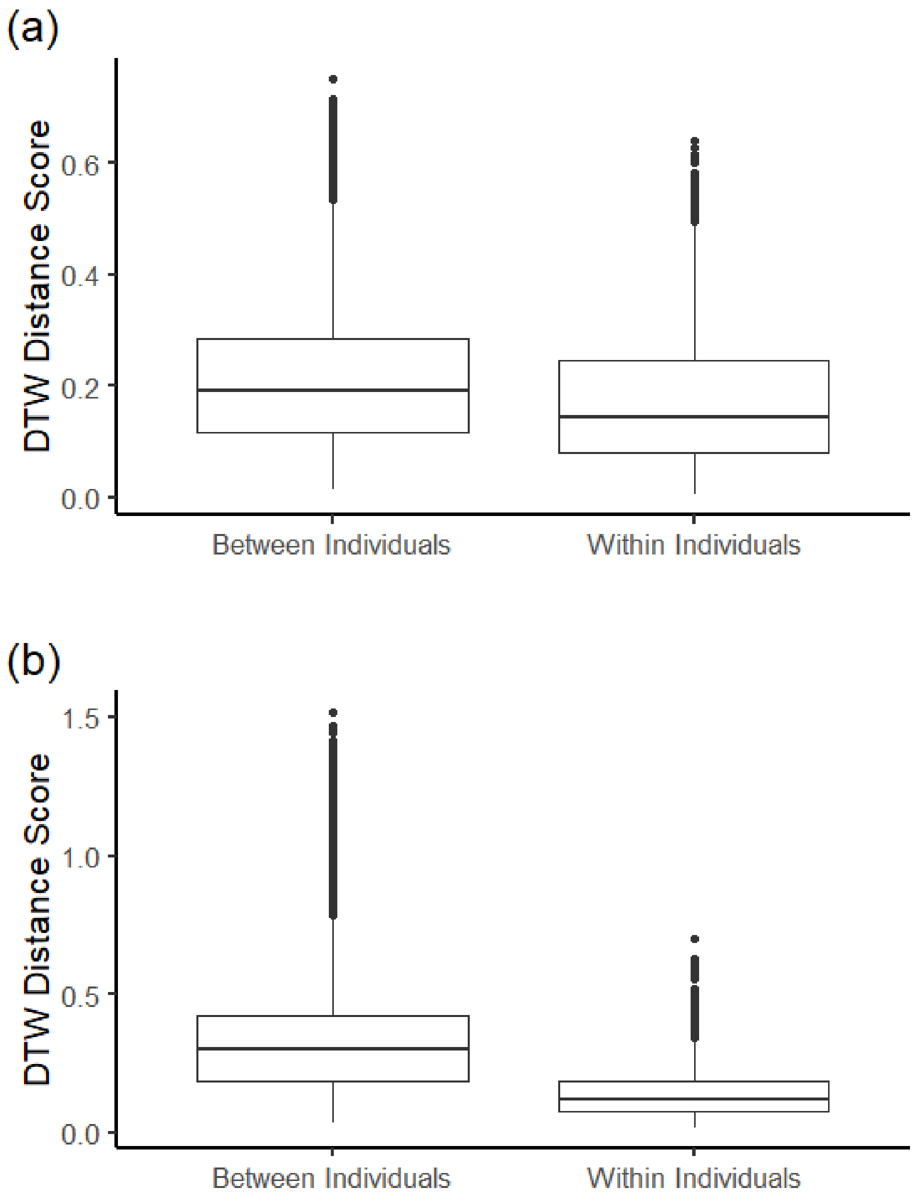

Figure 4.6. Dissimilarity of (a) the *churr* call (n = 907 calls from 54 birds) and (b) the *triple* call (n = 246 calls from 23 birds) within and between individual long-tailed tits, measured in distance score generated by DTW.

4.3.2 Repeatability of call parameters

Repeatability tests were carried out in order to quantify the amount of variation in churr and triple call parameters that could be attributed to the identity of the caller. This analysis used the subset of calls from 2015-2017 (churr: n = 54 birds, 907 calls, triple: n = 23 birds, 246 calls). Both the churr call and the triple call were repeatable within individuals based on all except one of the parameters tested, with frequency based parameters showing higher individual repeatability than temporal parameters (Table 4.2). Based on the frequency parameters, the first syllable of the churr call showed higher individual repeatability than the second syllable, with the frequency parameters of the triple call showing the highest individual repeatability overall. In both the churr and the triple, there was no effect of recording year (Table 4.3) or sex (Table 4.4) on any of the parameters.

Table 4.2. Repeatability of long-tailed tit call parameters based on caller identity (*churr* calls: n = 907 calls from 54 birds; *triple* calls: n = 246 calls from 23 birds).

Call	Parameter	R ± SE	CI	p
Churr	Total call duration (s)	0.33 ± 0.05	0.22, 0.42	< 0.001
	Syllable gap/length	0.25 ± 0.05	0.16, 0.34	< 0.001
	Syllable 1 duration (s)	0.39 ± 0.06	0.27, 0.48	< 0.001
	Syllable 2 duration (s)	0.36 ± 0.05	0.24, 0.46	< 0.001
	Syllable 2/syllable 1 duration (s)	0.32 ± 0.05	0.21, 0.41	< 0.001
	Syllable 2 repeats	0.07 ± 0.02	0.02, 0.09	< 0.001
	Fundamental frequency (Hz)	0.29 ± 0.05	0.19, 0.38	< 0.001
	Peak frequency (Hz)	0.27 ± 0.05	0.17, 0.35	< 0.001
	Max. fundamental frequency (Hz)	0.5 ± 0.07	0.35, 0.61	< 0.001
	Max. peak frequency (Hz)	0.5 ± 0.07	0.35, 0.61	< 0.001
	Bandwidth (Hz)	0.21 ± 0.05	0.13, 0.3	< 0.001
	Weiner Entropy	0.19 ± 0.04	0.11, 0.27	< 0.001
	Syllable 1 fundamental frequency (Hz)	0.45 ± 0.06	0.33, 0.55	< 0.001
	Syllable 1 peak frequency (Hz)	0.45 ± 0.06	0.31, 0.55	< 0.001
	Syllable 1 max. fundamental frequency (Hz)	0.49 ± 0.07	0.35, 0.6	< 0.001
	Syllable 1 max. peak frequency (Hz)	0.49 ± 0.07	0.35, 0.6	< 0.001
	Syllable 2 fundamental frequency (Hz)	0.31 ± 0.05	0.2, 0.41	< 0.001
	Syllable 2 peak frequency (Hz)	0.27 ± 0.05	0.18, 0.36	< 0.001
	Syllable 2 max. fundamental frequency (Hz)	0.45 ± 0.07	0.31, 0.58	< 0.001
	Syllable 2 max. peak frequency (Hz)	0.43 ± 0.08	0.28, 0.57	< 0.001
Triple	Total call duration (s)	0.31 ± 0.09	0.12, 0.47	< 0.001
	Syllable gap/length	0.2 ± 0.07	0.05, 0.35	< 0.001
	Element repeats	0 ± 0.01	0, 0.02	1
	Fundamental frequency (Hz)	0.69 ± 0.07	0.52, 0.78	< 0.001
	Peak frequency (Hz)	0.70 ± 0.09	0.48, 0.81	< 0.001
	Maximum fundamental frequency (Hz)	0.74 ± 0.08	0.56, 0.85	< 0.001
	Maximum peak frequency (Hz)	0.71 ± 0.08	0.51, 0.82	< 0.001
	Bandwidth (Hz)	0.2 ± 0.08	0.06, 0.35	< 0.001
	Weiner Entropy	0.22 ± 0.08	0.06, 0.37	< 0.001

Table 4.3. Within-year repeatability, or effect of year calls were recorded, on variation in long-tailed tit call parameters (n = 907 *churr* calls and 246 *triple* calls over 3 years).

Call	Parameter	R ± SE	CI	p
Churr	Total call duration (s)	0 ± 0.07	0, 0.06	1
	Syllable gap/length	0 ± 0.01	0, 0.04	1
	Syllable 1 duration (s)	0 ± 0.02	0, 0.07	1
	Syllable 2 duration (s)	0 ± 0.02	0, 0.06	1
	Syllable 2/syllable 1 duration (s)	0 ± 0.003	0, 0.01	1
	Syllable 2 repeats	0.02 ± 0.04	0, 0.12	0.37
	Fundamental frequency (Hz)	0.04 ± 0.05	0, 0.16	0.39
	Peak frequency (Hz)	0.03 ± 0.04	0, 0.14	0.35
	Max. fundamental frequency (Hz)	0.01 ± 0.7	0, 0.22	0.4
	Max. peak frequency (Hz)	0.07 ± 0.08	0, 0.27	0.4
	Bandwidth (Hz)	0.05 ± 0.05	0, 0.18	0.17
	Weiner Entropy	0 ± 0.01	0, 0.04	1
	Syllable 1 fundamental frequency (Hz)	0.01 ± 0.03	0, 0.11	0.45
	Syllable 1 peak frequency (Hz)	0.02 ± 0.04	0, 0.13	0.43
	Syllable 1 max. fundamental frequency (Hz)	0.06 ± 0.07	0, 0.25	0.38
	Syllable 1 max. peak frequency (Hz)	0.07 ± 0.08	0, 0.26	0.4
	Syllable 2 fundamental frequency (Hz)	0.01 ± 0.03	0, 0.1	0.49
	Syllable 2 peak frequency (Hz)	0.01 ± 0.02	0, 0.07	0.54
	Syllable 2 max. fundamental frequency (Hz)	0.1 ± 0.1	0, 0.32	0.37
	Syllable 2 max. peak frequency (Hz)	0.12 ± 0.11	0, 0.36	0.37
Triple	Total call duration (s)	0.04 ± 0.05	0, 0.16	0.18
	Syllable gap/length	0.06 ± 0.07	0, 0.25	0.08
	Element repeats	0 ± 0.01	0, 0.02	1
	Fundamental frequency (Hz)	0.01 ± 0.03	0, 0.09	0.25
	Peak frequency (Hz)	0.01 ± 0.03	0, 0.11	0.23
	Maximum fundamental frequency (Hz)	0.02 ± 0.03	0, 0.11	0.14
	Maximum peak frequency (Hz)	0.02 ± 0.04	0, 0.14	0.13
	Bandwidth (Hz)	0.11 ± 0.1	0, 0.35	0.06
	Weiner Entropy	0.08 ± 0.09	0, 0.31	0.09

Table 4.4. Sex differences in *churr* (n = 907 calls from 54 birds) and *triple* (n = 246 calls from 23 birds) call parameters, tested using Gaussian (continuous data) or Poisson (count data) GLMMs.

Call	Parameter	Est.	df	t / z	p
Churr	Log Total call duration (s)	-0.03	48.91	-0.92	0.36
	Syllable gap/length	-0.03	50.96	-1.64	0.11
	Log Syllable 1 duration (s)	-0.01	49.59	-0.5	0.62
	Syllable 2 duration (s)	0.04	50.46	1.24	0.22
	Syllable 2/syllable 1 duration (s)	0.04	51.6	1.43	0.16
	Syllable 2 repeats	-0.05	-	-0.99	0.32
	Fundamental frequency (Hz)	-2.89	49.81	-0.04	0.97
	Peak frequency (Hz)	-3.782	49.55	-0.06	0.95
	Max. fundamental frequency (Hz)	-36.26	51.10	-0.47	0.64
	Max. peak frequency (Hz)	-31.93	51.15	-0.43	0.67
	Bandwidth (Hz)	-64.28	50.95	-1.65	0.11
	Weiner Entropy (log)	0.01	50.03	1.89	0.06
	Syllable 1 fundamental frequency (Hz)	-22.21	52.64	-0.29	0.78
	Syllable 1 peak frequency (Hz)	-27.26	52.64	-0.35	0.73
	Syllable 1 max. fundamental frequency (Hz)	-38.35	51.09	-0.49	0.62
	Syllable 1 max. peak frequency (Hz)	-30.81	51.13	-0.41	0.68
	Syllable 2 fundamental frequency (Hz)	-24.46	51.06	-0.33	0.74
	Syllable 2 peak frequency (Hz)	-33.32	50.93	-0.49	0.62
	Syllable 2 max. fundamental frequency (Hz)	-35.22	51.45	-0.45	0.66
	Syllable 2 max. peak frequency (Hz)	-55.37	51.42	-0.73	0.47
Triple	Total call duration (s)	-41.06	19.06	1.36	0.19
	Syllable gap/length	-0.01	18.56	-0.33	0.75
	Element repeats	0.01	-	0.05	0.96
	Fundamental frequency (Hz)	48.71	19.48	0.62	0.54
	Peak frequency (Hz)	49.63	19.52	0.64	0.53
	Maximum fundamental frequency (Hz)	32.61	19.8	0.3	0.77
	Maximum peak frequency (Hz)	26.82	19.77	0.26	0.8
	Bandwidth (Hz)	1.59	17.85	0.1	0.92
	Weiner Entropy	0.17	19.59	0.67	0.51

4.4 Discussion

4.4.1 *Individuality of long-tailed tit calls*

The results of the SPCC analysis showed that churr calls made by the same individual are more similar in acoustic structure than those made by different individuals, consistent with previous findings (Sharp & Hatchwell 2005). SPCC is a powerful tool for comparing the overall similarity of sounds as it incorporates all parameters of the spectrogram into the analysis, providing a comprehensive overview of acoustic structure. However, there are limitations. The correlation values are sensitive to the fast Fourier transformation (FFT) length used to generate the spectrograms (Giret *et al*. 2011) and small differences in duration and modulation rate (Meliza *et al*. 2013). Increased background noise may also decrease correlation values (Janik 1999). A number of calls used in this analysis contain some level of background noise, which may explain why in this population, SPCC correlation values of within individual comparisons have a mean of just 0.54. However, even with this seemingly low similarity value, within individual comparisons had significantly higher correlation values than between individual comparisons. SPCC did not detect a significant difference in similarity among triple calls made by the same individual compared with calls made by different individuals, even though the difference in mean correlation coefficients was greater than in the churr call, but this was probably due the small sample size.

SPCC, in conjunction with multivariate analyses such as multi-dimensional scaling or more recently, principal coordinates analysis (Cortopassi & Bradbury 2000; Berg *et al*. 2011) is a commonly used method in studies of this kind, but DTW may offer a more sophisticated solution. DTW analysis showed that both churr and triple calls made by the same individual are more similar in acoustic structure than those made by different individuals. This approach may be more suitable than SPCC for the reasons already discussed, and will be used in all subsequenct comparisons of overall call structure. The resolution of comparative analyses can also be defined in Luscinia, without the need to re-measure or aggregate call similarity data manually. That is, DTW allows the user to select whether comparisons should be made at the level of the syllable, call, individual or group, providing a fast and simple way to compare call similarity between individuals and groups of individuals, which can then be compared to a hypothesis matrix of the same dimensions (e.g. kin or non-kin), or a matrix comprising correlation coefficients, such as genetic relatedness estimates.

However, in contrast to SPCC, mean DTW score is fairly low both within and between individual birds, indicative of churr and triple calls being very similar across the population. These values were similar under various combinations of parameter weightings tested. Therefore, although long-tailed tit calls are individually distinct, differences between individuals are likely to be small and subtle. An overall measure of similarity may not be sensitive enough to capture and quantify these differences at the fine scale required to correlate with genetic relatedness or kinship. Therefore, analyses of overall call similarity were followed by an analysis of specific call parameters, to determine which sound characteristics shape similarity patterns. Although selecting a finite number of pre-determined parameters could cause potentially important sound features to be missed and may be subject to researcher bias, used in conjunction with visual inspection of sonograms and more objective analyses such as DTW, this method provides a comprehensive bioacoustic analysis with which to compare call similarity and has been implemented in numerous comparable studies (e.g. Yasukawa *et al.* 2008; Keen *et al.* 2013, Dowling *et al.* 2016).

Parameter analyses indicated that both churr calls and triple calls made by the same individual were highly repeatable over time. These findings are again consistent with earlier studies on a different population of long-tailed tits using alternative methods (Sharp & Hatchwell 2005). Together, these results confirm that long-tailed tit calls are individually distinct and highly repeatable.

4.4.2 Call characteristics for kin recognition

Parameter analyses on the larger dataset from 2015-2017 indicated that long-tailed tit churr calls were significantly repeatable based on all sound characteristics measured, and that the triple calls were also significantly repeatable based on all except one of the sound characteristics measured, supporting the results of the DTW analysis. Although this makes excluding any sound characteristics for kin recognition difficult, this is perhaps unsurprising. One explanation for this uniformly high repeatability is that some of the sound parameters are correlated and measure similar aspects of the call. For example, call duration was separated into the three temporal parameters to identify which aspects of the call influence its duration, i.e. element length, the number of elements or the pauses between elements. The two measurements of frequency, peak frequency and fundamental frequency, are also very similar, particularly in syllables without multiple harmonics (such as these). Most studies of bird vocalisations identify fundamental frequency as one of the most informative characteristics of a sound (Riede *et al*. 2006), and this quantifies what is perceived as pitch. Peak frequency calculates the frequency of highest relative amplitude, indicating where most of the acoustic energy is concentrated across the call. One way to statistically identify highly correlated parameters would be to compare individual variation in one parameter with variation in another. A strong relationship may indicate collinearity and highlight any redundant measurements. However, individual recognition may not be based on one or two sound parameters, but many subtle characteristics may instead make up an individual’s signature (Beecher 1982), perhaps necessitating high repeatability of a large number of acoustic parameters.

Nevertheless, certain parameters had particularly high repeatability. Both frequency measurements had higher repeatability than any of the temporal measurements (fundamental frequency was slightly higher than peak frequency). Maximum fundamental and peak frequency showed higher repeatability than the mean of these values across the call or syllable. Other studies have identified frequency measurements as important indicators of caller identity (Wanker & Fischer 2001; Sharp & Hatchwell 2005; Thomsen *et al*. 2013), and these characteristics may be involved in recognition. Although discrimination is likely to be based on the most reliable cues (Sherman *et al*. 1997), it is possible that the call characteristics used in kin recognition are not entirely the same as those used in individual recognition, and to discard any acoustic measurements at this stage may be premature. Yet, several parameters had fairly low repeatability and appeared to be context dependent, such as the number of element repeats, both in the triple call and the second syllable of the churr call. Both Weiner entropy and bandwidth also had fairly low repeatability. These parameters refer to the noisiness of the signal, and although background noise was not measured, these aspects of the call were the most likely to be affected by recording quality and appear not to provide any additional information of biological relevance over that which can be taken from both the duration and frequency parameters. Separating the churr call into two syllables tended to reveal higher repeatability in the first syllable than the second. Previous studies have found that vocalisations can be partitioned into different components, each serving a separate function (Payne *et al*.

1988; Leedale *et al*. 2015) and the churr call may be both context-dependent and used to signal caller identity. Overall, the results suggest that the triple call is more individually distinct than the churr call. This is again consistent with previous findings (Sharp & Hatchwell 2005) and is perhaps unsurprising, as the triple call is a long-range contact call, often used in winter foraging flocks outside the breeding season and before nests sites are chosen and pairs are selected. Long-tailed tits may use cues in the triple call to recognise kin at this stage, possibly influencing their decision of who to pair with and where to build their nest. However, the churr call also has high repeatability and has previously been shown to be involved in kin recognition (Sharp *et al*. 2005) and both calls may potentially carry important information about identity and kinship.

These bioacoustic analyses have demonstrated individuality in the churr call of long-tailed tits, completing an important initial stage in investigating vocal kin recognition in this species. This work has demonstrated that churr calls have high intra-individual repeatability, which means that individuals for whom I have only a small number of calls can be included in further analyses. These calls were repeatable based on almost all acoustic parameters tested, but certain characteristics demonstrated particularly high repeatability and may be more reliable indicators of caller identity. These individuality patterns can be compared with call patterns observed among relatives when investigating how call similarity correlates with kinship.

Chapter 5

Mechanisms of kin recognition in long-tailed tits

Summary

In most cooperative breeders, individuals are organised in to family groups. In a minority, breeding populations comprise of extended networks of opposite sex adults of varying relatedness. Selection for effective kin recognition may be expected for individuals in such groups to gain indirect fitness and avoid inbreeding. Here, I examine how call similarity correlates with kinship and affects helping and pairing decisions in the long-tailed tit *Aegithalos caudatus*, a cooperative breeder in which help is redirected. I detect a positive correlation between call similarity and kinship. Native breeders are also more acoustically similar to other natives than to breeders that dispersed into the study site as adults. As predicted, failed breeders choose to help males with more similar calls, but do not adjust their provisioning rates according to vocal similarity, suggesting call similarity alone is not responsible for fine-tuning in helping effort. Contrary to our predictions, breeders prefer partners with similar calls to themselves, although this could be confounded by micro-geographic call variation. This result also suggests that vocal similarity *per se* is not used to avoid inbreeding. The results suggest that although vocalisations make up an important part of the recognition system of long-tailed tits, this is likely to be based on prior association, rather than phenotype-matching, and may involve a combination of learned cues.

5.1 Introduction

Kin selection is often invoked to explain the evolution of cooperation in kin-structured communities, and the ability to discriminate kin from non-kin is a crucial aspect of kin selection theory (Hamilton 1964). Kin recognition refers to the differential treatment of conspecifics differing in genetic similarity (Sherman *et al.* 1997). The adaptive functions of kin recognition for cooperative species are well established: inbreeding avoidance (Pusey & Wolf 1996, Koenig & Haydock 2004) and increased indirect fitness from helping relatives (Wilson 1975; Bourke 2011). The proximate mechanisms are often less clear; any cue that reliably co-varies with relatedness may be used to discriminate kin from non-kin (Komdeur & Hatchwell 1999), but any recognition system is prone to error (Sherman *et al.* 1997). Thus, kin recognition is likely to involve a certain rate of *acceptance errors*, where non-kin are perceived as kin, and *rejection errors*, where kin are perceived as non-kin (Reeve 1989). The accuracy of kin recognition, and hence the frequency of such errors, depends on the relative costs of each kind of error; these in turn will be determined by the probability of encountering a relative and the fitness consequences of the associated behaviour (Agrawal 2001).

This theoretical framework is supported empirically by intraspecific studies showing shifts in acceptance thresholds as the costs of error change (Downs & Ratnieks 2000), and by comparative analyses that demonstrate stronger kin discrimination in cooperatively breeding vertebrates where the benefits of helping are greater (Griffin & West 2003), and when the average relatedness within a group is lower and more variable (Cornwallis *et al.* 2009). This relationship between kin recognition accuracy, probability of encountering relatives, and fitness outcomes remains to be studied in the context of mate choice, but a similar pattern may be expected if kin-selected recognition mechanisms have been co-opted for inbreeding avoidance within species. However, it is likely that kin recognition thresholds may differ in the contexts of helping and mate choice if the relative fitness costs and benefits associated with acceptance and rejection errors vary.

