Investigating the Value of Incorporating Behavioural Measures in a Discriminant Function Developed for Sex Assignment in Black-capped Chickadees (*Poecile atricapillus*)

by

Sheeraja Sridharan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Ecology

Department of Biological Sciences
University of Alberta

Abstract

Black-capped chickadees (*Poecile atricapillus*) are a sexually monomorphic species that typically require molecular assays or observation of sex-specific behaviours in the breeding season for accurate sex assignment. We developed a discriminant function in a central Albertan chickadee population using 469 individuals (238 females and 231 males) for the purpose of future sex assignment in the same population. We used morphometric measurements in the development of our model and investigated the potential value of incorporating behavioural measurements to improve correct sex assignment rates in the discriminant function. In total, we evaluated the utility of five measures of morphological traits (i.e., body mass, wing length, tarsus length, bill length and bill depth) and four measures of behavioural traits (exploration, handling aggression, daily foraging rate and foraging inter-visit interval (IVI) for developing a discriminant function. We compared a model consisting of all morphological traits with a model consisting only of highly repeatable morphological traits (i.e., body mass, wing length, tarsus length) and found they yielded comparable accuracy rates of 86% (84.8% Males, 87.2% Females) and 89.2% (91.3% males and 87.2% females), respectively. This demonstrates that at large enough sample sizes, inclusion of additional traits does not result in considerable increase in model accuracy and in fact, may run the risk of introducing error into model as a consequence of measurement errors. The model using only repeatable morphological traits also revealed clear cut off points in the discriminant score to assign sex to males (discriminant scores >81) and females (discriminant scores <77). However, 31.1% of individuals within our population fall within the

intermediate score range and consequently will require alternative sex confirmation techniques. We also ran a third model on a subset of our study population (155 individuals: 82 males, 73 females) using all morphological traits and the only behavioural trait (foraging IVI) for which significant sex differences were found, to evaluate the utility of using behavioural traits to improve sex assignment rates in discriminant functions. This model produced an overall accuracy of 90% (93.8% males, 85.7% females) with similarly high model accuracies obtained when compared to alternative models with i) all the morphological traits and no behavioural traits, and ii) only highly repeatable morphological traits and no behavioural traits. In our study, the measured behavioural traits did not improve the accuracy of the discriminant function. We discuss reasons why this might be the case and propose other behaviours that might be useful to consider for future work. Ultimately, in this study we were able to exhibit the value of capitalising on simple, easy to measure morphometric traits to accurately assign sex in species with cryptic morphological differences between sexes.

Preface

This thesis is an original work by Sheeraja Sridharan. No part of this thesis has been previously published. The data collected part of the thesis research received animal care approval from the University of Alberta Animal Care Committee under permit numbers AUP00002210 and AUP00002542.

Acknowledgements

I would first like to thank my supervisor Dr. Kimberley Mathot for taking me on as Master's student and for providing me with endless support and guidance throughout my project. I would also like to express my gratitude to my committee members Dr. David Coltman and Dr. Corey Davis for their invaluable advice and feedback through regular committee meetings that have been paramount to shaping my project, particularly in the molecular aspects of the project. I am also grateful to Dr. Coltman for being gracious enough to share his lab space and other lab resources to enable me to complete the molecular aspects of this project. Much appreciation also goes to the Coltman lab technician, Anh Dao, for her help in setting me up within the Coltman lab space, as well as for her guidance on how to execute different molecular techniques. Similarly, I would also like to thank MBSU technician Sophie Dang for loaning equipment and for her advice with regards to conducting lab work. I would also like to thank Dr. Chris Sturdy and Dr. Alison Murray for kindly agreeing to take on the roles of external examiner and examining chair in my committee, respectively. I am grateful to the field technician of the Mathot lab, Jan Wijmenga, for his extensive involvement in the fieldwork components of the black-capped chickadee foraging ecology project, especially with regards to the collection of all the morphological data that was critical to the successful execution of this project. I would also like to acknowledge the other Mathot lab members and students, both former and present, that have contributed towards the fieldwork or analysis of the data from the black-capped chickadee foraging ecology project, as their work has directly or indirectly contributed towards the data used in this project. I would like to specifically recognise both Josue Arteaga-Torres and Elène Haave Audet for their contributions towards the fieldwork in the black-capped chickadee foraging ecology project, and Elène in particular, for her role in scoring cage test videos, which generated data that was subsequently used in this project. Special thanks to all Mathot lab members for their support, encouragement, and friendship, throughout the course of my Masters. Finally, I am indebted to my family for their love and support throughout this journey.