Most animal societies exhibit strong kin structure, whereby individuals are organised in to discrete family units of parents and their retained offspring (Ligon & Burt 2004). Because the probability of encountering a relative is high, individuals can maximise indirect fitness by indiscriminately cooperating within the group and avoid inbreeding by selecting partners from outside the group. In more complex societies, such spatial cues to kinship may be unreliable. For example, in a small number of cooperatively breeding birds, cooperation occurs after natal dispersal, across extended networks known as *kin neighbourhoods* (Dickinson & Hatchwell 2004). Here, the relatedness among spatially clustered individuals is less predictable, so kin recognition is likely to be based on phenotypic cues instead (Gamboa *et al.* 1991).

Vocalisations are the most widespread recognition cues among birds (Sharp *et al.* 2005; McDonald & Wright 2011; Açkay *et al.* 2013), although olfactory kin recognition has been described in a few species (Coffin *et al.* 2011; Krause *et al.* 2012; Bonadonna & Sanz-Aguilar 2012). Avian vocalisations are learned (Sharp *et al.* 2005) or imprinted (Bateson 1978) during the nestling period, a sensitive stage in development when all individuals present are likely to be kin (Komdeur & Hatchwell 1999). Once recognition cues are fixed, individuals are potentially able to recognise familiar kin outside of the association context. Phenotype matching is an alternative mechanism of kin recognition that involves an assessment of relatedness, rather than associative learning. Here, an individual's own phenotypic cues, or those of a subset of familiar kin, are generalised to form an internal template (Lacey & Sherman 1983) against which the phenotypes of other individuals are compared (Greenberg 1979; Holmes & Sherman 1983). Phenotype matching does not require a period of previous association (Tang-Martinez 2001). Instead, it relies on a positive correlation between template-phenotype similarity and degree of genetic relatedness (Mateo 2004).

Whether kin are recognised through prior association or phenotype matching can be difficult to determine. Both require active recognition based on phenotypic cues that co-vary with relatedness, and both mechanisms result in discrimination patterns that reflect phenotypic similarity between individuals. Functionally, both recognition mechanisms provide a similar outcome, and many studies have suggested that both associative learning and phenotype matching could work together to mediate kin recognition within a single species or population (Greenberg 1979; Holmes & Sherman 1983; Komdeur & Hatchwell 1999). But, crucially, mechanisms involving phenotype matching would permit individuals to recognise unfamiliar kin, and distinguish between kin of varying relatedness.

In general, kin recognition based on familiarity may be sufficient for individuals to maximise inclusive fitness by directing help towards relatives, and associative learning is indeed the most commonly identified mechanism of kin recognition in cooperatively breeding birds (Curry 1988; Hatchwell *et al.* 2001b; Komdeur 2004). However, studies on long-tailed tits *Aegithalos caudatus* (Nam *et al.* 2010) and bell miners *Manorina melanophrys* (Wright *et al.* 2010), suggest that helpers may be able to modify provisioning effort according to their degree of relatedness to recipient broods. In both species, help is extended beyond discrete, nuclear family groups, so the risk of caring for non-kin is high. Thus, kin recognition mechanisms with low error rates are likely to be under strong selection. Moreover, finely tuned adjustment of provisioning behaviour in relation to kinship could indicate a relatively sophisticated mechanism of kin recognition, which may involve phenotype matching. McDonald & Wright (2011) subsequently identified a relationship between genetic relatedness and vocal similarity in bell miners, but whether this relationship exists in other species remains to be tested.

As with helping behaviour, relatively simple decision rules can alleviate inbreeding in viscous populations with complete, sex-biased dispersal (Koenig & Haydock 1994). However, incest avoidance mechanisms that rely solely on associative learning may be prone to acceptance errors if opposite-sex kin are not sufficiently segregated. Yet, inbreeding is extremely rare in natural populations (Pusey & Wolf 1996), leading some researchers to propose phenotype matching as a likely avoidance mechanism (Bateson 1983; Bonadonna & Sanz-Aguilar 2012). For example, a recent study on the decorated cricket *Gryllodes sigillatus* found that manipulating a female's own phenotype affects their choice of partner (Capodeanu-Nägler *et al*. 2014). Among social birds, kin avoidance during mate choice has been identified in western bluebirds *Sialia mexicana* (Dickinson *et al.* 2016), but the mechanisms remain undetermined and there is no conclusive evidence for phenotype matching.

This study aimed to identify a recognition mechanism that permits active inbreeding avoidance and flexible helper investment in long-tailed tits, a kin-neighbourhood cooperative breeder shown to exhibit effective kin recognition in the absence of spatial cues (Russell & Hatchwell 2001; Leedale *et al.* 2018). Helpers are failed breeders that redirect their care following unsuccessful attempts at independent breeding. Around 40% of nests receive help (Hatchwell *et al.* 2004), typically from one or two helpers, but not all failed breeders choose to become helpers (Hatchwell *et al.* 2013). Although kin remain clustered during breeding, most neighbours are non-kin and help is directed towards close kin more often than expected by indiscriminate helping (Leedale *et al.* 2018). Furthermore, helpers provision more closely related broods at higher rates (Nam *et al.* 2010). Helpers are overwhelmingly male, and gain indirect fitness benefits by increasing the productivity of broods (MacColl & Hatchwell 2004; Hatchwell *et al.* 2014). In contrast, no direct fitness benefits of helping have been identified (McGowan *et al.* 2003; Meade & Hatchwell 2010). Long-tailed tits also demonstrate effective avoidance of kin as partners, despite the inbreeding risk generated by fine-scale population genetic structure (Chapter 3). Long-tailed tits have two individually distinct contact calls: the *churr* and the *triple* (Sharp & Hatchwell 2005). Both are learned in the nest, and playback experiments have shown that birds can discriminate between the churr calls of kin and non-kin (Sharp *et al.* 2005). However, if calls provide cues to kinship, the fine-tuning of helper care in relation to kinship (Nam *et al.* 2010) suggests a positive correlation between vocal similarity and relatedness. Furthermore, whether these vocalisations are actually used to assess relatedness when making helping and pairing decisions remains untested.

I investigated the use of churr and triple calls for kin recognition in two ways. First, the association between vocal similarity and relatedness was assessed by comparing churr and triple call similarity among breeders of varying relatedness. As these calls are learned in the nest, a positive correlation between call similarity and relatedness was predicted. Vocal similarity was correlated against two measures of relatedness: genetic relatedness estimates, and kinship coefficients from a social pedigree. Although genetic relatedness closely matches kinship in this species, there may be some discrepancies due to the variation in genetic similarity among kin and non-kin (Chapter 3, Fig. 3.1). Moreover, as social pedigrees are incomplete in open populations, some kinships may go undetected. There is also a small proportion (<8%) of extra-pair offspring (Hatchwell *et al.* 2002). If differences occur, I expect vocal similarity to correlate more strongly with kinship than genetic relatedness, because these vocalisations are learned during development (Sharp *et al.* 2005). Specific vocal characteristics that correlate particularly strongly with relatedness were also identified. I also compared the similarity of churr calls of *native* breeders (birds hatched in the study site) and *immigrant* breeders (birds that dispersed into the study site as adults). Vocalisations were expected to be more similar among natives than between natives and immigrants, due to natal philopatry. Second, I examined whether the degree of similarity in churr and triple calls influenced helping and pairing decisions by analysing: (i) the vocal similarity of helpers to the breeders they helped and the nearby breeders they did not help, and (ii) the vocal similarity within breeding pairs and among potential, unpaired partners. Finally, I investigated whether long-tailed tits helpers adjust their provisioning effort according to how similar their churr calls are to the helped breeders.

5.2 Methods

5.2.1 Study site and population monitoring

Fieldwork was carried out on a population of 31-46 breeding pairs of long-tailed tits in a section of the Rivelin Valley, Sheffield, UK (53º38’N 1º56W) from 2015-2017. The site is approximately 2.5km^2 and comprises a variety of suitable breeding habitat. This population of 17-72 (mean c.50) pairs has been studied extensively during the breeding season (February-June) since 1994. Almost all individuals (>95%) were marked with a BTO ring and a unique combination of two colour rings for field identification. Native birds were ringed as 11-day old nestlings and immigrants were captured in mist nests under BTO licence before or during breeding. When ringed, individuals were weighed, and wing and tarsus length measured. A sample of 5-30µl of blood was taken by brachial venepuncture under Home Office licence. All nesting attempts were closely monitored to record breeding events and life-history traits such as incubation period and clutch size, and the identity of parents and helpers. Nest locations were recorded using GPS receivers to an accuracy of 8m. For most visible nests, provisioning behaviour was observed every two days from day two to fledging or until nest failure. Most observation periods lasted 1h, during which the identities and visit rate of all carers were recorded. For further details of provisioning observations, see MacColl & Hatchwell (2003) and Nam *et al.* (2010).

5.2.3 Genetic and social relatedness

Individuals were genotyped at 17 microsatellite loci (Appendix 3). Genetic relatedness was estimated using Queller and Goodnight's (1989) r_{QG} coefficient of relatedness in *SPAGeDi version 1.1.5* (Hardy & Vekemans 2002). This relatedness estimate is reliable when tested against our social pedigree (Nam *et al*. 2010). For further details on genotyping, see Simeoni *et al*. (2007) and Adams *et al*. (2015). The population allele frequencies used in analyses were generated using all genotyped individuals (1994-2017, n = 3304) in *CERVUS version 3.0.7* (Kalinowski *et al*. 2007) to ensure non-zero frequencies for all alleles. To calculate social relatedness among dyads, an additive relationship matrix was generated from the social pedigree (1994-2017, n = 3068) in *R version 3.5.0* (R Core Team 2018), using the *R* package, *nadiv* (Wolak 2012). For further details on pedigree construction see Appendix 1. Hereafter, *genetic* *relatedness* refers to the r_{QG} coefficients calculated from the microsatellite markers, whereas *kinship* refers to social relationships derived from the pedigree.

5.2.4 Field recordings

Adult vocalisations of two contact calls, the *churr* and the *triple*, were recorded using a *Sennheiser ME67/K6* shotgun microphone fitted with a *Rycote* windjammer. Recordings were made onto a *Roland R-05 version 1.03* WAV/MP3 recorder with a 6GB *SanDisk* memory card, set to a sample rate of 48kHz with WAV-16bit accuracy. The microphone input level was set to 60db, to optimise sensitivity, with a low-cut frequency of 400Hz, to cut out extreme low frequency noise. All recordings were made in approximately similar conditions between 06:00 and 18:00 BST. Birds were recorded at a distance of approximately 3-15m, to minimise sound degradation and reverberation. Birds were recorded at the nest and identified by their unique colour ring combinations. If more than one bird was present, vocalisations were assigned to individuals by observing movements of the bill and throat feathers. At the start of each recording, date, time, nest number and recording number were dictated into the microphone. When caller ID could be identified with certainty, this was dictated into the microphone after each call. In total, 213 recordings were made, containing 1116 churr calls from 98 birds (mean ± SD = 11.39 ± 10.24 per bird; range 1 – 42) and 400 triple calls from 55 birds (mean ± SD = 7.02 ± 5.99 per bird; range 1 – 23).

5.2.5 Bioacoustic analysis

The sampling frequency was converted to 22.05 KHz and recordings were visualised spectrographically to assess call quality, with a frequency resolution of 188Hz and a time resolution of 2.7ms in *Avisoft SAS-Lab Pro version 4.52* (Avisoft Bioacoustics). Recordings with extreme background noise were excluded. All useable calls were isolated, stored and measured in *Luscinia version 2.16.10.29.01* (Robert Lachlan, Queen Mary University of London, UK). Vocal similarity was assessed by: (i) dynamic time-warping analysis (DTW) implemented in Luscinia (Lachlan 2007), and (ii) the difference in a range of repeatable acoustic parameters (Table 5.1), measured as Euclidean distances using the *R* package, *spaa* (Zhang 2016). Parameters were selected that have individual repeatability estimates >0.2 (Chapter 4). DTW analysis generates a score representing the amount of warping required to match one signal to another. Therefore, the lower the DTW score, the greater the vocal similarity between individuals. The acoustic features used in the DTW analysis were weighted as: time = 1, fundamental frequency = 2, change in fundamental frequency = 1, change in peak frequency = 1. These settings generated a DTW algorithm which correctly matched visually similar vocalisations. For further details on bioacoustics, see Chapter 4.

5.2.6 Call similarity and relatedness

To investigate how vocal similarity varied with relatedness, I tested for a relationship between vocal similarity and: (i) genetic relatedness, (ii) kinship. For the latter, three degrees of kinship were considered: first-order (r = 0.5), second-order (r = 0.25), or non-kin (r < 0.25); non-kin relationships included only those birds for which the parentage of both birds in the dyad was known. This analysis was carried out among all breeders, and separately among males and females. To determine how vocal similarity varied in relation to the dispersal status of breeders in our long-tailed tit population, I compared the vocal similarity (based on churr DTW score) of three categories of dyads: (i) among native breeders (including all parent-offspring relationships), (ii) among immigrant breeders, and (iii) between native and immigrant breeders. Kinship among immigrants was based on sibship reconstructions using genetic relatedness estimates (Appendix 1).

Table 5.1. List and definitions of the selected acoustic parameters used to assess vocal similarity in two contact calls of breeding long-tailed tits; the *churr* and the *triple*.

Call	**Acoustic Parameter**	**Definition**
Churr	Bandwidth (Hz)	Difference between the first and final frequency of maximum intensity.
	DTW score	Similarity in overall acoustic structure based on dynamic-time warping analysis (Lachlan 2007).
	Duration (ms)	Time between start and end of entire call.
	Frequency (Hz)	Mean fundamental frequency. This is the frequency of pure-tone signals and common denominator frequency of harmonic signals.
	Max. frequency (Hz)	Maximum fundamental frequency across call.
	S1 duration (s)	Time between start and end of syllable 1.
	S1 frequency (Hz)	Fundamental frequency of syllable 1.
	S1 max. frequency (Hz)	Maximum fundamental frequency across syllable 1.
	S2 duration (s)	Time between start and end of syllable 2.
	S2 frequency (Hz)	Fundamental frequency of syllable 2.
	S2 max. frequency (Hz)	Maximum fundamental frequency across syllable 2.
	S2/S1 duration (ms)	Ratio of syllable 2 length to syllable 1 length.
Triple	Bandwidth (Hz)	Difference between the first and final frequency of maximum intensity.
	DTW score	Similarity in overall acoustic structure based on dynamic-time warping analysis. (Lachlan 2007)
	Duration (ms)	Time between start and end of entire call.
	Frequency (Hz)	Mean fundamental frequency. This is the frequency of pure-tone signals and common denominator frequency of harmonic signals.
	Max. frequency (Hz)	Maximum fundamental frequency across call.

5.2.7 Call similarity and helping decisions

Of the nests for which I had churr call recordings of the helper and the breeding male (n = 19), 26% of helpers were known first order relatives of the male, 16% were second order relatives of the male, and 55% were apparently unrelated to the male. In the remaining 5% of nests, kinship could not be determined from the social pedigree. However, genetic relatedness of helpers to male breeders was $r = 0.17 \pm 0.2$ (mean ± SD, n = 17), suggesting that kinship may be underestimated in our pedigree, as is often the case in open populations. Among nests for which I had churr call recordings of the helper and the breeding female (n = 15), there were no cases of help given to known female kin, and the genetic relatedness of helpers to females was $r = -0.04 \pm 0.12$ (mean ± SD, n = 13). The number of helped nests for which I had triple call recordings of the helper and the breeding male was smaller (n = 5). In two of these cases, the helper was a first or second order relative of the male. Again, there were no cases of help given to known female kin (n = 7). Therefore, analyses focused on the helper's vocal similarity to the breeding male.

If individuals use vocal similarity as a cue to relatedness, in order to direct helping effort towards close kin, I expected helpers to be more vocally similar to the breeders they helped than the breeders they did not help. For each helper, vocal similarity to male breeders at their first *chosen* nest in a given year was compared with their mean vocal similarity to a sample of *potential* males (excluding those helped) nesting within 750m that year, the range in which the majority of failed breeders travel to provide aid (mean ± SD = 337.4m ± 253.4, 95% CI = 744.1m, n = 220). Helping distance was calculated as the distance between a helper's last failed breeding attempt and the nest at which they first appeared as a helper in the same year.

To investigate whether helpers use vocal similarity to modify their provisioning effort, I tested for a relationship between the provisioning rates of helpers and their vocal similarity to the helped males. Because I predicted that vocal similarity is used as a cue to relatedness, I also tested for a relationship between provisioning rate and relatedness, using genetic relatedness estimates and kinship from the social pedigree. Although the fitness consequences of helping depend on the genetic relatedness, pedigree data is essential for understanding how accurately individuals are able to recognise kin, particularly when the mechanism depends on socially learned cues (Sharp *et al.* 2005). Provisioning rate was therefore expected to correlate most strongly with kinship.

5.2.8 Call similarity and pairing decisions

Among the breeding pairs for which I had churr (n = 51) or triple (n = 14) call recordings of both breeders, there were no cases of pairing among first order or second order kin. Mean genetic relatedness among recorded breeding pairs was $r = 0.06 \pm 0.15$ (n = 49). If individuals use vocal similarity to assess relatedness during mate choice to avoid inbreeding, vocal similarity within breeding pairs should be lower than that expected under random mate choice. For each focal individual, vocal similarity to the *chosen* partner was compared with their mean vocal similarity to a sample of *potential* partners (excluding chosen partner) within 900m that year, the range in which the majority of pairs are formed (Chapter 3). Potential partners were first-year, widowed or divorced breeders of the opposite-sex present in the breeding population in the same year as the focal individual. The distance between adults was based on the location of an individual’s first breeding attempt in a given year. Analyses were carried out from the perspective of male and female breeders separately.

5.2.9 Statistical analysis

All statistical analyses were carried out in *R version 3.5.0* (R Core Team 2018). The relationships between vocal similarity and relatedness among adult breeders were analysed using Mantel tests (Schnell *et al.* 1985) based on Spearman correlations of ranked distances with 10000 permutations using the R package, *ecodist* (Goslee & Urban 2007). In the context of both helping and pairing decisions, observed vocal similarity within chosen dyads was compared to mean vocal similarity within potential dyads for each focal individual using Wilcoxon signed-rank tests. All reported significance values are based on two-tailed tests and the sample sizes for each test are reported with the results.

I investigated whether helper provisioning rates varied with respect to their vocal similarity to male breeders using linear mixed effects models fitted by restricted maximum likelihood (REML) in the *R* package, *lme4* (Bates *et al.* 2008). As genetic relatedness, kinship and vocal similarity are expected to be closely correlated, their effect on provisioning rate was analysed in separate models. In each model, the *provisioning rate of helpers* (number of visits/hour) was the response variable. In the first model, the explanatory variables were: *nestling age, brood size*, *group size* and *genetic relatedness,* all of which influence the provisioning rates of helpers (Nam *et al.* 2010). In the second model, the explanatory variables were: *nestling age*, *brood size*, *group size* and *kinship.* In the final model, the explanatory variables were: *nestling age*, *brood size*, *group size* and *vocal similarity*. *Genetic relatedness* was the r_{QG} estimate between helpers and male breeders, measured as a continuous variable. *Kinship* was the relationship between helpers and male breeders according to the social pedigree (three factor levels: $r = 0$, $r = 0.25$ and $r = 0.5$). *Vocal similarity* was the DTW score of churr calls between helpers and male breeders. *Nestling age* was measured in days from hatching (day 0; long-tailed tit broods hatch synchronously). *Brood size* was the number of chicks in the nest on day 11, a good indicator of brood size from hatching because nestling starvation is rare (Hatchwell *et al.* 2004). *Group size* was the number of adults at the nest (parents and helpers). *Bird identity* and *nest identity* were included as random effects, to control for non-independence of repeated observations of feeding rates by the same birds, and repeated observations of feeding rates at the same nest. All explanatory covariates were initially included in full models and then dropped sequentially unless doing so significantly reduced the amount of variance explained, generating three minimum adequate models (MAMs) from full models containing either genetic relatedness, social relatedness or vocal similarity as explanatory variables.

5.3 Results

5.3.1 Call similarity and relatedness

There was substantial variation in vocal similarity among breeders in all pedigree categories. This was evident among males, among females and among all birds for churr (Fig. 5.1) and triple calls (Fig. 5.2). As predicted, kinship correlated with churr similarity in several acoustic parameters: DTW score (Mantel test: $R = -0.07$, $n = 80$, $p < 0.01$), difference in fundamental frequency (Mantel test: $R = -0.04$, $n = 80$, $p < 0.05$) and difference in frequency bandwidth (Mantel test: $R = -0.04$, $n = 80$, $p < 0.05$; Fig. 5.3a). Relationships between call similarity and kinship were also evident when tested separately among males and females, although the parameters that correlated with kinship differed between the sexes. Among males, kinship was correlated with churr similarity based on the difference in call duration (Mantel test: $R = -0.06$, $n =$

46, $p < 0.05$), duration of the 1st syllable (Mantel test: $R = -0.07$, $n = 46$, $p < 0.05$) and frequency bandwidth (Mantel test: $R = -0.09$, $n = 46$, $p < 0.01$; Fig. 5.3a). Among females, kinship correlated with churr similarity based on DTW score (Mantel test: $R = -0.1$, $n = 34$, $p < 0.05$) and the difference in frequency (Mantel test: R -0.11, $n = 34$, $p < 0.05$; Fig. 5.3a). In contrast, churr call similarity did not correlate with genetic relatedness, regardless of breeder sex (Fig. 5.3b). Moreover, triple call similarity did not correlate with kinship (Fig. 5.4a) or genetic relatedness in either sex (Fig. 5.4b).

The similarity of churr calls also varied with patterns of philopatry. Churr calls were most similar among native breeders (mean pairwise DTW score ± SD = 0.24 ± 0.08, $n = 300$ dyads), slightly less similar among immigrant breeders (0.25 ± 0.08, $n = 1377$), and least similar between native and immigrant breeders (0.29 ± 0.11, $n = 3076$ dyads; Mantel test: $R = -0.18$, $n = 98$, $p < 0.001$; Figure 5.5).

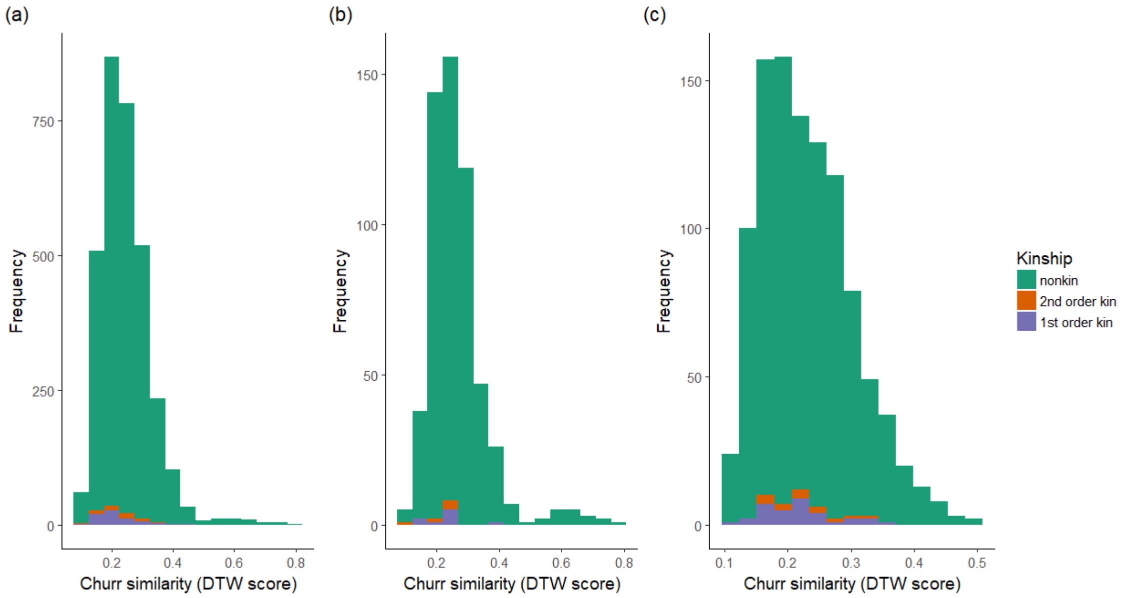

Figure 5.1. Distributions of similarity in long-tailed tit *churr* calls among (a) both sexes (n = 80), (b) males (n = 46), and (c) females (n = 34) of three levels of kinship: first-order kin (purple bars, n = 71 dyads), second-order kin (orange bars, n = 32 dyads) and non-kin (green bars, n = 3057 dyads), measured using dynamic-time warping analysis (DTW).

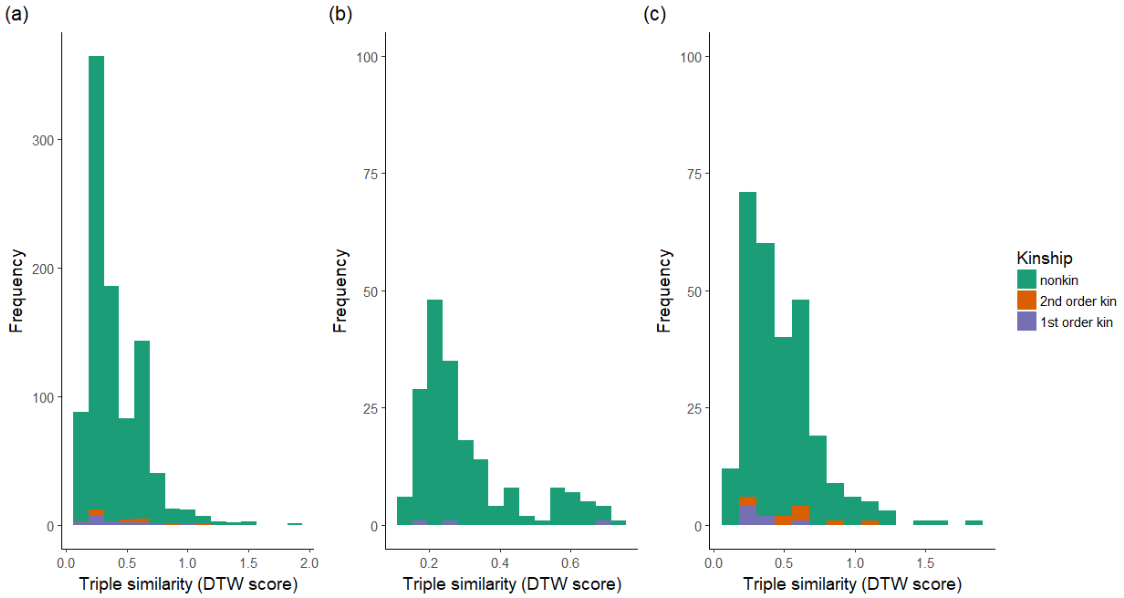

Figure 5.2. Distributions of similarity in long-tailed tit *triple* calls among (a) both sexes (n = 44), (b) males (n = 24), and (c) females (n = 20) of three levels of kinship: first-order kin (purple bars, n = 19 dyads), second-order kin (orange bars, n = 12 dyads) and non-kin (green bars, n = 915 dyads), measured using dynamic-time warping analysis (DTW).

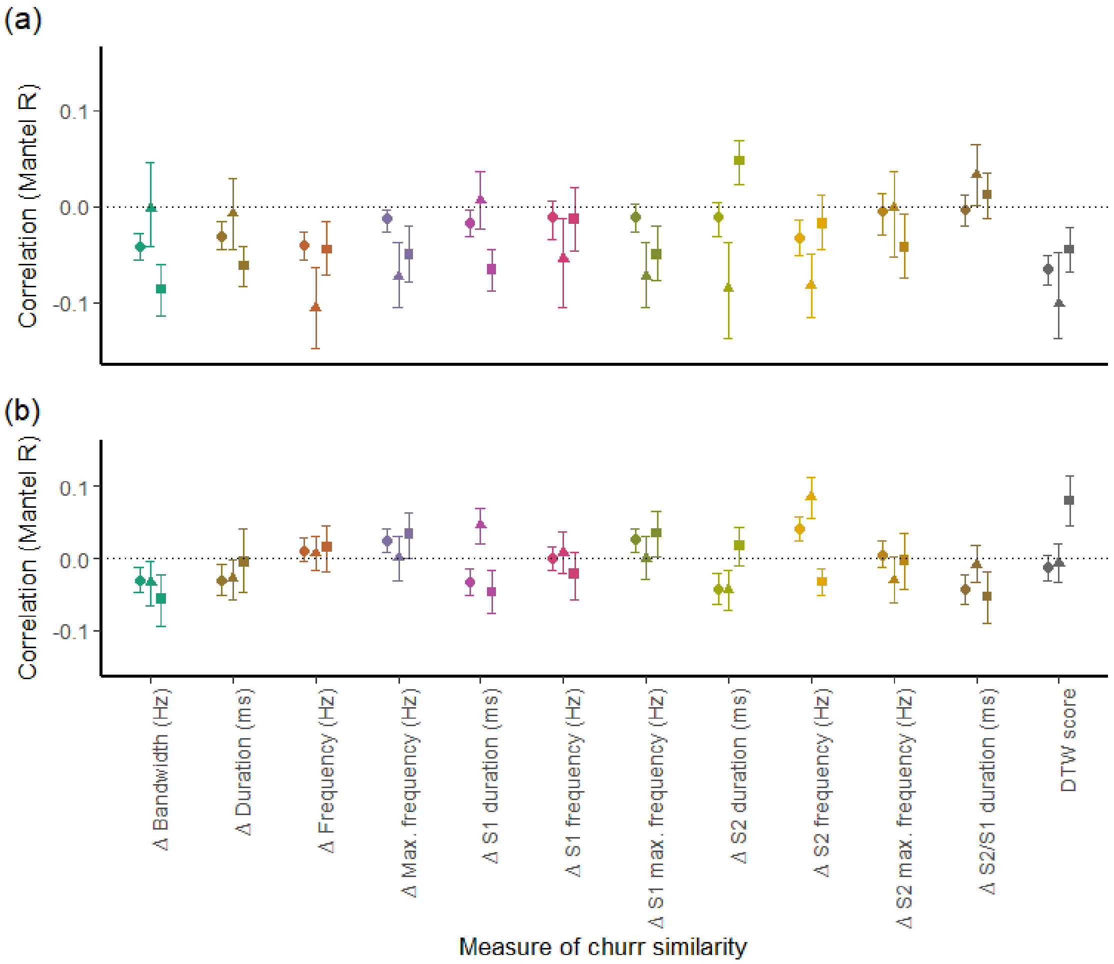

Figure 5.3. The relationship between *churr* call similarity and relatedness in long-tailed tits based on dynamic time-warping analysis (DTW score) and the difference (Δ) in a range of acoustic characteristics. Mantel R correlations are shown for dyadic comparisons among breeders of both sex (circles), among males (squares) and among females (triangles), based on (a) degree of kinship calculated from the social pedigree (all breeders: n = 80, males: n = 46, females: n = 34), and (b) genetic relatedness estimates (all birds: n = 88, males: n = 45, females: n = 43).

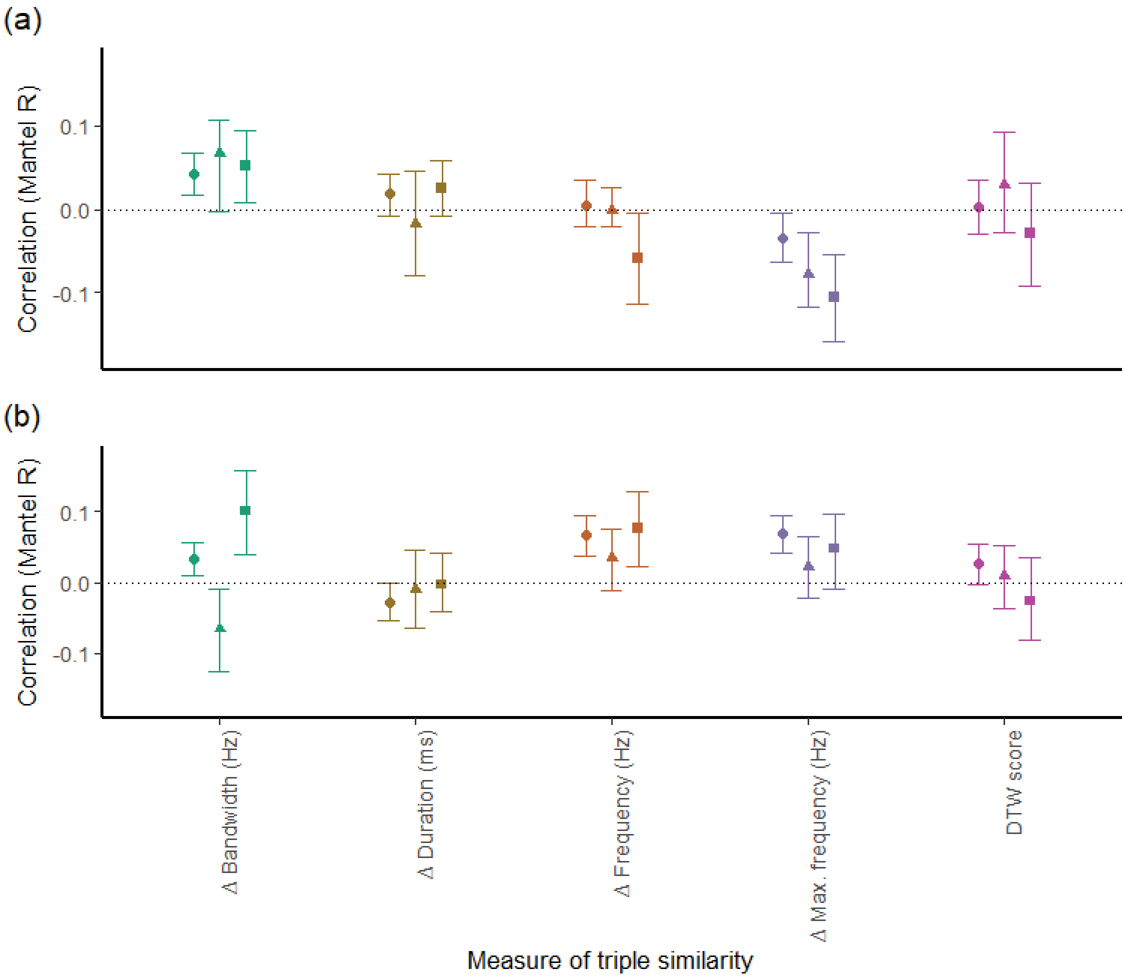

Figure 5.4. The relationship between *triple* call similarity and relatedness in long-tailed tits based on dynamic time-warping analysis (DTW score) and the difference (Δ) in a range of acoustic characteristics. Mantel R values are shown for dyadic comparisons among breeders of both sex (circles), among males (squares) and among females (triangles), based on (a) degree of kinship calculated from the social pedigree (all birds: n = 44, males: n = 24, females: n = 20), and (b) genetic relatedness estimates (all birds: n = 50, males: n = 26, females: n = 24).

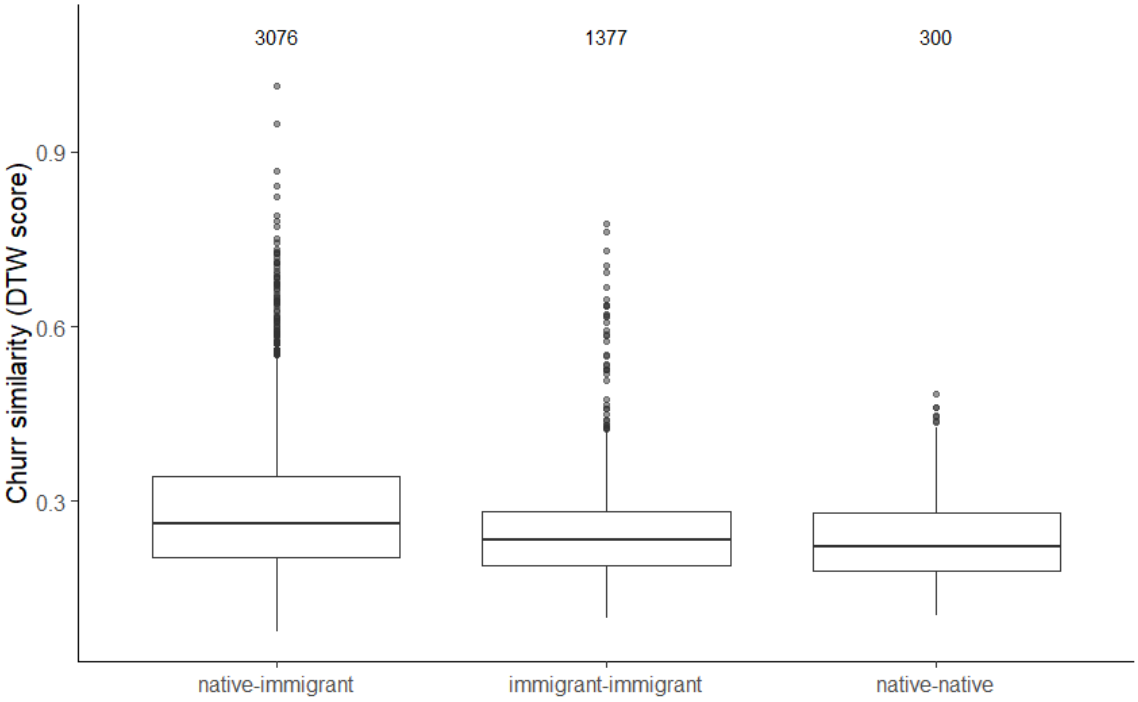

Figure 5.5. Vocal similarity of long-tailed tit breeders in relation to patterns of philopatry: median (± IQR) *churr* call similarity (based on DTW score) among native breeders (n = 300 dyads); among immigrant breeders (n = 1377 dyads) and between native and immigrant breeders (n = 3076).

5.3.2 Call similarity and helping decisions

The genetic relatedness of helpers to breeding males within 750m that were not helped was $r = 0.07 \pm 0.18$ (mean ± SD, n = 230; 10.6% of relationships in which kinship was known (n = 226) were first order kinships, 4.4% were second order kinships, 84.9% dyads were unrelated). As predicted, failed breeders chose to help males with more similar churr calls, based on DTW score (Wilcoxon signed-rank test: v = 15, n = 19, $p < 0.001$) and the difference in maximum frequency (Wilcoxon signed-rank test: v = 44, n = 19, $p < 0.05$; Table 5.2). However, there was no difference in helpers' triple call similarity to the male breeders they chose to help and those they chose not to help (Table 5.3), although the sample size for this comparison was small.

Provisioning data were available for 14/19 cases of helping, with 41 observation periods of 14 helpers at 11 nests over 3 years (mean duration of observation ± SD = 180.06 min ± 145.3 per nest, range = 1–8h, mean feeding rate (visits per hour) ± SD = 5.05 ± 2.56 per helper, range = 1–10.36). When *genetic relatedness* was included as a response variable in the full model, the minimum adequate model (MAM) for helper provisioning rate included *genetic relatedness*, *group size* and *nestling age* as explanatory variables (Table 5.4). Similarly, when *kinship* was included as a response variable in the full model instead, the MAM included *kinship* and *nestling age* as explanatory variables (Table 5.4). In contrast, when *vocal similarity* was included as a proxy for both relatedness estimates in the full model, only *nestling age* was retained an explanatory variable in the MAM (Table 5.4).

Table 5.2. Similarity of *churr* calls between helpers and helped male breeders compared with the mean churr similarity of those helpers to the potential males they did not help. Potential males were those breeding within 750m of the helpers last failed nest in the same year. Churr similarity was measured using dynamic time warping analysis (DTW), and the difference in (Δ) multiple acoustic parameters.

Similarity measure	Helped males (n = 19 dyads) mean ± SD	Potential males (n = 252 dyads) mean ± SD	Wilcoxon's signed rank v	p
DTW score	0.17 ± 0.05	0.22 ± 0.05	15	**< 0.001**
Δ Duration (ms)	25.9 ± 16.66	30.18 ± 11.57	66	0.26
Δ Mean frequency (Hz)	267.98 ± 211.95	313.8 ± 122.26	63	0.21
Δ Maximum frequency (Hz)	243.64 ± 179.51	340.68 ± 118.56	44	**< 0.05**
Δ Bandwidth (Hz)	150.06 ± 138.18	200.94 ± 101.64	57	0.13

Table 5.3. Similarity of *triple* calls between helpers and helped male breeders compared with the mean triple similarity of those helpers to the potential males they did not help. Potential males were those breeding within 750m of the helpers last failed nest in the same year. Triple similarity was measured using dynamic time warping analysis (DTW), and the difference in (Δ) multiple acoustic parameters.

Similarity measure	Helped males (n = 5 dyads) mean ± SD	Potential males (n = dyads) mean ± SD	Wilcoxon's signed rank v	p
DTW score	0.52 ± 0.37	0.32 ± 0.1	10	0.63
Δ Duration (ms)	84.47 ± 52.77	73.93 ± 47.19	10	0.63
Δ Mean frequency (Hz)	309.51 ± 262.75	184.74 ± 100.57	10	0.63
Δ Maximum frequency (Hz)	268.67 ± 262.75	206.84 ± 262.75	10	0.63
Δ Bandwidth (Hz)	58.19 ± 21.25	53.39 ± 32.86	8	1

Table 5.4. The relationship between: (i) vocal similarity, (ii) genetic relatedness and (iii) kinship between helpers and breeding males, and the provisioning rate of helpers (n = 41 observation periods, 14 helpers, 11 nests). Test statistics and significance terms are reported after backward step-wise removal of predictor variables. Degrees of freedom from the minimum adequate model are reported. All biologically meaningful two-way interaction terms were also tested, and none were significant (p > 0.05).

Model	Predictor variable	Sum sq	df	F	p
Vocal similarity	*Vocal similarity*	0.11	1, 39	0.02	0.90
	Brood size	6.72	1, 39	1.04	0.33
	Group size	22.79	1, 39	3.52	0.07
	Nestling age	30.17	1, 39	4.85	**< 0.05**
Genetic relatedness	*Genetic relatedness*	43.53	3, 37	5.83	**< 0.05**
	Brood size	6.89	3, 37	0.92	0.34
	Group size	65.87	3, 37	8.82	**< 0.01**
	Nestling age	67.09	3, 37	8.99	**< 0.01**
Kinship	*Kinship*	93.6	3, 37	6.43	**< 0.01**
	Brood size	0.51	3, 37	0.07	0.79
	Group size	5.87	3, 37	0.83	0.36
	Nestling age	60.62	3, 37	8.33	**< 0.01**

5.3.3 Call similarity and pairing decisions

The genetic relatedness of recorded, unpaired dyads within pairing range was *r* = 0.007 ± 0.14 (mean ± SD, n = 1049; 1.7% of these dyads in which kinship was known (n = 702) were first order kinships, 1.0% were second order kinships, and 97.3% dyads were unrelated). Contrary to predictions, the similarity of females' churr calls to their chosen partner was greater than the mean similarity to their potential partners, based on the difference in frequency (Wilcoxon signed-rank test: v = 410, n = 50, p < 0.05), maximum frequency (Wilcoxon signed-rank test: v = 312, n = 50, p < 0.001) and duration (Wilcoxon signed-rank test: v = 409, n = 50, p < 0.05, Table 5.5). From the male perspective, churr call similarity to chosen partner was also higher than the mean similarity to potential partners, although this was the case for maximum frequency only (Wilcoxon signed-rank test: v = 445, n = 51, p < 0.05, Table 5.5). In contrast, there was no difference in triple call similarity between selected and potential partners from either the female or male perspective (Table 5.6).

Table 5.5. Similarity in *churr* calls of focal breeders to their chosen partner compared with their mean similarity to potential partners. Potential partners are opposite-sex birds breeding within 900m of the focal breeders first nesting attempt, and is the range in which pairs are usually formed. Call similarity was measured using dynamic time warping analysis (DTW), and the difference in (Δ) set acoustic parameters.

Sex of focal breeder		Chosen partner (n = 51 pairs)	Potential partner (n = 690)	Wilcoxon's signed rank	
	Similarity measure	mean ± SD	mean ± SD	v	p
Female	DTW score	0.241 ± 0.12	0.26 ± 0.09	440	0.06
	Δ Duration (ms)	32.1 ± 32	36.17 ± 15.91	409	< **0.05**
	Δ Mean frequency (Hz)	316.65 ± 309.52	366.46 ± 184	410	< **0.05**
	Δ Maximum frequency (Hz)	273.3 ± 269.48	376.52 ± 151.76	312	< **0.01**
	Δ Bandwidth (Hz)	218.13 ± 207.58	224.36 ± 148.73	557	0.44
Male	DTW score	0.244 ± 0.12	0.25 ± 0.05	456	0.05
	Δ Duration (ms)	33.41 ± 33.02	39.67 ± 21.19	457	0.05
	Δ Mean frequency (Hz)	326.25 ± 313.98	355.75 ± 183.37	467	0.07
	Δ Maximum frequency (Hz)	282.57 ± 274.86	349.1 ± 111.96	445	< **0.05**
	Δ Bandwidth (Hz)	226.3 ± 213.62	215.32 ± 122.77	607	0.6

Table 5.6. Similarity in *triple* calls of focal breeders to their chosen partner compared with their mean similarity to potential partners. Call similarity was measured using dynamic time warping analysis (DTW), and the difference in (Δ) set acoustic parameters.