Table of Contents

Abstract ii

Preface iv

Acknowledgements v

List of Tables vii

List of Figures viii

Chapter 1 Investigating the value of incorporating behavioural measures in a discriminant function developed for sex assignment in black-capped chickadees (*Poecile atricapillus*) 1

1. Introduction 1
2. Methods 8
   - 2.1 Study Area 8
   - 2.2 Morphometric measurements 9
   - 2.3 Behavioural measurements: Standardised assays 10
   - 2.4 Behavioural measurements: Natural behaviours 11
   - 2.5 DNA extraction and molecular sexing 12
   - *2.6* Statistical analysis: 13
3. Results 14
   - 3.1 Univariate analyses 14
   - 3.2 Discriminant functions 14
4. Discussion 16
   - 4.1 Conclusions and future directions 23
5. References 26
6. Appendix 37

List of Tables

Table 1**:** Comparison between male and female Black-caped chickadees for all A) morphometric and B) behavioural measurements. Repeatability for each trait was calculated by group-level variance for each trait by the sum of group level variance and residual variance. The definition of repeatability varies slightly depending on the context of the trait being measured. In traits for which we would expect there to be no variation in the measurement, repeatability solely represents the extent of measurement error in the trait. In traits for we would expect variation in the measurements over time/context, repeatability not only accounts for measurement error but also accounts for variation at the among-individual level. Differences in sample sizes across traits occurred because we were unable to obtain data for certain traits for some individuals either due to individuals escaping prior to the trait being measured or data entry issues. 31

Table S 1: Percentage of individuals correctly classified and comparison with percentage of individuals correctly scored when dataset was cross-validated with a leave-one-out analysis, when using a model with all morphological traits, or all highly repeatable morphological traits, on the same data set as was used in LD3 (Presented in Table 2). 37

List of Figures

Figure 1: Location of the eight feeders within the study site at the University of Alberta Botanical Gardens in Devon, Alberta, Canada during the Fall 2018 and Fall 2019 field season. The grey region within the map indicates the bounds of the public areas of the garden which visitors were able to access, and the white region indicates the forested areas which members of the public were not able to access. The small map at the bottom left of the figure represents the map of Canada with the approximate location of the study site. Each feeder was placed equidistantly at approximate 300m intervals (as determined from GPS coordinates) within the forested areas. Feeders were kept stocked with black-oil sunflower seeds prior to, throughout and immediately after the catching period to ensure regular usage of the feeders by chickadees. 33
Figure 2: Distribution of discriminant scores for male and female chickadees calculated based on LD2 for the total study population consisting of 469 individuals (238 females and 231 males). Note, these discriminant scores were not centered and standardized to facilitate application of discriminant scoring in the field. 36

Chapter 1

Investigating the value of incorporating behavioural measures in a discriminant function developed for sex assignment in black-capped chickadees (*Poecile atricapillus*)

1. Introduction

Darwin's original proposal of sexual selection theory (Darwin 1859, 1871) and Fisher's hypothesis of 'runaway selection' (Fisher 1915, Fisher 1930, O'Donald 1980) both offer a possible explanation for the development of dramatic phenotypic differences between the sexes of the same species. Indeed, the phenomenon of sexual dimorphism is one which is well established and widely studied within the field of ecology (Hedrick and Temeles 1989, Shine 1990a, Badyaev and Hill 2003). Numerous studies provide empirical evidence demonstrating the existence of biologically significant differences between the sexes, which is a consequence of differential selection pressure experienced by males and females (Fairbairn and Preziosi 1994, Fairbairn 1997, Fairbairn et al. 2007).

Sexual selection theory posits that males and females often need to adopt different strategies to navigate variation in both external environmental pressures (Shine 1990a, b), as well as in inherent biological limitations, to increase their fitness output both in terms of fecundity and survivorship. Reproductive costs are a clear example of an inherent biological limitation that differs between males and females (Chu and Lee 2012). Males in general are considered to have lower reproductive costs in their ability to produce multiple gametes and therefore the potential to have more offspring and consequently higher fecundity (Darwin 1871, Kodric-Brown and Brown 1987, Parker 2011). In contrast, females are thought to have an innate limitation in the number of offspring they can produce due to the relatively higher energetic costs associated with gamete production. Not surprisingly, biological sex is a key driver of behavioural variation among-individuals through the evolution of sex-specific behaviours. This includes behaviours directly associated with reproduction (Wachtmeister and Enquist 2000, Anholt et al. 2020) such as courtship rituals, mating behaviours, egg laying, etc. In addition to behaviours directly associated with reproductive output, sex specific differences can also occur

in behaviours that are thought to indirectly influence fecundity such as territorial defence and mate guarding in species where high levels of conspecific competition exists within sexes (Parker 1970, Clutton-Brock 1989, Flecknell 2002, Yokoi et al. 2016).

Outside of behavioural variations that have been found to be associated with fecundity, several studies have also shown behavioural variation that segregate along sex even in behaviours that are not *a priori* expected to be associated with sex such as migration distance (Gow and Wiebe 2014), migration duration (Holberton 1993) and sleep (Steinmeyer et al. 2010, Stuber et al. 2015). Some studies have shown differences across sexes in behaviours measured through standardised assays (Videlier et al. 2015, Simeonovska-Nikolova 2016, Tuliozi et al. 2018). These standardized assays have been criticized for being overly simplified, crude measures that fail to account for complexities involved in the expression of a behavioural trait such as context and taxonomical differences, and consequently have been thought to be limited in their utility for understanding individual's behaviour in the natural environment or an individual's fitness (Carter et al. 2012, Beckmann and Biro 2013, Carter et al. 2013). Critiques of the viability of standardised assays aside, the fact that a few studies have found sex differences even in these artificially construed phenotypic traits, further supports the idea that sex differences impose such a strong selective pressure on behavioural variation that consistent differences between males and females can be resolved even in the coarsest measures of behaviour, and even in cases where the biological significance of the trait itself is not fully understood.

Identifying the sex of an individual is especially critical to the study of animal personality, which specifically focuses on understanding behavioural variation among individuals. Failure to account for sex when delimiting behavioural differences will result in an overinflation of variance that is assigned to the among-individual level, resulting in a biased estimate of the proportion of variation in a trait within a population due to individual identity alone (Réale et al. 2007, Bell et al. 2009, Réale et al. 2010, Schuett et al. 2010, Wolf and Weissing 2012, Sih et al. 2015). Hence, in behavioural studies, sex is a key variable to measure to truly be able to elucidate patterns in behavioural variation and the association of behaviour with individual fitness and selection. In species with conspicuous sexual dimorphism, sex determination of individuals poses no real challenges. These species are characterised by pronounced

morphological differences between sexes through the development of secondary sexual characteristics that allow for males and females to be easily distinguished from each other (Darwin 1871, Fisher 1915, Fisher 1930). In species that have morphologically similar males and females, the observation of well-documented, sex-specific behaviours can also serve as an alternative means through which the sex of an individual can be determined (Witte and Curio 1999, Gill 2003, Volodin et al. 2015). For example, observation of egg-laying would allow for conclusive characterization of an individual as female. However, in cases of sexually monomorphic species where there is no data on sex-specific behaviours available, where those behaviours are rare or difficult to observe, or where observations occur outside of the season when sex-specific behaviours are expressed (e.g., non-breeding season), sex determination of an individual can prove to be a challenging process.