Sex of focal breeder		Chosen partner (n = 14 pairs)	Potential partner (n = 229)	Wilcoxon's signed rank	
	Similarity measure	mean ± SD	mean ± SD	v	p
Female	DTW score	0.39 ± 0.25	0.37 ± 0.08	43	0.583
	Δ Duration (ms)	83.75 ± 43.68	116.08 ± 70.69	28	0.135
	Δ Mean frequency (Hz)	165.54 ± 148.59	186.65 ± 85.11	34	0.267
	Δ Maximum frequency (Hz)	201.27 ± 150.5	253.01 ± 108.2	27	0.118
	Δ Bandwidth (Hz)	41.73 ± 25.33	56.25 ± 19.86	30	0.172
Male	DTW score	0.39 ± 0.25	0.42 ± 0.24	27	0.118
	Δ Duration (ms)	83.75 ± 43.68	94.78 ± 35.94	35	0.295
	Δ Mean frequency (Hz)	165.54 ± 148.59	195.63 ± 143.71	36	0.325
	Δ Maximum frequency (Hz)	201.27 ± 150.5	227.9 ± 138.97	41	0.501
	Δ Bandwidth (Hz)	41.73 ± 25.33	59.72 ± 18.58	21	0.5

5.4 Discussion

5.4.1 Call similarity and relatedness

Of the two long-tailed tit calls studied, the churr and the triple, a positive association between churr call similarity and kinship was identified. This is in line with Sharp & Hatchwell's (2006) finding that the vocalisations of siblings are more similar than those of non-siblings in this species. This study shows that churr call similarity correlated with relatedness when including an intermediate level of kinship. However, no significant relationship between kinship and similarity in triple calls was detected, although the proportion of kin in this sample was small (Fig. 5.2). Interestingly, churr call similarity correlated with degree of kinship according to the social pedigree, but not genetic relatedness estimates from microsatellite markers. Genetic relatedness estimates have been shown to be reliable when tested against our social pedigree (Nam *et al.* 2010), and extra-pair paternity and brood parasitism are rare in this species (Hatchwell *et al.* 2002). However, genetic relatedness estimates among first-order, second-order and non-kin vary and there is a lot of overlap (Chapter 3, Fig 3.1). As these calls develop in the nest (Sharp *et al.* 2005), it is unsurprising that vocal similarity correlates more strongly with degree of kinship than estimated genetic relatedness. Although a genetic influence cannot be ruled out completely, this supports earlier suggestions that family-specific churr calls in long-tailed tits are unlikely to have a strong heritable component (Sharp *et al.* 2005; Sharp & Hatchwell 2006).

Although call similarity among second-order kin is an intermediate level between first order kin and non-kin (mean pairwise DTW score ± SD = 0.217 ± 0.07), it is very close to the value for first-order kin (0.216 ± 0.06). Therefore, while there is a correlation with kinship, whether this is a threshold or continuous relationship remains unclear. Nestlings may learn calls from any adult present at the nest during development. Half-siblings may sound similar because they share a parent, from whom they learn their calls. But nestlings could also learn their calls from helpers, who are often second-order relatives. This may explain why the calls of first and second-order kin are relatively similar, compared with the calls of non-kin.

As well as DTW score, churr call similarity correlated with kinship based on mean fundamental frequency and frequency bandwidth. However, frequency alone didn't correlate more strongly with kinship than DTW score, and although frequency parameters are the most repeatable characteristics (Chapter 4), their relatively greater influence on vocal signature is accounted for in the DTW weightings (see Methods). Moreover, when considering males and females separately, temporal parameters correlated with relatedness in males, whereas frequency parameters correlated with relatedness in females. Yet, there are no overall sex differences in churr calls (Chapter 4) and kinship does not correlate more strongly with call similarity in one sex. So, although frequency is clearly important, a single measure of overall similarity (e.g. DTW score) is still the most suitable way to assess vocal kin recognition in this species.

As expected, vocal similarity among native breeders was more similar than that between native and immigrant breeders, reflecting an increase in churr call variation at larger spatial scales. Geographic variation in bird song has been well-studied; the rate of song sharing declines sharply with distance in many species (e.g. great tits *Parus major*, McGregor & Krebs 1982; chaffinches *Fringilla coelebs*, Lynch *et al.* 1989). Playback experiments suggest birds are able to detect these differences, often showing a preference for local songs or dialects (Searcy 1997). Regional differences in contact calls have also been described in parrots (Wright 1996; Bradbury *et al.* 2001). Although there are functional differences between songs and calls, any vocalisations that are learned rather than innate are expected to show micro-geographic variation. In species such as long-tailed tits, that don't have a territorial song (Sharp & Hatchwell 2005), this variation may play an important role in mate choice, beyond the context of inbreeding avoidance. That local differences in churr call similarity have been detected based on DTW score, further establishes this as a reliable and informative measure of vocal similarity in long-tailed tits.

5.4.2 Call similarity and helping decisions

Crucially, helpers chose to help male breeders with more similar churr calls. This supports vocal similarity as a mechanism of kin recognition that permits kin-directed helping in long-tailed tits. The similarity of churr calls, but not triple calls, drives these decisions. This is not surprising, as there was no relationship between triple call similarity and relatedness. This does not rule out the use of triple calls for kin recognition, but suggests similarity in the churr alone is enough to discriminate kin when making helping decisions. As well as overall call similarity (DTW score), difference in maximum frequency also appeared to influence the decision to help. As with relatedness, the relationship between helping and DTW score was stronger than maximum frequency alone, yet in these analyses, and elsewhere (Sharp *et al.* 2005), frequency has been consistently revealed as the most distinguishing feature of long-tailed tit churr calls.

Kin-biased cooperation among social animals is often an all-or-nothing investment strategy in which perceived kin are helped at a certain rate (Russell & Hatchwell 2001; Komdeur 1994; Dickinson *et al.* 1996). However, recent studies on long-tailed tits (Nam *et al.* 2010) and bell miners (Wright *et al.* 2010) have shown that helpers can adjust their provisioning rate contingent on kinship. In bell miners, *mew* call similarity was identified as the cue to relatedness that allows helpers to make fine-scale facultative adjustments in their provisioning effort (McDonald & Wright 2011). Contrary to our predictions, although relatedness to male breeders explained a considerable amount of the variation in the provisioning rates of individual helpers, churr call similarity did not. However, there are important differences in the social organisation of bell miners and long-tailed tits, which may have important consequences for the evolution of their recognition systems. Bell miners live in complex societies in which membership of a coterie does not guarantee kinship and there is no evidence of a period of call learning, leading to the suggestion that mew calls are innate (McDonald & Wright 2011). Such genetically acquired cues would permit bell miners to recognise relatives in a population where familiarity does not signal kinship. The social structure of long-tailed tits on the other hand, is relatively simple. Although the close proximity of non-kin and kin of variable relatedness within breeding populations necessitates an active kin recognition mechanism, individuals can maximise fitness by directing their help towards close kin. Indeed, long-tailed tits are very good at discriminating first-order kin, but not second-order kin (Leedale *et al.* 2018), and a period for associative learning from close kin has already been identified (Sharp *et al.* 2005; Komdeur & Hatchwell 2009).

Therefore, my results suggest that while call similarity alone may permit individuals to recognise unfamiliar kin via phenotype-matching in bell miners, in long-tailed tits, vocal similarity is part of a combination of cues allowing individuals to recognise familiar individuals. This mechanism still appears to be error prone because non-relatives are sometimes helped. However, when the costs of helping are relatively low, and the potential benefits of helping kin are relatively high, a low threshold for acceptance of another individual as kin may be selected for (Hatchwell *et al.* 2014). The positive relationship between provisioning effort and relatedness to the brood may reflect a decision to help known kin at a higher rate than unknown individuals that could be more distant kin. A recognition mechanism based on prior association or familiarity would permit this adjustment.

5.4.3 Call similarity and pairing decisions

It was predicted that long-tailed tits use vocal similarity as a recognition cue in mate choice, allowing breeders to avoid mating with kin. Thus, vocal similarity was expected to be lower among breeding pairs than among unpaired opposite-sex dyads. Contrary to this prediction, long-tailed tit breeders of both sex paired with opposite-sex birds with more similar calls. Given that long-tailed tits avoid mating with close kin very effectively (Chapter 3), this result suggests that a combination of cues allows breeders to discriminate familiar kin, as appears to be the mechanism in the context of helping, rather than vocal similarity alone.

There are three related, biological explanations for finding greater call similarity between members of breeding pairs than expected by random mating. First, vocal similarity may be used to select unfamiliar, distant kin as partners, in order to maximise inclusive fitness, as predicted by models of optimal outbreeding (Bateson 1983) or so-called *kinbreeding* (Puurtinen 2011). Our results also show that native breeders that remained in their natal area sounded more similar to one another than to immigrant breeders. Hence, a second possibility is that long-tailed tits use vocal similarity to select unrelated partners that are from the local area. These breeders are more likely to have neighbours that are close relatives, and therefore an increased probability of attracting helpers. Third, during the non-breeding season, native birds remain in the natal area in large flocks with pair formation potentially occurring long before the breeding season starts. Immigrants, on the other hand, often disperse into the area at the start, or even during the breeding season, by which time many native breeders have already paired. This difference in timing could result in a tendency for locally hatched, similar sounding individuals to form pairs without any assessment of call similarity. This third explanation is perhaps the most parsimonious one, for both the greater call similarity and closer relatedness among breeding pairs reported here.

An alternative explanation for the greater call similarity between members of breeding pairs concerns to sampling. Interestingly, within this particular subset of pairs, for which there is vocal data, genetic relatedness is higher within pairs (r = 0.06 ± 0.15, n = 49) than between unpaired dyads (r = 0.007 ± 0.14, n = 1049). This was surprising, because this is higher than the within-pair relatedness observed when analysing the whole dataset, with which we have shown that relatedness among pairs is much lower than expected by random mating (Chapter 3). In two of the three years during which acoustic data was collected (2014-2015), the breeding population was much smaller than the annual average, and the effect this may have had on mating decisions is unknown. Indeed, fine-scale genetic structure does fluctuate between years (Appendix 4). Nevertheless, these findings do support earlier results that demonstrate a positive relationship between call similarity and relatedness.

4.3.4 Kin recognition mechanism for helping and pairing

Previous studies have shown that long-tailed tits recognise and direct help towards first-order kin (Russell & Hatchwell 2001, Sharp *et al.* 2005). The proportion of help given to second-order kin is low, and no different from that expected by chance

(Leedale *et al.* 2018). Similarly, mate choice models suggest long-tailed tits effectively avoid incest, but do not appear to distinguish second order or more distant kin during mate choice (Chapter 3). Nevertheless, call similarity among second-order kin is higher than among non-kin, suggesting that kin recognition is not based on call similarity alone. Instead, a combination of cues may be used to recognise familiar individuals. These may include auditory, visual, olfactory or behavioural cues. Indeed, long-tailed tits usually help at the nest of close relatives of whom they have had prior experience (Sharp *et al.* 2005; Napper & Hatchwell 2016). A kin recognition mechanism that relies on learned association is much less error-prone than phenotype matching, as long as there is a period of association. This requirement is satisfied in long-tailed tits, along with most cooperatively breeding birds (Komdeur & Hatchwell 1999). In contrast, phenotype-matching mechanisms may be open to cheats that have the phenotypic cues but not the associated kinship, and may be short-lived if the phenotypic cues are related to fitness, because directional selection will reduce phenotypic variation among conspecifics (Gardner & West 2007).

It is possible that long-tailed tits use different recognition mechanisms in the context of helping and pairing. Alternatively, the same mechanism may be used in both contexts, but with different outcomes depending on the position of the acceptance threshold, which may shift if the relative fitness costs and benefits associated with acceptance and rejection errors vary (Reeve 1989). The benefits of helping have been well-documented in this species (Hatchwell *et al.* 2014), and inbreeding has substantial fitness costs (Chapter 3). Furthermore, whereas non-kin are sometimes helped, breeding among close kin is extremely rare. So, although the same recognition mechanism is probably used in both contexts, this disparity in recognition errors may reflect a permissive recognition threshold in the context of helping, and a more restrictive recognition threshold in the context of mate choice.

In conclusion, I have demonstrated a relationship between relatedness and call similarity, and a preference for helpers to help breeders with more similar churr calls. This suggests that call similarity acts as a recognition cue to facilitate adjustments in helping behaviour with respect to relatedness. However, provisioning effort does not correlate with call similarity *per se*, and breeders do not appear to select partners with less similar calls. This suggests that vocal similarity is only part of a more complex mechanism of kin recognition in long-tailed tits.

Chapter 6
General Discussion

6.1 Introduction

The importance of kin selection in the evolution of cooperation has been debated because of the numerous examples of cooperation among non-kin (Clutton-Brock 2002; Riehl *et al*. 2013). In some cooperative species, direct fitness benefits explain the presence of unrelated helpers (Reyer 1984; Magrath & Whittingham 1997). However, even in systems that are thought to be entirely kin-selected, cooperation may occur among non-kin (Meade & Hatchwell 2010; Wright *et al.* 2010). The acceptance threshold model illustrates a trade-off between acceptance errors and rejection errors which shows that having a permissive threshold that allows individuals to always accept desirable recipients, means that some non-desirable recipients may also be accepted. If the costs of the acceptance errors do not outweigh the costs of the rejection errors, this would be an adaptive mechanism. Other situations may warrant a more restrictive threshold. For example, if the risk of costly inbreeding is high, strong kin discrimination during mate choice with low rates of acceptance error would be predicted. This strategy would come at the expense of a relatively high rate of rejection errors, in which compatible non-kin may be rejected as partners. These putative thresholds would depend not only on the cost/benefit ratio of the behaviour, and the relatedness of interacting individuals, but also on the distribution of kin and the variance in relatedness within populations (Agrawal 2001).

The redirected helping system of long-tailed tits creates breeding populations within which adults must actively choose who to help and who to mate with from a pool of potential social partners of varying degrees of kinship. The aims of this study were to determine the fine-scale genetic structure underlying breeding populations, and subsequently the strength of kin discrimination required for individuals to maximise fitness in this unusual, deceptively simple, cooperative breeding system. I also aimed to identity the recognition mechanisms used to direct cooperation towards close relatives and to avoid inbreeding; and whether the same cues and discrimination rules are used in both contexts. In this final chapter, the results are summarised and the implications with regard to the evolution of kin recognition mechanisms in long-tailed tits and other social species are discussed.

6.2 Summary of results

In Chapter 2, I used genetic, pedigree and behavioural data from a long-term dataset (1994-2016) to determine the fine-scale genetic structure of long-tailed tit breeding populations. Spatial autocorrelation analyses revealed fine-scale genetic structure, such that related adults were spatially clustered during breeding. Relatedness among neighbouring males was higher than that among neighbouring females, and genetic structure was also observed among opposite-sex breeders. I then compared the proportion of close, distant and non-kin helped, to those proportions expected by indiscriminate helping, based on this newly quantified genetic structure. Importantly, most close neighbours were unrelated and help was redirected towards first-order relatives much more often than one would expect if help was given randomly, or based on spatial distribution. However, failed breeders helped second-order kin at a similar rate to that expected by chance. This suggests that long-tailed tits are able to recognise first-order kin, but perhaps not second-order kin. However, over 40% nests were helped by second-order or non-kin, suggesting a permissive recognition threshold for helping behaviour.

Besides the indirect fitness benefits of helping, kin recognition may also be used to avoid inbreeding, particularly in species where opposite-sex kin interact as reproductive adults, as shown in Chapter 2. In Chapter 3, I investigated the fitness costs, risk and avoidance of inbreeding in long-tailed tits using the same long-term data as in Chapter 2. Examination of the typical pairing ranges of males and females revealed that mates are selected from within the range that kin remain clustered after dispersal, generating a substantial inbreeding risk. Importantly, this population was not inbred, and incest was extremely rare; only 0.5% pairings were between first or second-order kin. Despite its rarity, I was able to quantify the fitness costs of inbreeding in this population by examining the association between fitness and standardised multi-locus heterozygosity (H_s), an effective proxy for inbreeding depression in this population. I detected a weak, but positive correlation between H_s and direct fitness calculated from lifetime reproductive success data. Together, these results indicate strong selection for inbreeding avoidance in long-tailed tits.

To determine whether kin were actively avoided during mate choice, I compared the relatedness of chosen partners to that expected under null mate choice models, in which available birds of the opposite sex were selected randomly from within expanding spatial ranges. I found that relatedness to a chosen partner was significantly lower than expected under random mate choice within the range that most partners are selected, for both males and females. To determine the strength of kin discrimination that would permit this effective inbreeding avoidance, the relatedness within pairs was compared to that expected under mate choice models in which: (i) close, or (ii) close and distant kin were avoided as partners. The results were striking:

within-pair relatedness closely matched a mate choice model in which close kin, but not distant kin were avoided as partners. That this level of kin discrimination is enough to alleviate inbreeding was somewhat unexpected, but biologically sensible: because of the way kin are distributed, the probability of encountering second-order kin within pairing range is fairly low (Leedale *et al.* 2018).

The remainder of the thesis was dedicated to investigating the proximate mechanism(s) that permit the observed level of kin discrimination observed in long-tailed tits when making helping and pairing decisions. Vocalisations have already been shown to have an important function in kin recognition in long-tailed tits, and in Chapter 4, I developed a comprehensive method of measuring vocal similarity within and between individuals, in order to investigate how vocal cues are used to recognise relatives. Dynamic time warping (DTW) was selected as the most appropriate method to quantify overall similarity in churr and triple calls within and between individuals. DTW scores were used alongside specific call parameters to determine which sound characteristics shape similarity patterns. These analyses showed that both churr and triple calls were individually repeatable over time, and sexually monomorphic. Frequency measures, particularly the maximum frequency of calls, were the most repeatable call parameters. As these characteristics are particularly consistent within individuals, they may be particularly important for kin recognition.

In Chapter 5, I aimed to identify the recognition mechanism that permits kin-directed helping and active inbreeding avoidance in long-tailed tits. In particular, I asked whether vocal similarity may be used to assess degree of relatedness, and whether vocal similarity influences helping and pairing decisions. I found substantial variation in vocal similarity among breeders and churr call similarity did indeed correlate with kinship: the churr calls of first-order kin were more similar than the churr calls of second-order kin, and second-order kin were more similar than non-kin.

Although I found a similar pattern when using genetic relatedness estimates, this relationship was weaker and not statistically significant. Churr call similarity varied with philopatry: native breeders were more acoustically similar than those that dispersed into the study site as adults. As predicted, failed breeders chose to help males with more similar churr calls, but they did not adjust their provisioning rates according to vocal similarity, suggesting churr call similarity alone is not responsible for fine-tuning in helping effort. Finally, contrary to our predictions, breeders preferred partners with similar churr calls to themselves, although this could be confounded by micro-geographic variation in churr calls. This result also suggests that vocal similarity *per se* is not used by long-tailed tits to recognise kin in the context of mate choice.

6.3 Kin recognition in long-tailed tits

6.3.1 Vocal similarity and relatedness

Vocalisations are clearly an essential part of the kin-selected cooperative breeding system of long-tailed tits and I have shown, along with previous studies, that these calls are family- and individual-specific (Sharp & Hatchwell 2005; 2006). The novelty here is the detection of a positive correlation between churr call similarity and degree of kinship. This association provides a potential cue by which individuals could discriminate not only kin from non-kin but also kin or varying relatedness, an ability that has been indicated by the flexible provisioning effort of helpers contingent on genetic relatedness to the brood (Nam *et al.* 2010).

I did not find a strong relationship between vocal similarity and genetic relatedness estimated from microsatellite data. This does not exclude a genetic influence on vocal variation in long-tailed tits; indeed, high individual repeatability suggests some innate individual differences. However, the relatively stronger correlation between call similarity and kinship is in line with studies confirming through cross-fostering experiments that calls are learned (Sharp *et al.* 2005). A correlation between call similarity and degree of relatedness can still be acquired environmentally if calls are learned from both parents: logically, half-siblings that share one parent may be half as similar as full siblings that share both parents. However, the relationship between call similarity and relatedness is not linear: the difference in vocal similarity among first- and second-order kin is smaller than the difference in vocal similarity among second-order and non-kin. Which adults nestlings choose to imitate, and whether they can distinguish between helpers and parents during vocal learning, is still unknown. Further investigation into vocal learning would increase our understanding of this relationship and is a worthwhile avenue for further study.

Churr call similarity among native breeders was also greater than that between natives and immigrants, alluding to micro-geographic acoustic variation in long-tailed tits. It would be interesting to investigate the nature of this variation by examining how call similarity varies with distance and fine-scale genetic structure. Finally, although churr calls were individually repeatable and correlated with kinship, there was also a lot of acoustic variation, both within individuals and within different categories of kinship. This variation may make a recognition mechanism that relies entirely on vocal similarity unreliable. It is possible that a combination of cues could be used to assess kinship that alone would be ineffective. These may include: acoustic, visual, chemical, behavioural, or even spatial-temporal information.

6.3.2 Phenotype matching or associative learning?

Long-tailed tit nestlings learn their recognition templates from familiar kin during an associative learning period, when the cues themselves develop (Sharp *et al.* 2005). I wanted to determine the specificity of these templates; that is, whether they permit discrimination between close and distant kin, as well as kin and non-kin.

Nam *et al.* (2010) found that long-tailed tit helpers modified their effort according to their genetic relatedness to the helped brood, suggesting phenotype-matching may be used in a continuous assessment of kinship. This is in contrast to the binary decision rule based on prior association that was previously thought to be the putative mechanism of kin recognition in this species (Hatchwell *et al.* 2001b; Sharp *et al.* 2005). I found a similar effect of relatedness on helper provisioning effort (Chapter 5), corroborating Nam *et al.*'s (2010) study. Yet, although helpers directed their care towards nests belonging to acoustically similar males, we found no evidence that call similarity influenced the rate at which helpers provisioned nests. Neither did we find that breeders chose partners with less similar calls. Furthermore, we found strong evidence for very effective discrimination of first-order kin, but not second-order kin, both in the context of helping behaviour and mate choice.

In this population, it is very likely that first-order kin (parents, offspring and full-siblings) associate during the crucial period in which recognition templates are acquired. There are two instances in which this is not the case, both of which are relatively rare: extra-pair paternity produces half-siblings raised together (Hatchwell *et al.* 2002), and pair-bonds lasting more than one year may produce full-siblings that have not been raised together (Hatchwell *et al.* 2000). In contrast, it is less likely that second-order kin associate during this period (half-siblings, grand-parent/offspring and aunt-uncle/niece-nephew relations). Consequently, it seems highly probable that kin recognition is based on associative learning, rather than phenotype-matching. The important, and relatively frequent exception to this predictable relationship between familiarity and kinship is of course helper-offspring relationships: offspring presumably acquire recognition templates from those second-order kin or non-kin that helped them as a nestling. This may also explain the apparent existence of acceptance errors, in which non-kin help without gaining any known benefit (Meade & Hatchwell 2010). This argument is also supported by the observation that failed breeders sometimes help non-kin that provisioned them as a nestling (A. E. Leedale & B. J. Hatchwell, unpublished data). To tease apart these mechanisms, playback experiments of both familiar and unfamiliar kin of variable relatedness would be worthwhile, although gathering enough acoustic data for each of these categories would be extremely challenging.

Another important aspect of long-tailed tits social lives is their associations outside the breeding season. Previous studies have shown that winter flocks are made up of both kin and non-kin (Hatchwell *et al.* 2001a; McGowan *et al.* 2007; Napper *et al.* 2013). While template formation appears to occur before fledgling dispersal, whether or not these templates can be *updated* during these long periods of association is unknown. Association during winter is known to influence helping decisions the following spring (Napper & Hatchwell 2016), and this could have important consequences in terms of kin recognition.

6.3.3 Acceptance thresholds and kin discrimination rules

Previous studies on long-tailed tits have shown kin-biased helping and an ability to discriminate kin based on learned vocal cues (Hatchwell *et al.* 2001b; Russell & Hatchwell 2001; Sharp *et al.* 2005; Nam *et al.* 2010; Hatchwell *et al.* 2014). Whether, and how, long-tailed tits discriminate kin during mate choice was untested. That helpers provision more closely related broods at higher rates led to the prediction that adults can distinguish between kin of varying relatedness. The effective inbreeding avoidance demonstrated in Chapter 3 led to the suggestion that sophisticated kin discrimination may also be apparent during mate choice. However, throughout this thesis, the results suggest that long-tailed tits recognise close kin, with which they are likely to be familiar, but not more distant kin, with which they are more likely to be unfamiliar. Importantly, I have shown that this level of kin discrimination allows individuals to avoid incest and matches the observed patterns of helping.

In terms of acceptance thresholds, this level of kin discrimination does lead to a certain amount of error. Failed breeders may prefer to help close kin, but because kin cluster within the range that most helping occurs, there might still be a reasonable chance of gaining some indirect fitness by helping an unfamiliar individual because they could be distantly related. In this scenario, a permissive threshold would be adaptive. But, when non- and distant kin are helped, they are provisioned at a lower rate than close kin. These results fit a decision rule in which known, close kin are helped at a higher rate than less familiar or unfamiliar individuals who may be more distant kin. It would be interesting to investigate this further by seeing whether the helpers that choose to help non-kin did so because there were no close kin available to help.

Although this discrimination rule leads to acceptance errors in the context of helping, it actually creates a restrictive threshold for inbreeding avoidance. Because of the population structure, the probability of encountering a first-order relative when selecting a partner by chance is fairly low. The probability of encountering a first-order relative with which you are unfamiliar is even lower. However, there is good evidence of strong selection to avoid even this reasonably low chance. But, as the results show, this does not require recognition of distant kin. A more detailed understanding of exactly which birds are available to pair with would be useful, but this may not be possible without tracking individuals throughout their lives.

Overall, the results suggest that the same discrimination rule is used in both helping and pairing decisions. The exact cues are less clear because call similarity did not seem to be the only factor driving helping and pairing decisions. Importantly, fine-scale genetic structure is an important component when considering kin recognition evolution in this species. In this thesis, I have explored some of the fitness costs associated with inbreeding. Hamilton's rule has previously been tested, and supported, in the context of helping (Hatchwell *et al.* 2014), but applying Hamilton's rule to mate choice as well, would allow one to predict precisely how strong kin discrimination needs to be in each context for individuals to maximise fitness. A similar approach could be used to predict the strength of kin discrimination required across species, and, along with more empirical studies, could provide a deeper understanding of how recognition mechanisms have evolved across social animals.

6.4 Future directions in kin recognition research

In the previous section, I have discussed how my results fit into our current understanding of the long-tailed tit cooperative breeding system in particular. In the following section, I discuss my findings in relation to the implications for the wider field of kin recognition evolution, focusing on two emergent themes: (i) whether kin recognition involves a threshold or continuous assessment of relatedness, and (ii) the role of familiarity in kin recognition.

6.4.1 Threshold or continuous kin recognition?

The extent to which recognition cues permit kinship to be perceived on a continuous or binary scale is an important aspect of the kin recognition mechanism. How the information provided by such cues is utilised is still somewhat of a black box, but in most cooperative breeders, kin discrimination is considered binary: conspecifics are categorised as either kin or non-kin, based on a template-phenotype similarity threshold. But, kin discrimination could also be continuous, whereby individuals are able to assess their degree of relatedness to conspecifics, based on the extent of template-phenotype similarity.

When spatial location or group membership is used to recognise relatives, such as in Arabian babblers *Turdoides squamiceps* (Wright 1998), kin discrimination is binary. When recognition is based on phenotype, e.g. white-fronted bee eaters *Merops bullockoides* (Emlen and Wrege 1988), it may be binary or continuous, depending on the algorithm used to assess kinship. In order to tease apart these mechanisms, some studies have focused on whether the assumed recognition cues could be considered discrete or continuous traits. For example, green woodhoopoes (Radford 2005), Mexican jays *Aphelocoma wollweberi* (Hopp *et al.* 2001), stripe-backed wrens (Price

1998; 1999), superb starlings *Lamprotornis superbus* (Keen *et al.* 2013) and western bluebirds have family- or group-specific vocalisations (Açkay *et al.* 2013; 2014). Although their use in kin recognition remains undetermined, this suggests that they may be used in a threshold-based assessment of kinship. And yet, many cooperative bird species, such as the chestnut-crowned babbler *Pomatostomus rifuceps* (Crane *et al.* 2014), have individually distinct vocalisations, even though there are still no conclusive examples of individual recognition. Indeed, continuous traits are not necessarily perceived as such (Caves *et al.* 2018).

Binary or threshold kin discrimination is effective in most cooperative breeders because they often live in stable territories that, at least with regard to the helping sex, are mostly made up of first-order relatives. In the Seychelles warbler (Komdeur *et al.* 2004), splendid fairy-wren (Payne *et al.* 1988) and Galápagos mockingbird (Curry 1988), the presence of helpers is based on a decision rule related to prior association. Although the cues that permit recognition of familiar individuals have yet to be identified, there is no evidence to suggest that individuals perceive kinship as a continuous measure. As discussed above, this thesis supports the view that kin discrimination in long-tailed tits is also binary (although see Nam *et al.* 2010). Within cooperative breeders, a recognition cue that permits individuals to discriminate kin varying in relatedness has been identified only in the bell miner (Wright *et al.* 2010). Bell miners form large colonies often comprising hundreds of individuals. Within colonies, individuals are organised into *coteries* of numerous breeding pairs assisted by non-breeding helpers of varying relatedness that provision multiple nests within their coterie. From an early age, young interact with both related and unrelated group members, making spatial or association-based recognition unreliable. Instead, the provisioning effort of helpers correlates with their vocal similarity to the breeding male; an innate signal that also correlates with genetic relatedness (McDonald & Wright 2011).

In the context of mate choice, disentangling whether recognition is continuous or binary may be more challenging, as the decision to mate or not is a binary one. However, this can be assessed by quantifying the frequency of matings among individuals of varying degrees of kinship (Bateson 1983), or by comparing mating decisions to models in which kin of differing relatedness are avoided (Chapter 3). As discussed, this thesis supports binary kin recognition in long-tailed tits during mate choice; but other studies, such as those on the Japanese Quail, suggest continuous recognition (Bateson 1983).

The ability to assess the relatedness of conspecifics continuously may have been overlooked in some cases because of the way in which cooperative behaviour is measured. For example, some studies focus on the probability of helping, (Curry 1988; Creel *et al.* 1991; Dickinson *et al.* 1996), whereas others measure the amount of help given (Wright *et al.* 1999; Dunn *et al.* 1995; Clutton-Brock *et al.* 2001), and both have been measured in just a few (Emlen & Wrege 1988; Komdeur 1994; Russell & Hatchwell 2001; Nam *et al.* 2010). Similarly, some inbreeding avoidance studies measure incest rates, rather than the genetic relatedness of paired individuals (Dikcinson *et al.* 2016). In order to assess whether kin discrimination is a binary or continuous, the associated behaviour must be measured as a continuous response. More sophisticated observations and experiments are required to determine how relatedness is perceived. Understanding the sophistication of the recognition system is not only important in the context of cooperative breeding and mate choice, but for any situation that requires some form of recognition.

6.4.2 Familiarity

Determining the role of familiarity has proved to be a persistent problem in kin recognition studies. Commonly, this results from an inability to experimentally exclude the possibility that conspecifics have had a prior association. For example, in cross-fostering experiments, there is still some period of association before parents and offspring are separated, and without knowing precisely when kin recognition templates begin to form, it is impossible to rule out familiarity in these cases. Kin recognition may even begin during gestation (Hepper 1987) or incubation (Dowling *et al.* 2016). Furthermore, kin associations often persist into adulthood (Covas *et al.* 2006; Drobniak *et al*. 2015; Dickinson *et al.* 2016). How these associations may affect recognition cues is not fully understood, but adult associations can have a positive effect on cooperative behaviour (Carter & Wilkinson 2013; Napper & Hatchwell 2016).

Genetic relatedness has been shown to play more of an important role than familiarity for inbreeding avoidance in mammals (e.g. Green *et al.* 2015), and seabirds (Bonadonna & Sanz-Aguilar 2012). Although, in cooperatively breeding western bluebirds, incest avoidance is achieved by recognition of familiar individuals (Dickinson *et al.* 2016). There is also good evidence of humans avoiding familiar individuals as sexual partners (Shepher 1971). Future studies that examine the effect familiarity on mate choice, rather than controlling for it, could be rewarding.

The underlying issue is that familiarity is difficult to define and quantify. In the kin recognition literature, familiarity generally refers to some previous social interaction among individuals, particularly during early life stages (Hepper 1986; Komdeur & Hatchwell 1999). But, the nature of this interaction is often vague. Precisely when the interactions took place, how many interactions there were, their duration, and the specific behaviour and information transfer that took place during these interactions may influence how individuals are recognised and treated later in life. Social network analysis has been used to quantify the degree of familiarity between individuals and can give a broader view of how individuals associate during adulthood (Kurvers *et al.* 2013), particularly in kin-structured societies (Napper & Hatchwell 2016; McDonald *et al.* 2016). Some of these studies have shown that social interactions do not necessarily reflect genetic relatedness (Godfrey *et al.* 2014). Importantly, this can depend on the nature of the behaviour used to build the social network (Madden *et al.* 2012). More studies are required which implement social network analysis to quantify familiarity and assess the relative importance of both familiarity and kinship.

Another challenge when determining the role of familiarity in kin recognition is that how individuals actually recognise familiar individuals is often unknown. Future work should aim to correlate phenotypic cues with genetic relatedness, kinship and degree of familiarity quantified using social network analysis. Even then, it may be difficult to determine cause and effect: frequent interactions may lead to an increase in phenotypic similarity among interacting individuals in a positive feedback loop. Vocal convergence can be adaptive for coordinated foraging (Bradbury & Balsby 2016), particularly when birds forage in annual winter flocks that are disbanded each spring. For example, black-capped chickadees *Parus atricapillus*, exhibit vocal plasticity throughout adulthood and vocal convergence can occur within a week of winter flock formation (Nowicki 1989). However, such species do not breed cooperatively and individuals do not gain indirect fitness benefits from associating with kin. In kin-selected systems, it may be unlikely for kin recognition cues to be *updated* during adulthood, particularly if interactions among kin and non-kin were frequent, although there are some examples (Keen *et al.* 2013; Radford 2005). In long-tailed tits, vocalisations do not differ significantly over an individual's lifetime once learned, but how individuals recognise familiar kin is still unknown. More studies that investigate the plasticity of putative recognition cues are needed. It is also essential to control for familiarity or kinship when considering how these cues develop.

6.5 Conclusion

In conclusion, this is the first investigation in to a recognition mechanism that may be used for both kin-selected cooperation and inbreeding avoidance in a wild vertebrate population. These results demonstrate how population structure, shaped by life-history, dispersal and mating system, creates selection for recognition mechanisms of varying sophistication. The strength of discrimination required to maximise fitness may vary according to context, and this can be quantified through empirical observations of natural populations. The unusual cooperative system of the long-tailed tit has made this species an excellent system in which to study kin recognition and to consider the social and ecological drivers of variation in kin discrimination complexity across cooperative species.

Appendices

Appendix 1. Reconstruction of the long-tailed tit social pedigree

1.1 Parentage analysis

The long-tailed tit social pedigree was constructed using detailed field observations of marked birds in the Rivelin Valley population from 1994-2016 (n = 3486). A small proportion (<2%) of recruits first ringed as adults were philopatric birds that were not ringed as chicks because they fledged from inaccessible nests. To assign parentage to these birds, I used the likelihood approach implemented in *CERVUS version 3.0.7* (Kalinowski *et al.* 2007) to perform a parentage analysis on all genotyped putative immigrants that appeared in the population from 1997-2017; analysis was done on a year by year basis, with all adults present in the population in the recruit's birth-year considered as potential parents, based on the assumption that long-tailed tits disperse from their natal area during their first winter (McGowan *et al.* 2003; Sharp *et al.* 2008a). Due to incomplete sampling early in the project, 1994-1996 were excluded. For each year, simulations were run with 100,000 offspring, assuming: 80% of candidate parents were sampled (proportion of recruits genotyped), 98.3% of loci were typed (calculated from the allele frequency data) and a mistyping rate of 0.01, with 10 as the minimum number of typed loci. Previous analyses have shown that the true error rate is less than 1% (M. Simeoni, S. P. Sharp & B. J. Hatchwell, unpublished data). I then carried out parent pair analyses with reference to critical delta scores for 99% confidence derived from the simulations. Having identified parent-offspring relationships, I also compared this to the r_{QG} estimates (Queller & Goodnight 1989) calculated for each pair. Parent-offspring relationships were considered to be genuine only if the parents were known to have paired together and fledged unringed chicks in the recruit's birth-year.

1.2 Sibship reconstruction

As this is an open population, our social pedigree is incomplete. However, long-tailed tits are known to disperse in sibling coalitions during their first year (McGowan *et al.* 2003; Sharp *et al.* 2008b). I tested for the presence of full siblings among each yearly cohort of genotyped immigrants, using the *Descending Ratio* sibship reconstruction algorithm implemented in *KINGROUP version 2* (Konovalov *et al.* 2004). A primary hypothesis of full siblings was tested against a null hypothesis of unrelated pairs; separate analyses were carried out for each year from 1995 to 2016. Having identified sibships, I performed likelihood ratio tests based on pairwise r_{QG} estimates using the same hypotheses and carried out 10,000 simulations. The results from the sibship reconstruction were considered to match those from likelihood ratio tests if individuals in a dyad with a significant likelihood ratio ($p < 0.05$) were placed in the same sibling group (or if those in a dyad with a non-significant likelihood ratio were placed in different groups). Sibling relationships were added to the social pedigree only if all siblings in a given group matched and had high pairwise r_{QG} estimates. Birds were included in our reconstructed social pedigree only if their parentage was known, or if they had been assigned a sibling group (n = 3068).

Appendix 2. Distribution of kinships among breeding long-tailed tits

Using both genetic data from microsatellites (n = 1022) and the social pedigree (n = 866), I calculated the frequency of first order kin ($r_A \geq 0.5$; $r_{QG} \geq 0.25$), second order kin ($0.5 > r_A \geq 0.25$; $0.25 \geq r_{QG} \geq 0.125$) and non-kin ($r_A < 0.25$; $r_{QG} < 0.125$) available to helpers in the breeding population. The frequency of these relationships was measured over four defined distance bands, based on the nest at which birds first appeared as an adult: 0-300m, 300-600m, 600-900m, >900m. The relationship between distance and the frequency of kin relationships was analysed using a Pearson's Chi-squared test. Within distance bands, the distribution of male and female kin was compared using a series of Pearson's Chi-squared tests.

Based on genotypes, on average the long-tailed tit breeding population was made up of 5.1% first order relationships (n = 3702), 13.3% second order relationships (n = 9715) and 81.6% non-kin relationships (n = 59652). Based on the social pedigree, the population contained just 1.8% first order (n = 1038) and 1.2% second order kin relationships (n = 677), with the remaining 97% relationships between non-kin (n = 55546). Among males, relationships measured using genetic data (6.1% first order, n = 1219; 12.7% second order, n = 2584; 81.2% non-kin, n = 16476) were again higher than those estimated from the social pedigree (2.2% first order, n = 358; 1.6% second order, n = 271; 96.2% non-kin, n = 16300). Among females, the same pattern emerged, with genetic data (4.6% first order, n = 732; 13.5% second order, n = 2174; 81.9% non-kin, n = 13135) resulting in higher estimates of kinship than the social pedigree (1.8% first order, n = 206; 0.8% second order, n = 90; 97.4% non-kin, n = 11389). These rather low observed frequencies of kin relationships are consistent with the low mean r_{QG} estimates observed across the population as a whole.

As the distance between dyads of birds increased, the proportion of kin decreased (genetic data: $\chi^2 = 1101.3$, df = 6, $p < 0.001$, Fig. A2.1a; social pedigree: $\chi^2 = 2144.3$, df = 6, $p < 0.001$, Fig. A2.1b). This strong distance effect was apparent among males (genetic data: $\chi^2 = 696.7$, df = 6, $p < 0.001$, Fig. A2.1c; social pedigree: $\chi^2 = 1476.9$, df = 6, $p < 0.001$, Fig. A2.1d) and among females (genetic data: $\chi^2 = 160.91$, df = 6, $p < 0.001$, Fig. A2.1e; social pedigree: $\chi^2 = 279.92$, df = 6, $p < 0.001$, Fig. A2.1f). The proportion of kinships was greater among males than among females within 300m (genetic data: $\chi^2 = 51.17$, df = 2, $p < 0.001$; social pedigree: $\chi^2 = 67.42$, df = 2, $p < 0.001$), at 300-600m (genetic data: $\chi2 = 10.19$, df = 2, $p < 0.01$; social pedigree: $\chi2 = 10.65$, df = 2, $p < 0.01$), at 600-900m for the social pedigree ($\chi2 = 14.64$, df = 2, $p < 0.001$) but not for genetic data ($\chi2 = 0.51$, df = 2, $p = 0.78$), and over 900m (genetic data: $\chi2 = 6.43$, df = 2, $p < 0.05$; social pedigree: $\chi2 = 10.80$, df = 2, $p < 0.01$).

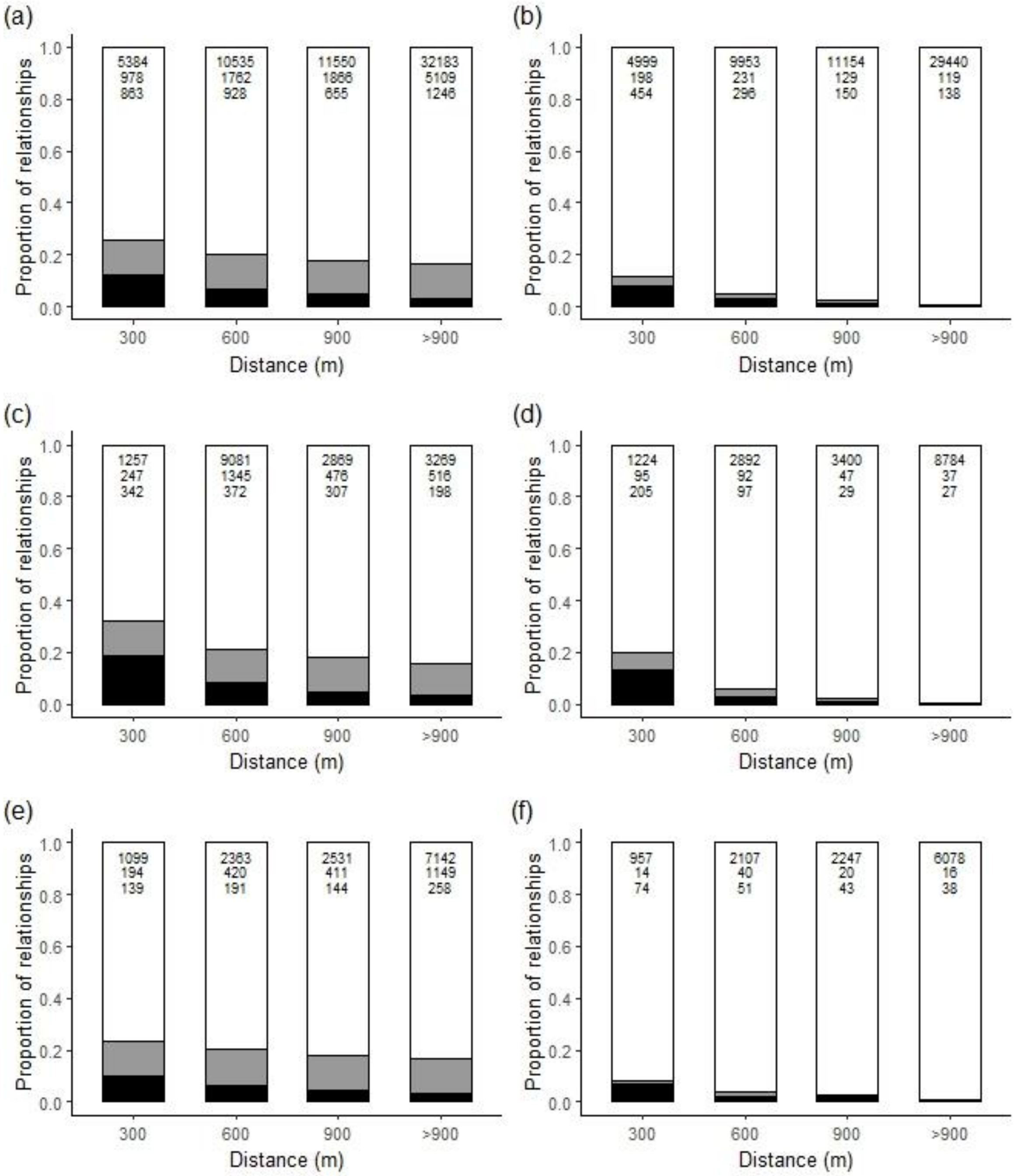

Fig. A2.1. Proportion of 1st order (black), 2nd order (grey) and non-kin (white) in the breeding population over four bands of distance between dyads. (a) genotype data for all birds, (b) social pedigree for all birds, (c) genotype data for males, (d) social pedigree for males, (e) genotype data for females, (f) social pedigree for females. The number of pairwise relationships in each distance band is displayed (top to bottom: non-, 2nd order and 1st order kin).

Appendix 3. Allele distributions across microsatellite markers

Table A3.1. Distribution of alleles among 17 microsatellite loci used to estimate genetic relatedness in long-tailed tits.

Locus	Number of alleles
CAM01	26
CAM23	5
Tgu_01.040	7
Tgu_04.012	5
Tgu_05.053	9
Tgu_13.017	5
Ase.37	18
CAM03	14
CAM15	8
Pca.4	13
Ase18	17
Ase64	6
Hru2	6
Hru6	53
Pca3	7
PmaD22	43
Ppi2	22
Total	**264**

Appendix 4. Annual male-female genetic structure

Table A4.1. Temporal variation in mean pairwise relatedness, r_{QG}, between male and female long-tailed tits. Approximate SE were calculated by jackknifing over loci.