Molecular sexing through DNA analysis is a well-established method that can assign sex to individuals with high accuracy (Dubiec and Zagalska-Neubauer 2006, Morinha et al. 2012, Thanou et al. 2013). The high accuracy of molecular sexing techniques has resulted in their widespread use in sexually monomorphic species, species with limited sexual dimorphism or in studies where sex information of individuals was unavailable, with reliable molecular markers developed to differentiate between sex across different taxa (Griffiths et al. 1998, Huynen et al. 2002, Shaw et al. 2003, Rovatsos and Kratochvíl 2017). However, despite being a highly reliable tool for sex determination, molecular sexing presents challenges of its own in its application. Firstly, collecting a blood or tissue DNA sample in the field is typically an invasive process which may cause stress to the study organism, and runs the risk of the occurrence of adverse physical effects such as the formation of hematomas, excessive blood loss and other injuries (Voss et al. 2010), which in extreme cases may have detrimental effects on survivorship (Brown and Brown 2009). Proper training of the field personnel is necessary to ensure the safe collection of a DNA sample through invasive methods while minimising harm to the study organism. Secondly, the processing the DNA sample in the lab after collection to assign sex to individuals can be a time consuming and expensive procedure in terms of the material and equipment needed to collect, store and process the sample in the lab (reviewed in Morinha et al. 2012) compared to some commonly used non-molecular sex determination methods such as behavioural and

morphometric measurements. Molecular sexing also requires significant training to acquire the necessary lab work skills. Lastly, it must be acknowledged molecular sexing techniques are also subject to sources of error (Gebhardt and Waits 2008). Error rates in molecular sex assignment techniques can vary depending on the technique used, as well as with regards to the conditions involved pertaining to collection and storage of the DNA samples, as well as the execution of the lab work protocol.

In light of these challenges, discriminant functions can be implemented in lieu of molecular sexing tools to assay the sex of individuals in a sexually monomorphic species, by capitalising on small but consistent phenotypic differences that exist between males and females as a result of differential sexual selection (Price 1984, Delestrade 2001, Murphy 2007). While the influence of differential sexual selection on highly visible phenotypic traits like secondary sexual characteristics and sex-specific behaviours are immediately apparent, differential selection pressures can also result in occurrence of more nuanced phenotypic differences between sexes that are biologically significant, but are not immediately discernable or conclusive enough to serve as reliable indicators of sex on their own (Murphy 2007, Zefania et al. 2010). The inherent differences in strategies adopted by males and females in order to adapt to varying environmental and selection pressures and maximise their fitness results in the formation of small but consistent differences between sexes even in species that are deemed to be sexually monomorphic. Examples of these phenotypic differences could be in morphological traits such as body size and tarsus length which have been shown segregate differentially across sex in certain bird species such as black-capped chickadees (*Poecile atricapillus*) (Desrochers 1990), common mynas (*Acridotheres tristis*) (James et al. 1994) and black-headed gulls (*Chroicocephalus ridibundus*) (Indykiewicz et al. 2019).

Traditionally, discriminant functions have been built and used with great success using measurements of morphological traits (Desrochers 1990, Muriel et al. 2010). While any one morphological feature on its own may be insufficient to accurately distinguish between males and females, analysing several morphological features collectively can be effective for developing discriminant functions that allow males and females to be distinguished with a high degree of accuracy across many species. One study looking at a population of buff-breasted

wrens (*Cantorchilus leucotis*) achieved a 95.5% accurate sex assignment with a discriminant function from a single morphological measurement (wing chord), though it must be noted that the study had a relatively small sample size of 68 individuals (Gill and Vonhof 2006). More recently published discriminant functions have also begun including measures of non-morphological traits such as vocalisations (Stirnemann et al. 2015) in an attempt to increase the power of the function to be able to accurately assign sex to a larger proportion of individuals within the study population.