Year	r_{QG}	SE	n (birds)	n (comparisons)
1994	0.2459	0.0501	10	24
1995	0.0195	0.0213	28	196
1996	0.0214	0.0109	48	551
1997	0.0221	0.0115	57	806
1998	0.0079	0.0112	55	756
1999	0.0062	0.0156	68	1152
2000	0.0017	0.0128	93	2160
2001	0	0.019	39	374
2002	0.0079	0.0101	75	1386
2003	-0.0027	0.0071	87	1862
2004	0.0082	0.0064	126	3944
2005	0.0026	0.0087	91	2070
2006	-0.0001	0.0081	98	2385
2007	0.0033	0.0084	89	1968
2008	0.0168	0.0061	119	3498
2009	0.0189	0.0088	75	1404
2010	0.0185	0.011	106	2808
2011	0.0139	0.0081	104	2703
2012	0.007	0.0091	111	3068
2013	0.0153	0.0096	55	756
2014	0.0019	0.0118	56	783
2015	0.0146	0.013	58	841
2016	0.0207	0.0142	71	1254

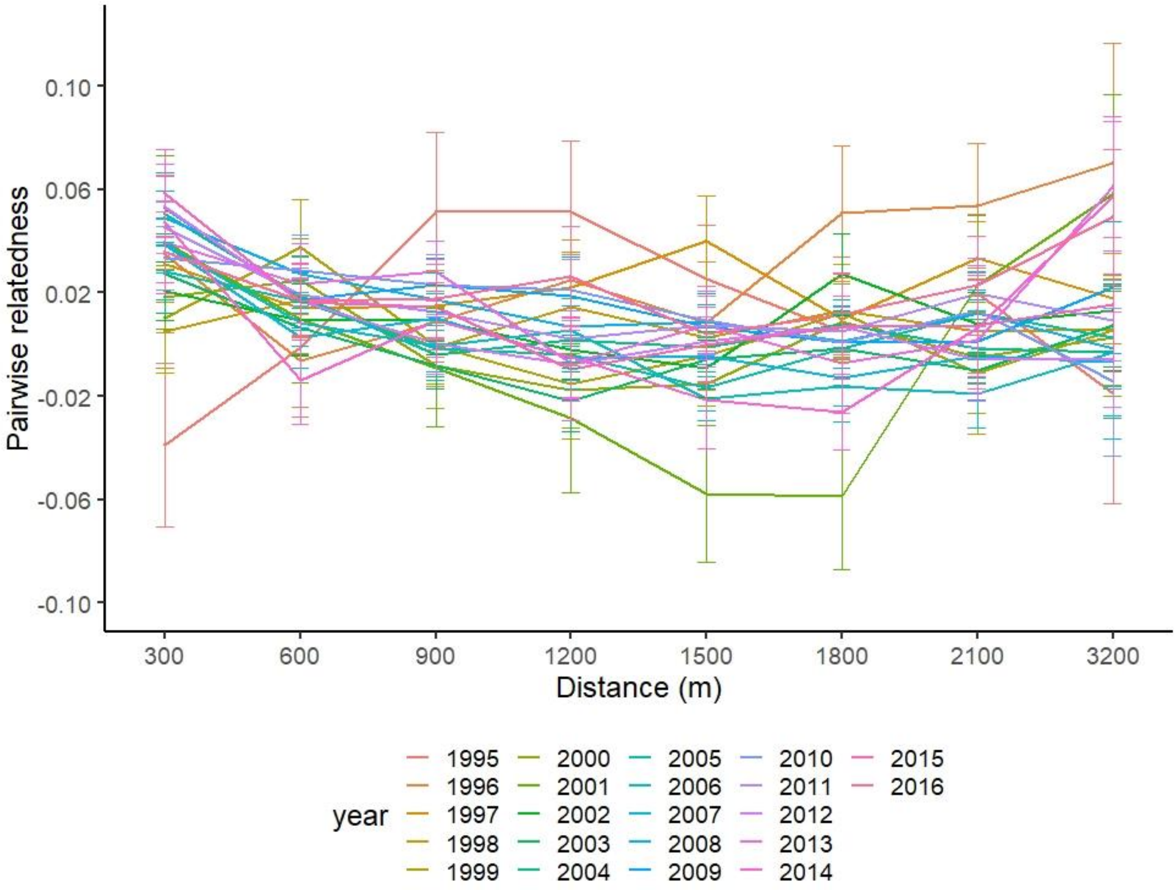

Fig. A4.1. Mean pairwise relatedness (r_{QG}) in a long-tailed tit population over eight bands of distance between males and females from 1995-2016. Data from 1994 is excluded due to low sample size. Error bars approximate SE of relatedness estimates from jackknifing over loci.

Appendix 5. Cases of inbreeding in long-tailed tits

Table A5.1. Cases of inbreeding in the Rivelin valley long-tailed tit population (1994-2016) according to the social pedigree ($r > 0.125$). Genetic relatedness estimates (r_{QG}) and social relationships are also reported.

Year/Pair	Male	Female	r_{QG}	r	Relationship
2007/21	1945	1943	0.253	0.5	Full siblings (reared together)
2000/04	797	1080	0.219	0.25	Maternal ½ siblings; male was cross-fostered
2011/15	2339	2408	0.212	0.25	Paternal uncle/niece
1996/19	698	633	0.273	0.125	Maternal ½ aunt/nephew
2004/50	1516	1611	-0.151	0.125	Paternal cousins
2007/12	1761	2052	0.267	0.125	Paternal cousins

Table A5.2. Cases of inbreeding in the Rivelin valley long-tailed tit population (1994-2016) according to genetic relatedness estimates ($r_{QG} > 0.125$). Pedigree relatedness (r) and social relationships are also reported.

Year/Pair	Male	Female	r_{QG}	r	Relationship (if known)
2012/28	2283	468	0.571	0	
2003/27	265	266	0.452	0	
2008/28	435	458	0.372	0	
2004/06	290	264	0.351	0	
2015/21	3079	3185	0.347	0	
2005/01	290	356	0.332	0	
2014/05	3042	2947	0.327	0	
2004/04	307	306	0.322	0	
2007/71	446	2090	0.301	0	
2011/05	1978	520	0.297	0	
2015/11	3171	3170	0.283	0	
2004/37	294	242	0.282	0	
1996/19	698	633	0.273	0.125	Maternal ½ aunt/nephew
2016/19	3286	3285	0.269	0	
2015/08	2909	3182	0.267	0	
2007/12	1761	2052	0.267	0.125	Paternal cousins
2015/18	3179	2756	0.263	0	
2007/07	354	368	0.262	0	
2007/21	1945	1943	0.254	0.5	Full-sibs (reared together)
2007/69	448	390	0.25	0	
2012/51	2592	2762	0.244	0	
2006/32	1756	313	0.241	0	
2000/42	181	1098	0.229	0	
2016/17	3132	3074	0.226	0	Unrelated
2015/06	3092	3059	0.226	0	
2007/60	443	442	0.224	0	
2012/55	2712	2758	0.221	0	
1998/16	108	123	0.22	0	
2004/65	327	352	0.219	0	
2000/04	797	1080	0.219	0.25	Maternal ½ siblings; male was cross-fostered
2013/12	2779	2782	0.219	0	
2002/04	195	1192	0.217	0	
2003/09	1475	256	0.215	0	

2009/02	465	2124	0.215	0	
2011/04	2688	2687	0.213	0	
2011/15	2339	2408	0.212	0.25	Paternal uncle/niece
2014/08	2967	2782	0.213	0	
2002/30	1339	249	0.209	0	
2004/27	301	285	0.201	0	
2005/38	1589	368	0.198	0	
2009/17	2097	520	0.197	0	
2008/68	2137	496	0.195	0	
1998/11	848	819	0.195	0	unrelated
2015/05	2975	3055	0.193	0	
2005/35	1601	281	0.192	0	
2004/18	299	1591	0.191	0	
2013/29	2877	2916	0.184	0	
2007/55	2068	441	0.183	0	
2008/30	431	485	0.182	0	
1999/21	149	148	0.18	0	
2008/62	367	469	0.178	0	
2002/19	1367	207	0.176	0	
1999/26	166	140	0.174	0	
1998/22	136	46	0.174	0	
2011/07	3162	2693	0.172	0	
2016/04	3165	3274	0.172	0	
2007/14	404	324	0.169	0	
2007/39	505	504	0.168	0	
2011/41	2700	2701	0.167	0	
2012/13	2738	2739	0.166	0	
2012/62	2229	520	0.165	0	
2003/29	1347	242	0.163	0	
2002/12	229	228	0.161	0	
2001/25	214	1042	0.160	0	
2008/16	453	2060	0.159	0	
2005/31	1602	363	0.157	0	
2008/21	2050	463	0.157	0	
2005/08	1689	323	0.157	0	
2016/05	3206	3276	0.155	0	
2006/22	354	270	0.155	0	
2008/13	381	1941	0.153	0	
2013/15	2723	2722	0.153	0	
2008/20	1930	1402	0.151	0	unrelated
2009/01	367	2090	0.151	0	
2003/24	1349	217	0.149	0	
2004/16	303	302	0.147	0	
2009/27	2159	483	0.147	0	
2012/12	2740	2695	0.146	0	
2004/54	308	320	0.146	0	
2005/47	1739	377	0.146	0	
2010/09	2381	2380	0.145	0	
2016/26	3087	3061	0.145	0	
2011/25	2717	2718	0.143	0	
2000/29	1071	169	0.141	0	
2010/43	429	520	0.141	0	
2012/01	2714	2705	0.14	0	
2014/03	3007	3056	0.138	0	
2006/48	1512	350	0.137	0	

2008/50	476	270	0.133	0	
1999/30	965	863	0.133	0	unrelated
1997/13	108	107	0.132	0	
2014/37	2907	3009	0.132	0	
2006/10	386	356	0.132	0	
2004/39	1641	332	0.131	0	
2006/59	409	1402	0.127	0	
2007/42	429	430	0.126	0	
1997/31	671	103	0.126	0	
2001/01	201	200	0.125	0	

References

Adams M. J., Robinson M. R., Mannarelli M-E. & Hatchwell B. J. (2015). Social genetic and social environment effects on parental and helper care in a cooperatively breeding bird. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20150689.

Agrawal A. (2001). Kin recognition and the evolution of altruism. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 1099-1104.

Akçay Ç., Hambury K. L., Arnold J. A., Nevins A. M. & Dickinson J. L. (2014). Song sharing with neighbours and relatives in a cooperatively breeding songbird. *Animal Behaviour*, **92**, 55-62.

Akçay Ç., Swift R. J., Reed V. A. & Dickinson J. L. (2013). Vocal kin recognition in kin neighborhoods of western bluebirds. *Behavioral Ecology*, **24**, 898-905.

Arnold K. E. & Owens I. P. (1998). Cooperative breeding in birds: a comparative test of the life history hypothesis. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 739-745.

Axelrod R. & Hamilton W. D. (1981). The evolution of cooperation. *Science*, **211**, 1390-1396.

Balloux F. & Lugon-Moulin N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155-165.

Barnard C. J. (1991). Kinship and social behaviour: The trouble with relatives. *Trends in Ecology & Evolution*, **6**, 310–312.

Barnard C. J., Hurst J. L. & Aldhous P. (1991). Of mice and kin: the functional significance of kin bias in social behaviour. *Biological Reviews*, **66**, 379-430.

Bateman A. J. (1948). Intra-sexual selection in Drosophila. *Heredity*, **2**, 349-368.

Bates D., Maechler M., Bolker B. & Walker S. (2015). Fitting Linear Mixed-Effects Models Using *lme4*. *Journal of Statistical Software*, **67**, 1-48.

Bateson P. (1978). Sexual imprinting and optimal outbreeding. *Nature*, **273**, 659-660.

Bateson P. (1982). Preferences for cousins in Japanese quail. *Nature*, **295**, 236–237.

Bateson P. (1983). Optimal outbreeding. In: *Mate Choice* (ed. P. Bateson), pp. 257-277. Cambridge, UK: Cambridge University Press.

Beckerman A. P., Sharp S. P. & Hatchwell B. J. (2011). Predation and kin-structured populations: an empirical perspective on the evolution of cooperation. *Behavioral Ecology*, **22**, 1294-1303.

Beecher M. D. (1982). Signature systems and kin recognition. *American Zoologist*, **22**, 477-490.

Beecher M. D. (1991). Successes and failures of parent-offspring recognition. In: *Kin Recognition* (ed. P. G. Hepper), pp. 94-124. Cambridge, UK: Cambridge University Press.

Bekoff M. (1992). Kin recognition and kin discrimination. *Trends in Ecology & Evolution*, **7**, 100.

Benabentos R., Hirose S., Sucgang R., Curk T., Katoh M., Ostrowski E. A., Strassmann J. E., Queller D. C., Zupan B., Shaulsky G. & Kuspa A. (2009). Polymorphic Members of the *lag* Gene Family Mediate Kin Discrimination in *Dictyostelium*. *Current Biology*, **19**, 567–572.

Bérénos C., Ellis P. A., Pilkington J. G. & Pemberton J. M. (2016). Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. *Molecular Ecology*, **25**, 3152–3168.

Berg K. S., Delgado S., Okawa R., Beissinger S. R. & Bradbury J. W. (2011). Contact calls are used for individual mate recognition in free-ranging green-rumped parrotlets, *Forpus passerinus*. *Animal Behaviour*, **81**, 241-248.

Bergmüller R. & Taborsky M. (2005). Experimental manipulation of helping in a cooperative breeder: helpers “pay to stay” by pre-emptive appeasement. *Animal Behaviour*, **69**, 19–28.

Bergmüller R., Heg D. & Taborsky M. (2005). Helpers in a cooperatively breeding cichlid stay and pay or disperse and breed, depending on ecological constraints. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 325-331.

Biedrzycki M. L & Bais H. P. (2010). Kin recognition in plants: a mysterious behaviour unsolved. *Journal of Experimental Botany*, **61**, 4123-4128.

Blyton M. D. J., Banks S. C. & Peakall R. (2015). The effect of sex-biased dispersal on opposite-sexed spatial genetic structure and inbreeding risk. *Molecular Ecology*, **24**, 1681–1695.

Bonadonna F. & Sanz-Aguilar A. (2012). Kin recognition and inbreeding avoidance in wild birds: the first evidence for individual kin-related odour recognition. *Animal Behaviour*, **84**, 509-513.

Bonneaud C., Chastel. O., Fedeici P., Westerdahl. H & Sorci G. (2006). Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1111-1116.

Boomsma J. J. (2009). Lifetime monogamy and the evolution of eusociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 3191–3207.

Bourke A. F. G. (2011). *Principles of Social Evolution*. Oxford Series in Ecology and Evolution (eds. P. H. Harvey, R. M. May, C. H. Godfray & J. A. Dunne). Oxford: Oxford University Press.

Boyse E. A., Beauchamp G. K., Yamazaki K. & Bard J. (1991). Genetic components of kin recognition in mammals. In: *Kin Recognition* (ed. P. G. Hepper), pp.148-161. Cambridge, UK: Cambridge University Press.

Bradbury J. W. & Balsby T. J. S. (2016). The functions of vocal learning in parrots. *Behavioral Ecology and Sociobiology*, **70**, 293–312.

Bradbury J. W., Cortopassi K. A. & Clemmons J. R. (2001). Geographical Variation in the Contact Calls of Orange-Fronted Parakeets. *The Auk*, **118**, 958–972.

Brooker M. G., Rowley I., Adams M. & Baverstock P. R. (1990). Promiscuity: An inbreeding avoidance mechanism in a socially monogamous species? *Behavioral Ecology and Sociobiology*, **26**, 191-199.

Brown J. L. (1987). *Helping and communal breeding in birds*. Princeton, NJ: Princeton University Press.

Byers J. A. & Bekoff M. (1986). What does “kin recognition” mean? *Ethology*, **72**, 342-345.

Cant M. A., Vitikainen E. & Nichols H. J. (2013). Demography and Social Evolution of Banded Mongooses. *Advances in the Study of Behavior*, **45**, 407–445.

Capodeanu-Nägler A., Rapkin J., Sakaluk S. K., Hunt J. & Steiger S. (2014). Self-recognition in crickets via on-line processing. *Current Biology*, **24**, R1117-R1118.

Caspers B. A., Hoffman J. I., Kohlmeier P., Krüger O. & Krause E. T. (2013). Olfactory imprinting as a mechanism for nest odour recognition in zebra finches. *Animal Behaviour*, **86**, 85-90.

Catchpole C. K. & Slater P. J. B. (2008). *Bird Song: Biological Themes and Variations*. 2nd Edition. Cambridge, UK: Cambridge University Press.

Caves E. M., Green P. A., Zipple M. N., Peters S., Johnsen S. & Nowicki S. (2018). Categorical perception of colour signals in a songbird. *Nature*, **560**, 365–367.

Chapman J. R. & Sheldon B. C. (2011). Heterozygosity is unrelated to adult fitness measures in a large, noninbred population of great tits (*Parus major*). *Journal of Evolutionary Biology*, **24**, 1715-1726.

Chapman J. R., Nakagawa S., Coltman D. W., Slate J. & Sheldon B. C. (2009). A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology*, **18**, 2746–2765.

Charlesworth B. & Charlesworth D. (1999). The genetic basis of inbreeding depression. *Genetical Research*, **74**, 329-340.

Charlesworth D. & Charlesworth A. (1987). Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology and Systematics*, **18**, 237-268.

Charlesworth D. & Willis J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783–796.

Choe J. C. & Crespi B. J. (1997). *The evolution of mating systems in insects and arachnids*. Cambridge: Cambridge, UK: Cambridge University Press.

Clark C. W., Marler P. & Beeman K. (1987). Quantitative-analysis of animal vocal phonology - an application to swamp sparrow song. *Ethology*, **76**, 101-115.

Clarke F. M. & Faulkes C. G. (1999). Kin discrimation and female mate choice in the naked mole-rat *Heterocephalus glaber*. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1995–2002.

Clarke M. F., Fitz-Gerald G. F. (1994). Spatial organisation of the cooperatively breeding bell miner *Manorina melanophrys*. *Emu*, **94**, 96–105.

Clutton-Brock T. H., Brotherton P. N. M., Russell A. F., O'Riain M. J., Gaynor D., Kansky R., Griffin A., Manser M., Sharpe L., McIlrath G. M., Small T., Moss A. & Monfort S. (2001). Cooperation, control, and concession in meerkat groups. *Science*, **291**, 478–481.

Clutton-Brock T. H. & Lukas D. (2012). The evolution of social philopatry and dispersal in female mammals. *Molecular Ecology*, **21**, 472–492.

Clutton-Brock T. H. (2016). *Mammal Societies*. Chichester, West Sussex, UK: John Wiley & Sons.

Clutton-Brock T. H., Hodge S. J., Flower T., Spong G. F. & Young A. J. (2010). Adaptive suppression of subordinate reproduction in cooperative mammals. *The American Naturalist,* **176**, 664–673.

Cockburn A. (2004). Mating systems and sexual conflict. In: *Ecology and Evolution of Cooperative Breeding in Birds* (eds. W.D. Koenig & J. L. Dickinson), pp 81-101. Cambridge, UK: Cambridge University Press.

Cockburn A., Osmond H. L., Mulder R. A., Green D. J. & Double M. C. (2003). Divorce, dispersal and incest avoidance in the cooperatively breeding superb fairy-wren *Malurus cyaneus*. *Journal of Animal Ecology*, 72, 189–202.

Coffin H. R., Watters J. V. & Mateo J. M. (2011). Odour-Based Recognition of Familiar and Related Conspecifics: A First Test Conducted on Captive Humboldt Penguins (*Spheniscus humboldti*). *Plos One*, 6, 1-4.

Collins S. A. (2004). Vocal fighting and flirting: the functions of birdsong. In: *Nature's Music: The Science of Birdsong* (eds. P. Marler & H. Slabbekoorn), pp. 39-79. London, UK: Elsevier Academic Press.

Coltman D. W., Pilkington J. G., Smith J. A. & Pemberton J. M. (1999). Parasite-mediated selection against inbred Soay sheep in a free-living island population. *Evolution*, **53**, 1259-1267.

Cornwallis C. K., West S. A. & Griffin A. S. (2009). Routes to indirect fitness in cooperatively breeding vertebrates: kin discrimination and limited dispersal. *Journal of Evolutionary Biology*, **22**, 2445-2457.

Cornwallis C. K., West S. A., Davis K. E. & Griffin A. S. (2010). Promiscuity and the evolutionary transition to complex societies. *Nature*, **466**, 969.

Cortopassi K. A. & Bradbury J. W. (2000). The comparison of harmonically rich sounds using spectrographic cross-correlation and principal coordinates analysis. *Bioacoustics*, **11**, 89-127.

Coster A. (2013). *Pedigree: pedigree functions*. R package version 1.4. Available at: http://CRAN.R-project.org/.

Couvillon M. J., Robinson E. J. H., Atkinson B., Child L., Dent K. R. and Ratnieks F. (2008). En garde: rapid shifts in honeybee, *Apis mellifera* guarding behaviour are triggered by onslaught of conspecific intruders. *Animal Behaviour*, **76**, 1653-1658.

Covas R. & Griesser M. (2007). Life History and the Evolution of Family Living in Birds. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1349-1357.

Covas R., Dalecky A., Caizergues A., Doutrelant C. (2006). Kin associations and direct vs indirect fitness benefits in colonial cooperatively breeding sociable weavers *Philetairus socius*. *Behavioral Ecology and Sociobiology*, **60**, 323–331.

Cramp S. & Perrins C. M. (1993). *Birds of the Western Palearctic, Vol. 7*. Oxford, UK: Oxford University Press.

Crane J. M. S., Pick J. L., Tribe A. J., Vincze E., Hatchwell B. J. & Russell A. F. (2015). Chestnut-crowned babblers show affinity for calls of removed group members: a dual playback without expectancy violation. *Animal Behaviour*, **104**, 51–57.

Creel S. R., Monfort S. L., Wildt D. E. & Waser P. M. (1991). Spontaneous lactation is an adaptive result of pseudopregnancy. *Nature*, **351**, 660–662.

Crochet P. A. (2000). Genetic structure of avian populations - allozymes revisited. *Molecular Ecology*, **9**, 1463–1469.

Crozier R. H. (1986). Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution*, **40**, 1100-1101.

Curry R. L. & Grant P. R. (1990). Galápagos Mockingbirds. In: *Cooperative breeding in birds: Long-term studies of ecology and behavior* (eds. P. B Stacey & W. D. Koenig), pp.291-331. Cambridge, UK: Cambridge University Press.

Curry R. L. (1988). Influence of kinship on helping behaviour of Galápagos Mockingbirds. *Behavioral Ecology and Sociobiology*, **22**, 141-152.

Daniels S. & Walters J. (2000). Inbreeding Depression and Its Effects on Natal Dispersal in Red-Cockaded Woodpeckers. *The Condor,* **102**, 482-491.

Darwin C. (1859). *On the Origin of Species by Means of Natural Selection*. London, UK: Murray.

David P., Delay B., Berthou P. & Jarne P. (1995). Alternative models for allozyme-associated heterosis in the marine bivalve Spisula ovalis. *Genetics*, **139**, 1719-1726.

Dawkins R. (1976). *The Selfish Gene*. Oxford, UK: Oxford University Press.

Dawkins R. (1982). *The Extended Phenotype: The long reach of the gene*. Oxford, UK: Oxford University Press.