While discriminant functions have been heralded as a cheap, relatively easy way to overcome the hurdle of assigning sex to individuals in a sexually monomorphic species (Dechaume-Moncharmont et al. 2011), given that it would not incur the expenses associated with using molecular tools to assign sex, nor would it require research personnel that are trained to execute the necessary lab work protocol, the measurement of morphological features pose challenges of their own. Notably, not all morphological features are equally easy to measure with high accuracy, and research personnel must still be trained to take measurements across the different morphological features with a high degree of accuracy and consistency (Barrett et al. 1989, Yezerinac et al. 1992). For example, Perktas and Gosler (2010) found a large disparity in the measurement errors that occurred when measuring bill length (4.11%- 6.28%) compared to bill width (33.41%-38.58%), in a study that analysed morphometric measurements of common chaffinches (*Fringilla coelebs*) from museum samples. This demonstrates the risk that more “difficult to measure” morphological features may result in an overall decrease in the accuracy of those morphological features as a predictors or variables in any analysis. Measurement errors have been shown to increase the error rates in discriminant function and significantly diminish its ability to correctly assign sex to individuals (Dechaume-Moncharmont et al. 2011). Therefore, it is imperative that field personnel taking morphological measurements are properly trained and demonstrate high levels of accuracy and consistency in their measurements both within- and among- observers, to minimise the effect of observer identity and the resultant measurement error on the accuracy of the discriminant function that is generated from that morphological data (Barrett et al. 1989). Other sources of potential errors include incorporating highly correlated morphological features in the development discriminant

function as it skews the weight of the correlated morphological features in contributing sex differences in the population disproportionately within in the model, affecting the accuracy of the equation (Dechaume-Moncharmont et al. 2011). Additionally, collecting morphometric data in the field can also be time consuming and have the potential to stress individuals that often have to be handled extensively to complete measurements (O'Dell et al. 2014). Therefore, field personnel also need to prioritise measuring select important features to maximise measurement efficiency and minimise stress for each individual (Dechaume-Moncharmont et al. 2011). Thus, when developing a discriminant function, it is important to assess the type of phenotypic traits utilised in building the model and to determine the best combination of features to use in developing a discriminant function that accounts for the effort involved in acquiring reliable measurements of the selected phenotypic traits, as well as the relative risk of measurement errors associated with including each trait, against the relative improvement in discriminant rate sex with the inclusion of each particular phenotypic trait. Here, we develop a discriminant function for a population of black-capped chickadees (*Poecile atricapillus*) from central Alberta. Black-capped chickadees are small, non-migratory passerine birds that occur commonly over large range across North America (Smith 1997). They are sexually monomorphic and do not exhibit any sex-specific behaviours that can help to distinguish between the sexes in the non-breeding season (Desrochers 1990). Therefore, sex assignment for individuals in the non-breeding season can only occur either through molecular assays or through the development of a discriminant function built based on subtle morphological variations that occur between male and female chickadees.

Previously, a discriminant function based on three morphological traits (mass, wing length and rectrix length) was developed based on a chickadee population in Athabasca, Alberta, Canada, in the late 1980s with an overall 94% accuracy in sex assignments, with 92% and 95% of correct classification rate for females and males respectively (Desrochers 1990). The same discriminant function was subsequently adopted and utilised in sexing other chickadee populations around Canada (Christie et al. 2004b, van Oort and Otter 2005). Recent analyses have however suggested that the accuracy of a discriminant function can be reduced when applied to populations other than the original sample population that was used to develop the model, due

to population specific differences in the phenotypic traits used to build the model as well as possible differences in sex ratios that may exist between populations (Brennan et al. 1991, Ellrich et al. 2010, Sáez-Gómez et al. 2017, Indykiewicz et al. 2019). Long term temporal variation in population structure and morphological traits (Nowakowski 2000, Salewski et al. 2010) could also potentially result in waning accuracy of the discriminant function over time even when applied to the original sample population. Therefore, it is important to generate discriminant functions that are specific to the study populations and are developed based on relatively recent phenotypic data obtained from population to avoid potential sources of error in the application of the discriminant function.