Díaz-Muñoz S. L., DuVal E. H., Krakauer A. H. & Lacey E. A. (2014). Cooperating to compete: altruism, sexual selection and causes of male reproductive cooperation. *Animal Behaviour*, **88**, 67-78.

Dickinson J. L. & Hatchwell B. J. (2004). Fitness consequences of helping. In: *Ecology and Evolution of Cooperative Breeding in Birds* (eds. W. D. Koenig & J. L. Dickinson), pp 48-66. Cambridge, UK: Cambridge University Press.

Dickinson J. L., Akçay C., Ferree E. D. & Stern C. A. (2016). A hierarchical analysis of incest avoidance in a cooperative breeder. *Behavioral Ecology,* **27**, 1132-1140.

Dickinson J. L., Koenig W. D. & Pitelka F. A. (1996). Fitness consequences of helping behaviour in the western bluebird. *Behavioral Ecology*, **7**, 168-177.

Double M. C., Peakall R., Beck N. R. & Cockburn A. (2005). Dispersal, philopatry and infidelity: dissecting local genetic structure in superb fairy-wrens *Malurus cyaneus*. *Evolution*, **59**, 625-635.

Dowling J. L., Colombelli-Négrel D. & Webster M. S. (2016). Kin Signatures Learned in the Egg? Red-Backed Fairy-Wren Songs Are Similar to Their Mother's In-Nest Calls and Songs. *Frontiers in Ecology and Evolution*, **4**, 48.

Downs S. G. & Ratnieks F. L. W. (2000). Adaptive shifts in honey bee (*Apis mellifera L.*) guarding behavior support predictions of the acceptance threshold model. *Behavioral Ecology*, **11**, 326-333.

Drobniak S. M., Wagner G., Mourocq E. & Griesser M. (2015). Family living: an overlooked but pivotal social system to understand the evolution of cooperative breeding. *Behavioral Ecology*, **26**, 805–811.

Duarte L. C., Bouteiller C., Fontanillas P., Petit E. & Perrin N. (2003). Inbreeding in the greater white-toothed shrew, *Crocidura russula*. *Evolution*, **57**, 638-645.

Duffy J. E. & Macdonald K. S. (2009). Kin structure, ecology and the evolution of social organization in shrimp: a comparative analysis. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 575–584.

Dunn P. O., Cockburn A. & Mulder R. A. (1995). Fairy-wren helpers often care for young to which they are unrelated. *Proceedings of the Royal Society B: Biological Sciences*, **259**, 339–343.

Duthie A. B. & Reid J. M. (2015). What happens after inbreeding avoidance? Inbreeding by rejected relatives and the inclusive fitness benefit of inbreeding avoidance. *PLoS ONE*, **10**, e0125140.

Duthie A. B. & Reid J. M. (2016). Evolution of inbreeding avoidance and inbreeding preferences through mate choice among interacting relatives. *The American Naturalist*, **188**, 651-667.

Edelaar P., Alonso D., Lagerveld S., Senar J. C., Bjorklund M. (2012). Population differentiation and restricted gene flow in Spanish crossbills: not isolation-by-distance but isolation-by-ecology. *Journal of Evolutionary Biology*, **25**, 417-430.

Ekman J., Dickinson J. L., Hatchwell B. J. & Griesser M. (2004). Delayed dispersal. In: *Ecology and Evolution of Cooperative breeding in birds* (eds. W. D. Koenig & J. L. Dickinson), pp. 35–47. Cambridge, UK: Cambridge University Press.

Emlen S. T. & Wrege P. H. (1988). The role of kinship in helping decisions among white-fronted bee-eaters. *Behavioral Ecology and Sociobiology*, **23**, 305-315.

Emlen S. T. & Wrege P. H. (1991). Breeding biology of white-fronted bee-eaters at Nakuru - the influence of helpers on breeder fitness. *Journal of Animal Ecology*, **60**, 309-326.

Emlen S. T. & Wrege P. H. (1992). Parent–offspring conflict and the recruitment of helpers among bee-eaters. *Nature*, **356**, 331–333.

Emlen S. T. (1982). The evolution of helping. I. An ecological constraints model. *The American Naturalist*, **119**, 29-39.

Emlen S. T. (1991). Evolution of cooperative breeding in birds and mammals. In: *Behavioural ecology: An evolutionary approach* (eds. J. R. Krebs & N. B. Davies), pp. 301-337. Oxford: Blackwell Science.

Emlen S. T. (1995). An evolutionary theory of the family. *Proceedings of the National Academy of Science*, **92**, 8092-8099.

Emlen S. T. (1997). Predicting family dynamics in social vertebrates. In: *Behavioural Ecology: An Evolutionary Approach* (eds. J. R Krebs & N. B. Davies), pp.228-253. Cambridge, UK: Blackwell Science Ltd.

Evans C. S., Evans L. & Marler, P. (1993). On the meaning of alarm calls: functional reference in an avian vocal system. *Animal Behaviour,* **46**, 23–38.

Fisher R. A. (1954). Evolution and bird sociality. In: *Evolution as a Process* (eds. J. Huxley, A. Hardy & E. Ford), pp. 71-83. London, UK: Allen & Unwin.

Frankham R. (1995). Inbreeding and extinction: a threshold effect. *Conservation Biology*, **9**, 792-799.

Frantz A. C., Pope L. C., Etherington T. R., Wilson G. J. & Burke T. (2010). Using isolation-by-distance-based approaches to assess the barrier effect of linear landscape elements on badger (*Meles meles*) dispersal. *Molecular Ecology*, **19**, 1663-1674.

Franz M. & Goller F. (2003). Respiratory patterns and oxygen consumption in singing zebra finches. *Journal of Experimental Biology*, **206**, 967–978.

Gamboa G. J., Grudzien T. A., Espelie K. E. & Bura E. A. (1996). Kin recognition pheromones in social wasps: Combining chemical and behavioural evidence. *Animal Behaviour*, **51**, 625-629.

Gamboa G. J., Reeve H. K. & Holmes W. G. (1991). Conceptual issues and methodology in kin-recognition research - a critical discussion. *Ethology*, **88**, 109-127.

Gamboa G. J., Reeve H. K., Ferguson I. D. & Wacker T. L. (1986). Nestmate recognition in social wasps - the origin and acquisition of recognition odours. *Animal Behaviour*, **34**, 685-695.

Gardner A. & West S. A. (2007). Social Evolution: The Decline and Fall of Genetic Kin Recognition. *Current Biology*, **17**, R810–R812.

Giret N., Roy P., Albert A., Pachet F., Kreutzer M. & Bovet D. (2011). Finding good acoustic features for parrot vocalizations: The feature generation approach. *The Journal of the Acoustical Society of America*, **129**, 1089-1099.

Godfrey S. S., Ansari T. H., Gardner M. G., Farine D. R. & Bull C. M. (2014). A contact-based social network of lizards is defined by low genetic relatedness among strongly connected individuals. *Animal Behaviour*, **97**, 35–43.

Goslee S. & Urban D. (2007). The ECODIST package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1-19. 10.18637/jss.v022.i07.

Grafen A. (1990). Do animals really recognize kin? *Animal Behaviour*, **39**, 42-54.

Green J. P., Holmes A. M., Davidson A. J., Paterson S., Stockley P., Beynon R. J., & Hurst J. L. (2015). The genetic basis of kin recognition in a cooperatively breeding mammal. *Current Biology*, **25**, 2631-2641.

Greenberg. L. (1979). Genetic component of bee odour in kin recognition. *Science*, **206**, 1095-1097.

Greenwood P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28**, 1140–62.

Griesser M. & Ekman J. (2005). Nepotistic mobbing behaviour in the Siberian jay, *Perisoreus infaustus*. *Animal Behaviour*, **69**, 345-352.

Griffin A. S. & West S. A. (2003). Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science*, **302**, 634-636.

Griffiths R., Double M. C., Orr K. & Dawson R. J. G (1998). A DNA test to sex most birds. *Molecular Ecology*, **7**, 1071-1075.

Halpin Z. T. (1991). Kin recognition cues of vertebrates. In: *Kin Recognition*. (ed. P. G. Hepper), pp.220-258. Cambridge, UK: Cambridge University Press.

Hamilton W. D. (1963). The evolution of altruistic behavior. *The American Naturalist*, **97**, 354-356.

Hamilton W. D. (1964). The genetical evolution of social behaviour (I and II). *Journal of Theoretical Biology*, **7**, 1-52.

Hamilton W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, **228**, 1218.

Hansson B. & Westerberg L. (2002). On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11**, 2467–2474.

Hardy O. J. & Vekemans X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hatchwell B. J. & Komdeur J. (2000). Ecological constraints, life history traits and the evolution of cooperative breeding. *Animal Behaviour*, **59**, 1079-1086.

Hatchwell B. J. (2009). The evolution of cooperative breeding in birds: kinship, dispersal and life history. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 3217–3227.

Hatchwell B. J. (2010). Cryptic kin selection: Kin structure in vertebrate populations and opportunities for kin-directed cooperation. *Ethology*, **116**, 203–216.

Hatchwell B. J. (2016). Long-tailed tits: ecological causes and fitness consequences of redirected helping. In: *Ecology and Evolution of Cooperative Breeding in Birds* (eds. W. D. Koenig & J. L. Dickinson), pp. 39-57 Cambridge, UK: Cambridge University Press.

Hatchwell B. J., Anderson C., Ross D. J., Fowlie M. K. & Blackwell P. G. (2001a). Social organization of cooperatively breeding long-tailed tits: kinship and spatial dynamics. *Journal of Animal Ecology*, **70**, 820-830.

Hatchwell B. J., Gullett P. R. & Adams M. J. (2014). Helping in cooperatively breeding long-tailed tits: a test of Hamilton's rule. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**, 20130565.

Hatchwell B. J., Ross D. J., Chaline N., Fowlie M. K., Burke T. (2002). Parentage in the cooperative breeding system of long-tailed tits, *Aegithalos caudatus*. *Animal Behaviour*, **64**, 55-63.

Hatchwell B. J., Ross D. J., Fowlie M. K. & McGowan A. (2001b). Kin discrimination in cooperatively breeding long-tailed tits. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 885-890.

Hatchwell B. J., Russell A. F., Fowlie M. K. & Ross D. J. (1999). Reproductive success and nest-site selection in a cooperative breeder: effect of experience and a direct benefit of helping. *The Auk*, **116**, 355-363.

Hatchwell B. J., Russell A. F., MacColl A. D. C., Ross D. J., Fowlie M. K. & McGowan A. (2004). Helpers increase long-term but not short-term productivity in cooperatively breeding long-tailed tits. *Behavioral Ecology*, **15**, 1–10.

Hatchwell B. J., Russell A. F., Ross D. J. & Fowlie M. K. (2000). Divorce in cooperatively breeding long-tailed tits: a consequence of inbreeding avoidance? *Proceedings of the Royal Society B: Biological Sciences*, **267**, 813-819.

Hatchwell B. J., Sharp S. P., Beckerman A. P. & Meade J. (2013). Ecological and demographic correlates of helping behaviour in a cooperatively breeding bird. *Journal of Animal Ecology*, **82**, 486-494.

Hauber M. E. & Sherman P. W. (2001). Self-referent phenotype matching: theoretical considerations and empirical evidence. *Trends in Neurosciences*, **24**, 609-616.

Heg D., Bachar Z., Brouwer L. & Taborsky M. (2004). Predation risk is an ecological constraint for helper dispersal in a cooperatively breeding cichlid. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2367–2374.

Hepper P. G. (1986). Kin recognition: functions and mechanisms a review. *Biological Reviews*, **61**, 63-93.

Hepper P. G. (1987). The amniotic fluid: an important priming role in kin recognition. *Animal Behaviour*, **35**, 1343–1346.

Hepper P. G. (1991). *Kin Recognition*. Cambridge, UK: Cambridge University Press.

Hijmans R. J. (2016). *Raster: Geographic Data Analysis and Modeling*. R package version 2.5-8. https://CRAN.R-project.org/package=raster.

Holman L., Van Zweden J. S., Linksvayer T. A. & d'Ettorre P. (2013). Crozier's paradox revisited: maintenance of genetic recognition systems by disassortative mating. *BMC Evolutionary Biology*, **13**, 211.

Holmes W. G. & Sherman P. W. (1982). The ontogeny of kin recognition in two species of ground squirrels. *American Zoologist*, **22**, 491-517.

Holmes W. G. & Sherman P. W. (1983). Kin recognition in animals. *American Scientist*, **71**, 46-55.

Holmes W. G. (1986). Identification of paternal half-siblings by captive Belding ground-squirrels. *Animal Behaviour*, **34**, 321-327.

Holmes W. G. (2004). The early history of Hamiltonian-based research on kin recognition. *Annales Zoologici Fennici*, **41**, 691-711.

Hoogland J. L. (1992). Levels of Inbreeding Among Prairie Dogs. *The American Naturalist*, **139**, 591-602.

Hopp S. L., Jablonski P. & Brown J. L. (2001). Recognition of group membership by voice in Mexican jays, *Aphelocoma ultramarina*. *Animal Behaviour*, **62**, 297–303.

Huisman J., Kruuk L. E. B., Ellis P. A., Clutton-Brock T. & Pemberton J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Science*, **113**, 3585–3590.

Ihle M. & Forstmeier W. (2013). Revisiting the evidence for inbreeding avoidance in zebra finches. *Behavioral Ecology*, **24**, 1356-1362.

Ingram C. M., Troendle N. J., Gill C. A., Braude S. & Honeycutt R. L. (2015). Challenging the inbreeding hypothesis in a eusocial mammal: population genetics of the naked mole-rat, Heterocephalus glaber. *Molecular Ecology*, **24**, 4848–486.

Jamieson I. G., Taylor S. S., Tracy L. N., Kokko H. & Armstrong D. P. (2009). Why some species of birds do not avoid inbreeding: insights from New Zealand robins and saddlebacks. *Behavioral Ecology*, **20**, 575–584.

Janik V. M. (1999). Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods. *Animal Behaviour*, **57**, 133-143.

Jetz W. & Rubenstein D. R. (2011). Environmental uncertainty and the global biogeography of cooperative breeding in birds. *Current Biology*, **21**, 72-78.

Johnson M. L. & Gaines M. S. (1990). Evolution of Dispersal: Theoretical Models and Empirical Tests Using Birds and Mammals. *Annual Review of Ecology and Systematics*, **21**, 449–480.

Kalinowski S. T., Taper M. L. & Marshall T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099-1106.

Keen S. C., Meliza C. D. & Rubenstein D. R. (2013). Flight calls signal group and individual identity but not kinship in a cooperatively breeding bird. *Behavioral Ecology*, **24**, 1279–1285.

Keller L. & Ross K. G. (1998). Selfish genes: a green beard in the red fire ant. *Nature*, **394**, 573-575.

Keller L. F. & Waller D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, **17**, 230-241.

Kempenaers B. & Sheldon B. C. (1996). Why do male birds not discriminate between their own and extra-pair offspring? *Animal Behaviour*, **51**, 1165-1173.

Kennedy E. S., Grueber C. E., Duncan R. P. & Jamieson I. G. (2014). Severe inbreeding depression and no evidence of purging in an extremely inbred wild species- the Chatham island black robin. *Evolution*, **68**, 987–995.

Klump G. M. (1996). Bird communication in the noisy world. In: *Ecology and evolution of acoustic communication in birds* (eds. D. E. Kroodsma & E. H. Miller), pp. 321-338. Ithaca: Cornell University Press.

Koenig W. D. & Dickinson J. L. (2004). *Ecology and Evolution of Cooperative Breeding in Birds*. Cambridge, UK: Cambridge University Press.

Koenig W. D. & Dickinson J. L. (2016). *Cooperative Breeding in Vertebrates*. Cambridge, UK: Cambridge University Press.

Koenig W. D. & Haydock J. L. (2004). Incest and incest avoidance. In: *Ecology and Evolution of Cooperative Breeding in Birds* (eds. W. D. Koenig & J. L. Dickinson), pp. 142-156. Cambridge, UK: Cambridge University Press.

Koenig W. D., Haydock, J. & Stanback M. T. (1998). Reproductive Roles in the Cooperatively Breeding Acorn Woodpecker: Incest Avoidance versus Reproductive Competition. *The American Naturalist*, **151**, 243-255.

Koenig W. D., Pitelka F. A., Carmen W. J., Mumme R. L. & Stanback M. T. (1992). The evolution of delayed dispersal in cooperative breeders. *Quarterly Review of Biology*, **67**, 111–150.

Koenig W. D., Stanback M. T. & Haydock J. (1999). Demographic consequences of incest avoidance in the cooperatively breeding acorn woodpecker, *Animal Behaviour*, **57**, 1287-1293.

Kokko H. & Ekman J. (2002). Delayed Dispersal as a Route to Breeding: Territorial Inheritance, Safe Havens, and Ecological Constraints. *The American Naturalist*, **160**, 468–484.

Kokko H. & Ots I. (2006). When Not To Avoid Inbreeding. *Evolution*, **60**, 467-475.

Kokko H., Johnstone R. A. & Clutton-Brock T. H. (2001). The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 187–196.

Kokko H., Johnstone R. A. & Wright J. (2002). The evolution of parental and alloparental effort in cooperatively breeding groups: when should helpers pay to stay? *Behavioral Ecology*, **13**, 291–300.

Komdeur J. & Hatchwell B. J. (1999). Kin recognition: function and mechanism in avian societies. *Trends in Ecology & Evolution*, **14**, 237-241.

Komdeur J. (1994). The effect of kinship on helping in the cooperative breeding Seychelles warbler (*Acrocephalus-sechellensis*). *Proceedings of the Royal Society B: Biological Sciences*, **256**, 47-52.

Komdeur J., Richardson D. S. & Burke T. (2004). Experimental evidence that kin discrimination in the Seychelles warbler is based on association and not on genetic relatedness. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 963-969.

Konovalov D. A., Manning C. & Henshaw M. T. (2004). KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology*, **4**, 779-782.

Krause E. T., Kruger O., Kohlmeier P. & Caspers B. A. (2012). Olfactory kin recognition in a songbird. *Biology Letters*, **8**, 327-329.

Krebs J. R. & Davies N. B. (1997). The evolution of behavioural ecology. In: *Behavioural ecology: An evolutionary approach* (eds. J. R. Krebs & N. B. Davies), pp. 3-12. Oxford: Blackwell Science.

Kruuk L. E. B., Sheldon B. C. & Merila J. (2002). Severe inbreeding depression in collared flycatchers (Ficedula albicollis). *Proceedings of the Royal Society B: Biological Sciences*, **269**, 1581-1589.

Kurvers R. H. J. M., Adamczyk V. M. A. P., Kraus R. H. S., Hoffman J. I., van Wieren S. E., van der Jeugd, H. P., Amos W., Prins H. H. T., Herbert H. T. & Jonker R. M. (2013). Contrasting context dependence of familiarity and kinship in animal social networks. *Animal Behaviour*, **86**, 993–1001.

Kuznetsova A., Brockhoff P B. & Christensen R. H. B. (2017). *lmerTest* Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**, 1–26.

Lacey E. A. & Sherman P. W. (1991). Social organization of naked mole-rat colonies: evidence for divisions of labor. In: *The Biology of the Naked Mole-Rat* (eds. P. W.

Sherman P. W., J. U. M. Jarvis & R. D. Alexander), pp. 275–336. New Jersey, USA: Princeton University Press.

Lachlan R. F. (2007). *Luscinia: a bioacoustics analysis computer program* (v2.16.10.29.01). Retrieved from http://luscinia.sourceforge.net/.

Lacy R. C. & Sherman P. W. (1983). Kin recognition by phenotypic matching. *The American Naturalist*, **121**, 489-512.

Leclaire S., Van Dongen W. F., Voccia S., Merkling T., Ducamp C., Hatch S. A., Blanchard P., Danchin E. & Wagner R. H. (2014). Preen secretions encode information on MHC similarity in certain sex-dyads in a monogamous seabird. *Scientific reports*, **4**, 6920.

Lee J-W., Simeoni M., Burke T. & Hatchwell B. J. (2010). The consequences of winter flock demography for genetic structure and inbreeding risk in vinous-throated parrotbills, *Paradoxornis webbianus*. *Heredity*, **104**, 472-481.

Leedale A. E., Sharp S. P., Simeoni M., Robinson E. J. H. & Hatchwell B. J. (2018). Fine-scale genetic structure and helping decisions in a cooperatively breeding bird. *Molecular Ecology*, **27**, 1714–1726.

Lehmann L. & Perrin N. (2003). Inbreeding Avoidance through Kin Recognition: Choosy Females Boost Male Dispersal. *The American Naturalist*, **162**, 638–652.

Lehmann L. & Balloux F. (2007). Natural selection on fecundity variance in subdivided populations: Kin selection meets bet hedging. *Genetics*, **176**, 361-377.

Ligon J. D. & Burt D. B. (2004). Evolutionary origins. In: *Ecology and Evolution of Cooperative Breeding in Birds* (eds. W. D. Koenig & J. L. Dickinson), pp 5-34. Cambridge, UK: Cambridge University Press.

Ligon J. D. & Ligon S. H. (1978). Communal breeding in green woodhoopoes as a case for reciprocity. *Nature*, **276**, 496-498.

Ligon J. D. & Stacey P. B. (1991) The origin and maintenance of helping in birds. *The American Naturalist*, **138**, 254–258.

Lukas D. & Clutton-Brock T. H. (2017). Climate and the distribution of cooperative breeding in mammals. *Royal Society Open Science*, **4**, 160897.

Lukas D. & Clutton-Brock T. H. (2011). Group structure, kinship, inbreeding risk and habitual female dispersal in plural-breeding mammals. *Journal of Evolutionary Biology*, **24**, 2624-2630.

Lukas D. & Clutton-Brock T. H. (2012). Life histories and the evolution of cooperative breeding in mammals. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 4065-4070.

Lynch A., Plunkett G. M., Baker A. J. & Jenkins P. F. (1989). A model of cultural evolution of chaffinch song derived with the meme concept. *The American Naturalist*, **133**, 634-653.

Lynch M. (1991). The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution*, **45**, 622-629.

MacColl A. D. C. & Hatchwell B. J. (2004). Determinants of lifetime fitness in a cooperative breeder, the long-tailed tit *Aegithalos caudatus*. *Journal of Animal Ecology*, **73**, 1137–1148.

Madden J. R., Nielsen J. F. & Clutton-Brock T. H. (2012). Do networks of social interactions reflect patterns of kinship? *Current Zoology*, **58**, 319–328.

Marler P. (2004). Science and birdsong: the good old days. In: *Nature's Music: The Science of Birdsong* (eds. P. Marler & H. Slabbekoorn), pp.1-38. London, UK: Elsevier Academic Press.

Martín-Vivaldi M., Martínez J. J., Palomino J. & Soler M. (2002). Extrapair paternity in the Hoopoe *Upupa epops*: An exploration of the influence of interactions between breeding pairs, non-pair males and strophe length. *Ibis*, **144**, 236 - 247.