In this thesis, a black-capped chickadee population at the University of Alberta Botanical Garden in Devon, Alberta, Canada was studied with the goal of a developing discriminant function based on morphological traits for the purpose of future sex assignment in the study population. Morphological traits measured included mass, wing length, tarsus length, bill length and bill depth; all of which are morphological features that have been demonstrated to be effective in discriminating between sex in other discriminant functions developed in other sexually monomorphic bird species (Desrochers 1990, Sikora and Dubiec 2007, Muriel et al. 2010). A major aim in the construction of this discriminant function was to determine the efficacy of the discriminant function when developed with different combinations of morphological features, with the goal of balancing the effort invested in the accurate measurement of various morphological traits relative to the improvement in discriminant rate that occurred with the inclusion of specific morphological traits in the discriminant function.

There was also a secondary goal of investigating the effects of using behavioural traits measured in the development of discriminant functions to determine if the inclusion of such behavioural traits improves the power of the function to correctly assign sex. Recent studies have demonstrated the inclusion of non-morphological traits such as vocalisations (Stirnemann et al. 2015) to be useful for improving the resolution of discriminant function. Several behavioural field studies have also found the presence of significant sex differences in traits previously thought to not have sex-specific expression (Holberton 1993, Steinmeyer et al. 2010, Gow and Wiebe 2014, Stuber et al. 2015). These studies demonstrate the potential value in the

inclusion of behavioural traits in improving the accuracy of discriminant function models. Behavioural measures that were analysed in this study include two standardised behavioural assays; a novel environmental test that is designed to provide a measure of exploratory behaviour (Verbeek et al. 1994, Dingemanse et al. 2002, Huang et al. 2016) and a handling aggression test (Brommer and Kluen 2012, Dubuc-Messier et al. 2016); as well as two measurements of natural behaviours; foraging rate and a measure of inter-visit intervals between foraging events at the feeders set up at the study site. Ultimately, this study aims to assess the effectiveness of including different combinations of morphological and behavioural traits to discriminate male and female Black-capped chickadees with high accuracy and minimal error.

2. Methods

2.1 Study Area

The study was conducted at the University of Alberta Botanical Gardens (formerly known as the Devonian Botanic Garden) a 0.97km^2 property, located approximately 22km south west of the city of Edmonton, Alberta, Canada. The garden is composed of 0.32km^2 of managed, cultivated garden that is meant for public display, as well as an additional 0.65km^2 of natural, forested area. A marked population was established in Fall 2017, and has been maintained through catching effort each subsequent fall/winter. Morphometric and behavioural data of 469 individuals (238 females and 231 males) collected in the population were used to develop discriminant functions.

Blacked-capped chickadees were caught through the use of mist-nets in fall and winter and were immediately banded with a unique identifying aluminium ring as well an exclusive combination of colour bands across both legs to allow for visual identification of an individual. A subset of birds (156 individuals) also received a passive integrated transponder (PIT) tag attached to their leg band which would enable them to be detected when visiting feeders in the study site that were radio frequency identification (RFID) readers. After completion of the banding process, individuals were then immediately subjected to two standardised behavioural

assays, a standardised cage test and a handling aggression test (see below for more information on behavioural measures). Upon completion of the assays, morphometric data (mass, beak length, beak depth, tarsus length, wing length) was measured by a single observer (JJW) to avoid inter-observer variation (see below), and a small blood sample (<20µl) was obtained using a fine gauge needle from the brachial vein and transferred to a Whatman FTA card which was subsequently stored in at room temperature in a tightly sealed, dry environment until use in molecular analysis (see below).