Mateo J. M. & Johnston R. E. (2000). Kin Recognition and the 'Armpit Effect': Evidence of Self-Referent Phenotype Matching. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 695-700.

Mateo J. M. (2004). Recognition systems and biological organization: The perception component of social recognition. *Annales Zoologici Fennici*, **41**, 729-745.

Maynard Smith J. (1964). Group selection and kin selection. *Nature*, **201**, 1145-7.

McDonald P. G., Rollins L. A., & Godfrey S. (2015). The relative importance of spatial proximity, kin selection and potential “greenbeard” signals on provisioning behaviour among helpers in a cooperative bird. *Behavioral Ecology and Sociobiology*, **70**, 133–143.

McDonald P. G. & Wright J. (2011). Bell miner provisioning calls are more similar among relatives and are used by helpers at the nest to bias their effort towards kin. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 3403-3411.

McGowan A., Fowlie M. K., Ross D. J. & Hatchwell B. J. (2007). Social organization of co-operatively breeding Long-tailed Tits *Aegithalos caudatus*: flock composition and kinship. *Ibis,* **149**, 170-174.

McGowan A., Hatchwell B. J. & Woodburn R. J. W. (2003). The effect of helping behaviour on the survival of juvenile and adult long-tailed tits *Aegithalos caudatus*. *Journal of Animal Ecology,* **72**, 491-499.

Magrath R. D., & Whittingham L. A. (1997). Subordinate males are more likely to help if unrelated to the breeding female in cooperatively breeding white-browed scrubwrens. *Behavioral Ecology and Sociobiology*, **41**, 185–192.

Mcgregor P. K. & Krebs J. R. (1982). Song Types in a Population of Great Tits (Parus Major): Their Distribution, Abundance and Acquisition By Individuals. *Behaviour*, **79**, 126–152.

Meade J. & Hatchwell B. J. (2010). No direct fitness benefits of helping in a cooperative breeder despite higher survival of helpers. *Behavioral Ecology*, **21**, 1186-1194.

Meade J., Nam K.-B., Beckerman A. P. & Hatchwell B. J. (2010). Consequences of ‘load-lightening’ for future indirect fitness gains by helpers in a cooperatively breeding bird. *Journal of Animal Ecology*, **79**, 529-537.

Meliza C. D., Keen S. C., & Rubenstein D. R. (2013). Pitch-and spectral-based dynamic time warping methods for comparing field recordings of harmonic avian vocalizations. *The Journal of the Acoustical Society of America*, **134**, 1407-1415.

Mulder R. A., Dunn P. O., Cockburn A., Lazenby-Cohen K. A. & Howell M. J. (1994). Helpers liberate female fairy-wrens from constraints on extra-pair mate choice.

Proceedings of the Royal Society of London Series B: Biological Sciences, **255**, 223-229.

Nakagawa S. & Schielzeth H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, **85**, 935-956.

Nam K.-B., Simeoni M., Sharp S. P. & Hatchwell B. J. (2010). Kinship affects investment by helpers in a cooperatively breeding bird. *Proceedings of the Royal Society B: Biological Sciences,* **277**, 3299-3306.

Napper C. J. & Hatchwell B. J. (2016). Social dynamics in nonbreeding flocks of a cooperatively breeding bird: causes and consequences of kin associations. *Animal Behaviour*, **122**, 23-35.

Napper C. J., Sharp S. P., McGowan A., Simeoni S. & Hatchwell B. J. (2013). Dominance, not kinship, determines individual position within the communal roosts of a cooperatively breeding bird. *Behavioral Ecology and Sociobiology*, **67**, 2029-2039.

Nichols H. J. (2017). The causes and consequences of inbreeding avoidance and tolerance in cooperatively breeding vertebrates. *Journal of Zoology*, **303**, 1–14.

Nielsen R., Tarpy D. R. & Reeve H. K. (2003). Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology*, **12**, 3157-3164.

Nowicki S. (1989). Vocal plasticity in captive black-capped chickadees: the acoustic basis and rate of call convergence. *Animal Behaviour*, **37**, 64–73.

O'Riain M. J., Bennett N. C., Brotherton P. N. M., McIlrath G. & Clutton-Brock T. H. (2000). Reproductive suppression and inbreeding avoidance in wild populations of cooperatively breeding meerkats (*Suricata suricatta*). *Behavioral Ecology and Sociobiology*, **48**, 471–477.

Painter J. N., Crozier R. H., Poiani A., Robertson R. J. & Clarke M. F. (2000). Complex social organization reflects genetic structure and relatedness in the cooperatively breeding bell miner, *Manorina melanophrys*. *Molecular Ecology*, **9**, 1339-1347.

Parker G. A. (1979). Sexual selection and sexual conflict. In: *Sexual selection and reproductive competition in insects* (eds. M.S. Blum & N. A. Blum), pp. 123-166. New York: Academic Press.

Parker G. A. (2006). Sexual conflict over mating and fertilization: an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 235-259.

Payne R. B., Payne L. L. & Rowley I. (1988). Kin and social relationships in splendid fairy-wrens - recognition by song in a cooperative bird. *Animal Behaviour*, **36**, 1341-1351.

Pemberton J. (2004). Measuring inbreeding depression in the wild: the old ways are the best. *Trends in Ecology & Evolution*, **19**, 613–615.

Perrin N. & Mazalov V. (2000). Local competition, inbreeding, and the evolution of sex-biased dispersal. *The American Naturalist*, **155**, 116-127.

Pfefferl D., Ruiz-Lambides A. V. & Widdig A. (2013). Female rhesus macaques discriminate unfamiliar paternal sisters in playback experiments: support for acoustic phenotype matching. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20131628.

Pfennig D. W., Reeve H. K. & Sherman P. W. (1993). Kin recognition and cannibalism in spadefoot toad tadpoles. *Animal Behaviour*, **46**, 87-94.

Pizzari T., Lo H. & Cornwallis C. K. (2004). Sex-specific, counteracting responses to inbreeding in a bird. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2115–2121.

Preston S. A. J., Briskie J. V. & Hatchwell B. J. (2013). Adult helpers increase the recruitment of closely related offspring in the cooperatively breeding rifleman. *Behavioral Ecology*, **27**, 1617–1626.

Price J. J. (1998). Family- and sex-specific vocal traditions in a cooperatively breeding songbird. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 497-502.

Price J. J. (1999). Recognition of family-specific calls in stripe-backed wrens. *Animal Behaviour*, **57**, 483-492.

Pusey A. E. & Wolf M. (1996). Inbreeding avoidance in animals. *Trends in Ecology & Evolution*, **11**, 201-206.

Pusey A. E. (1987). Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution*, **2**, 295-299.

Puurtinen M. (2011). Mate choice for optimal (k)inbreeding. *Evolution*, **65**, 1501–1505.

Queller D. C. (1994). Genetic relatedness in viscous populations. *Evolutionary Ecology*, **8**, 70-73.

Queller D. C. & Goodnight K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, **43**, 258-275.

Queller D. C., Ponte E., Bozzaro S. & Strassmann J. E. (2003). Single-Gene Greenbeard Effects in the Social Amoeba *Dictyostelium discoideum*. *Science,* **299**, 105-106.

Queller D. C., Strassmann J. E., Solís C. R., Hughes C. R. & DeLoach D. M. (1993). A selfish strategy of social insect workers that promotes social cohesion. *Nature*, **365**, 639–641.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radford A. N. (2005). Group-specific vocal signatures and neighbour-stranger discrimination in the cooperatively breeding green woodhoopoe. *Animal Behaviour*, **70**, 1227-1234.

Reeve H. K. (1989). The evolution of conspecific acceptance thresholds. *The American Naturalist*, **133**, 407-435.

Reeve H. K., Westneat D. F., Noon W. A., Sherman P. W. & Aquadro C. F. (1990). DNA “fingerprinting” reveals high levels of inbreeding in colonies of the eusocial naked mole-rat. *Proceedings of the National Academy of Sciences*, **87**, 2496–2500.

Reid J. M., Arcese P., Cassidy A. L. E. V, Hiebert S. M., Smith J. N. M., Stoddard P. K., Marr A. B & Keller L. F. (2005). Fitness correlates of song repertoire size in free-living song sparrows (*Melospiza melodia*). *The American Naturalist*, **165**, 299–310.

Reyer H. U. (1984). Investment and relatedness: A cost/benefit analysis of breeding and helping in the pied kingfisher. *Animal Behaviour*, **32**, 1163-1178.

Reid J. M., Arcese P., Keller L. F., Elliott K. H., Sampson L. & Hasselquist D. (2007). Inbreeding effects on immune response in free-living song sparrows (*Melospiza melodia*). *Proceedings of the Royal Society B: Biological Sciences*, **274**, 697–706.

Richardson D. S., Burke T. & Komdeur J. (2003). Sex-specific associative learning cues and inclusive fitness benefits in the Seychelles warbler. *Journal of Evolutionary Biology*, **16**, 854-861.

Riede T., Suthers R. A., Fletcher N. H. & Blevins W. E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proceedings of the National Academy of Sciences*, **103**, 5543-5548.

Riehl C. (2013). Evolutionary routes to non-kin cooperative breeding in birds. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20132245–20132245.

Riehl C. & Stern C. A. (2015). How cooperatively breeding birds identify relatives and avoid incest: New insights into dispersal and kin recognition. *BioEssays*, **37**, 1303–1308.

Roff D. A. (2002). Inbreeding depression: tests of the overdominance and partial dominance hypotheses. *Evolution*, **56**, 768-775.

Rolshausen G., Segelbacher G., Hermes C., Hobson K. A. & Schaefer H. M. (2013). Individual differences in migratory behavior shape population genetic structure and microhabitat choice in sympatric blackcaps (*Sylvia atricapilla*). *Ecology and Evolution*, **3**, 4278-4289.

Rousset F. & Roze D. (2007). Constraints on the origin and maintenance of genetic kin recognition. Evolution. *International Journal of Organic Evolution*, **61**, 2320-2330.

Rowley I., Russell E. & Brooker M. (1986). Inbreeding: Benefits may outweigh costs. *Animal Behaviour*, **34**, 939–941.

Rubenstein D. R. & Abbot P. (2017). *Comparative Social Evolution*. Cambridge, UK: Cambridge University Press.

Russell A. F. & Hatchwell B. J. (2001). Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 2169-2174.

Russell A. F., Portelli D. J., Russell D. J. F. & Barclay H. (2010). Breeding ecology of the Chestnut-crowned Babbler: a cooperative breeder in the desert. *Emu*, **110**, 324–331.

Saccheri I., Kuussaari M., Kankare M., Vikman P., Fortelius W. & Hanski I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491-494.

Schielzeth H., Nakagawa S. & Stoffel M. (2011). *rptR: Repeatability for Gaussian and non-Gaussian data*. R package version 0.6.405.

Schnell G. D., Watt D. J. & Douglas M. E. (1985). Statistical comparison of proximity matrices: applications in animal behaviour. *Animal Behaviour*, **33**, 239–253.

Searcy W. A., Nowicki S. & Hughes M. (1997). The Response of Male and Female Song Sparrows to Geographic Variation in Song. *The Condor*, **99**, 651–657.

Sharp S. P. & Hatchwell B. J. (2005). Individuality in the contact calls of cooperatively breeding long-tailed tits (*Aegithalos caudatus*). *Behaviour*, **142**, 1559-1575.

Sharp S. P. & Hatchwell B. J. (2006). Development of family specific contact calls in the long-tailed tit *Aegithalos caudatus*. *Ibis*, **148**, 649-656.

Sharp S. P., Baker M. B., Hadfield J. D., Simeoni M. & Hatchwell B. J. (2008a). Natal dispersal and recruitment in a cooperatively breeding bird. *Oikos*, **117**, 1371-1379.

Sharp S. P., McGowan A., Wood M. J. & Hatchwell B. J. (2005). Learned kin recognition cues in a social bird. *Nature*, **434**, 1127-1130.

Sharp S. P., Simeoni M. & Hatchwell B. J. (2008b). Dispersal of sibling coalitions promotes helping among immigrants in a cooperatively breeding bird. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 2125-2130.

Sharp S. P., Simeoni M., McGowan A., Nam K.-B & Hatchwell B. J. (2011). Patterns of recruitment, relatedness and cooperative breeding in two populations of long-tailed tits. *Animal Behaviour*, **81**, 843-849.

Shen S.-F., Emlen S. T., Koenig W. D. & Rubenstein D. R. (2017). The ecology of cooperative breeding behaviour. *Ecology Letters,* **20**, 708–720.

Shepher J. (1971). Mate selection among second generation kibbutz adolescents and adults: Incest avoidance and negative imprinting. *Archives of Sexual Behavior*, **1**, 293–307.

Sherman P. W., Reeve H. K. & Pfennig D. W. (1997). Recognition Systems. In: *Behavioral Ecology: An Evolutionary Approach* (eds. J. R Krebs & N. B. Davies), pp. 69-96. Cambridge, UK: Blackwell Science Ltd.

Simeoni M., Dawson D. A., Ross D. J., Châline N., Burke T. & Hatchwell B. J. (2007) Characterization of 20 microsatellite loci in the long-tailed tit Aegithalos caudatus (Aegithalidae, AVES). *Molecular Ecology Notes*, **7**, 1319-1322.

Slate J., David P., Dodds K. G., Veenvliet B. A., Glass B. C., Broad T. E. & McEwan J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity*, **93**, 255-265.

Slate J., Kruuk L. E. B., Marshall T. C., Pemberton J. M. & Clutton-Brock T. H. (2000). Inbreeding depression influences lifetime breeding success in a wild population of red deer (*Cervus elaphus*). *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1657–1662.

Slatkin M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457-462.

Smallbone W., van Oosterhout C. & Cable J. (2016). The effects of inbreeding on disease susceptibility: *Gyrodactylus turnbulli* infection of guppies, *Poecilia reticulata*. *Experimental Parasitology,* **167**, 32–37.

Soler M., Soler J. J., Martinez J. G. & Møller A. P. (1995). Chick Recognition and Acceptance: A Weakness in Magpies Exploited by the Parasitic Great Spotted Cuckoo. *Behavioral Ecology and Sociobiology*, **37**, 243-248.

Stacey P. & Ligon J. (1987). Territory Quality and Dispersal Options in the Acorn Woodpecker, and a Challenge to the Habitat-Saturation Model of Cooperative Breeding. *The American Naturalist*, **130**, 654-676.

Stoffel M. A., Esser M., Kardos M., Humble E., Nichols H., David P. & Hoffman J. I. (2016). inbreedR: An R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, **7**, 1331–1339.

Suzuki T. N. (2013). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, **87**, 59–65.

Szulkin M., Bierne N. & David P. (2010). Heterozygosity-fitness correlations: a time for reappraisal. *Evolution*, **64**, 1202-1217.

Szulkin M., Garant D., McCleery R. H. & Sheldon, B. C. (2007). Inbreeding depression along a life-history continuum in the great tit. *Journal of Evolutionary Biology*, **20**, 1531–1543.

Szulkin M., Stopher K. V. Pemberton J. M. & Reid J. M. (2013). Inbreeding avoidance, tolerance, or preference in animals? *Trends in Ecology and Evolution*, **28**, 205–211.

Taborsky M. & Limberger D. (1981). Helpers in fish. *Behavioral Ecology and Sociobiology*, **8**, 143-145.

Tang-Martinez Z. (2001). The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: a critical re-evaluation. *Behavioural Processes*, **53**, 21-40.

Tarvin K., Webster M., Tuttle E. & Pruett-Jones S. (2005). Genetic similarity predicts the level of extra-pair paternity in splendid fairy-wrens. *Animal Behaviour*, **70**, 945-955.

Taylor P. D. (1992). Altruism in viscous populations: an inclusive fitness model. *Evolutionary Ecology*, **6**, 352-356.

Temple H. J., Hoffman J. I. & Amos, W. (2006) Dispersal, philopatry and intergroup relatedness: fine-scale genetic structure in the white-breasted thrasher *Ramphocinclus rachyurus*. *Molecular Ecology*, **15**, 3449-3458.

Thomsen H. M., Balsby T. J. & Dabelsteen T. (2013). Individual variation in the contact calls of the monomorphic Peach-fronted Conure, *Aratinga aurea*, and its potential role in communication. *Bioacoustics*, **22**, 215-227.

Townsend A. K., Clark A. B., McGowan K. J., Buckles E. L., Miller A. D. & Lovette I. J. (2009). Disease-mediated inbreeding depression in a large, open population of cooperative crows. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2057–2064.

Trivers R. (1972). Parental investment and sexual selection. In: *Sexual Selection and the Descent of Man 1871–1971* (ed. B. Campbell), pp. 139–179. Aldine Press, Chicago.

van Dijk R. E., Covas R., Doutrelant C., Spottiswoode C. N. & Hatchwell B. J. (2015). Fine-scale genetic structure reflects sex-specific dispersal strategies in a population of sociable weavers (*Philetairus socius*). *Molecular Ecology*, **24**, 4296-4311.

van Dijk R. E., Kaden J. C., Argüelles-Ticó A, Dawson, D. A., Burke, T. & Hatchwell, B. J. (2014). Cooperative investment in public goods is kin directed in communal nests of social birds. *Ecology Letters*, **17**, 1141-1148.

van Treuren R., Bijlsma R., Tinbergen J. M., Heg D. & van de Zande L. (1999). Genetic analysis of the population structure of socially organized oystercatchers (*Haematopus ostralegus*) using microsatellites. *Molecular Ecology*, **8**, 181–187.

Varian-Ramos C. W. & Webster M. S. (2012). Extrapair copulations reduce inbreeding for female red-backed fairy-wrens, *Malurus melanocephalus*. *Animal Behaviour*, **83**, 857-864.

Walling C. A., Nussey D. H., Morris A., Clutton-Brock T. H, Kruuk L. E. B. & Pemberton J. M. (2011). Inbreeding depression in red deer calves. *BMC Evolutionary Biology*, **11**, 318.

Walters J. R., Cooper C. B., Daniels S. J., Pasinelli G. & Schiegg K. (2004). Conservation biology. In: *Ecology and Evolution of Cooperative Breeding in Birds* (eds. W. D. Koenig & J. L. Dickinson), pp. 197–209. Cambridge, UK: Cambridge University Press.

Walters J. R., Doerr P. D. & Carter J. H. (2010). The Cooperative Breeding System of the Red-cockaded Woodpecker. *Ethology*, **78**, 275–305.

Wang C. & Lu X. I. N. (2011). Female ground tits prefer relatives as extra-pair partners: driven by kin-selection? *Molecular Ecology*, **20**, 2851-2863.

Wanker R. & Fischer J. (2001). Intra-and interindividual variation in the contact calls of spectacled parrotlets (*Forpus conspicillatus*). *Behaviour*, **138**, 709-726.

Ward S., Lampe H. M. & Slater P. J. B. (2004). Singing is not energetically demanding for pied flycatchers, *Ficedula hypoleuca*. *Behavioral Ecology*, **15**, 477–484.

Waser P., Austad S. & Keane B. (1986). When Should Animals Tolerate Inbreeding? *The American Naturalist*, **128**, 529-537.

Watts P. C., Rouquette J. R., Saccheri I. J., Kemp S. J. & Thompson D. J. (2004). Molecular and ecological evidence for small-scale isolation by distance in an endangered damselfly, *Coenagrion mercuriale*. *Molecular Ecology*, **13**, 2931-2945.

Weir B. S. and Cockerham C. C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

West S. A., Griffin A. S. & Gardner A. (2007a). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, **20**, 415–432.

West S. A., Griffin A. S. & Gardner A. (2007b). Evolutionary explanations for cooperation. *Current Biology*, **17**, R661-R672.

West S. A., Pen I. & Griffin A. S. (2002). Cooperation and competition between relatives. *Science*, **296**, 72-75.

Whittingham L. A., Dunn P. O. & Magrath R. D. (1997). Relatedness, polyandry and extra-group paternity in the cooperatively-breeding white-browed scrubwren (Sericornis frontalis). *Behavioral Ecology and Sociobiology*, **40**, 261-270.

Wilson E.O. (1975). *Sociobiology: The new synthesis*. Massachusetts, USA: Harvard University Press.

Wilson E.O. (1971). The Insect Societies. Cambridge, USA: Harvard University Press.

Winker K., McCracken K. G., Gibson D. D. & Peters J. L. (2013). Heteropatric speciation in a duck, *Anas crecca*. *Molecular Ecology*, **22**, 5922-5935.

Wolak M. E. (2012). nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution*, **3**, 792-796.

Wolfenden A., Jones C. G., Tatayah V., Züel N. & de Kort S. R. (2015). Endangered pink pigeons treat calls of the ubiquitous Madagascan turtle dove as conspecific. *Animal Behaviour*, **99**, 83-88.

Wolfenden G. E. (1975). Florida scrub jay helpers at the nest. *Auk*, **92**, 1-15.

Woxvold I. A., Adcock G. J. & Mulder R. A. (2006). Fine-scale genetic structure and dispersal in cooperatively breeding apostlebirds. *Molecular Ecology*, **15**, 3139-3146.

Wright J., McDonald P. G., te Marvelde L., Kazem A. J. N. & Bishop C. M. (2010). Helping effort increases with relatedness in bell miners, but 'unrelated' helpers of both sexes still provide substantial care. *Proceedings of the Royal Society B: Biological Sciences*, **227**, 437-445.

Wright J. (1998). Helping-at-the-nest and group size in the Arabian Babbler *Turdoides squamiceps*. *Journal of Avian Biology*, **29**, 105–112.

Wright J., Parker P. G. & Lundy K. J. (1999). Relatedness and chick-feeding effort in the cooperatively breeding Arabian babbler. *Animal Behaviour*, **58**, 779–785.

Wright L. I., Tregenza T. & Hosken D. J. (2008). Inbreeding, inbreeding depression and extinction. *Conservation Genetics*, **9**, 833-843.

Wright S. (1922). The Effects of Inbreeding and Crossbreeding on Guinea Pigs: I. Decline in Vigor: II. Differentiation Among Inbred Families. *USDA Bull*, **1121**, 1-60.

Wright S. (1943). Isolation by distance. *Genetics*, **28**, 114-138.

Wright S. (1949). Genetical structure of populations. *Nature*, **166**, 247–249.

Wright S. (1965). The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution*, **19**, 395–420.

Wright T. F. (1996). Regional Dialects in the Contact Call of a Parrot. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 867–872.

Yamazaki K., Boyse E.A., Mike V., Thaler H.T., Mathieson B.J., Abbott J., Boyse J. & Zayas Z.A. (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. *Journal of Experimental Medicine*, **144**, 1324-1335.

Yasukawa K., Urish J., Her A. & Light E. (2008). Similarity in the begging calls of nestling Red-winged Blackbirds. *Journal of Field Ornithology*, **79**, 254–262.

Zahavi A. (1995). Altruism as a handicap - the limitations of kin selection and reciprocity. *Journal of Avian Biology*, **26**, 1-3.

Zhang J. (2016). *spaa: SPecies Association Analysis*. R package version 0.2.1, URL https://CRAN.R-project.org/package=spaa.