2.2 Morphometric measurements

Five morphometric features (bill depth, bill length, tarsus length, wing length and mass) were measured by a single observer to avoid inter-observer variation. Bill length, bill depth and tarsus length were measured using ± 0.01mm calipers. Bill length was measured from the lowest point of feathering on the beak as the starting point till the end point of the upper beak. Bill depth was measured at the widest point of the beak close to the base of the beak. Tarsus length was measured from the notch of the intertarsal joint to the base of the toes. In order to obtain a measurement for wing length, the bend of a wing was pressed against the crevice of a ± 0.5mm stopped ruler, with the wing flattened along the length of the ruler, and the length of the longest primary feather was taken as a measurement of wing length. For subsequent use in the discriminant functions, the average measurement value for bill depth/bill length/tarsus length/wing length was taken for individuals for which we had multiple measurements (due to recaptures). There was generally high repeatability in the measurement values for all the measured features (Table 1).

Mass was recorded to the nearest 0.25 g using a Pesola scale. Seasonal and diurnal variation in mass have been demonstrated in Black-capped chickadees and hence, we considered using Best Linear Unbiased Predictors (BLUPs) for mass that accounted for these effects. BLUPs were predicted values of body mass for individuals derived from linear mixed effects models that accounted for seasonal (e.g., civil day length, average temperature etc.) and diurnal effects (e.g., time of day) on body mass. However, the accuracy in sex assignment of the discriminant function did not improve when the BLUP for mass was used compared to when the average mass for each individual was use (results not shown). Hence, the average mass for each

individual was used in all subsequent applications of the discriminant functions given the advantage that it can be measured directly in the field.

2.3 Behavioural measurements: Standardised assays

2.3.1 Cage Test: Measurement of exploration

A cage test to measure exploratory behaviour was designed with a similar protocol to that described by Stuber et al (2013). Individual chickadees were exposed to a novel environment in the form of cage (61cm length x 39cm width x 40 cm height) which had metal grills along the length of the cage on one side (which provided the only source of natural light and view to the external environment) and was completely closed off on all other sides with opaque plastic siding. The cage itself was completely empty and devoid of any features apart from three perches. During each cage test, a video recording of each individual was captured from a camera placed 2m away from the cage, while field personnel stayed out of the direct line of sight of the bird to minimise the effects of human interference during the course of the cage test. Observers scored the video recordings using the programme BORIS (Friard and Gamba 2016) to determine the total number of movements of an individual chickadee (as calculated by summing up the number of hops and flights) over the course of a two-minute period. Prior to the start of cage test, individual chickadees were kept in a small compartment adjacent to the cage and released into the cage by removing the barrier between the compartment and the cage area when the observation period began. To avoid familiarisation to the novel environment there was a minimum of a 3-week period implemented between each novel environment test for all recaptured individuals and an average cage test score was generated for all individuals with repeated measures.

2.3.2 Handling Aggression Test: Measurement of aggression

A handling aggression test to generate an aggression score for each individual was designed with a similar protocol to that described by Dubuc-Messier et al. (2016). Immediately after the cage test, individual chickadees were removed from the cage and held in an upright position by its legs, with its head and wings free to move (i.e., photographer's grip). The observer than held their forefinger 3cm in front of their bird (noting if the bird pecked at their finger or not) and then subsequently began to move their forefinger back and forth towards the bird, stopping

just short of the bird's bill. The aggression of the bird was then scored on a 4 point scale from 0 to 3; 0 = no aggression as indicating by the absence of pecking at the finger or any sort of reaction to the finger, 1 = low aggression as indicated by the bird pecked at the finger, but only when provoked, 2 = moderate aggression as indicated by the bird pecking at the finger spontaneously (i.e., without observer moving the finger towards the bird and spread its tail feathers, and 3 = high aggression that involved pecking at the finger, spreading the tail feathers, and flapping the wings.

2.4 Behavioural measurements: Natural behaviours

We also obtained data on natural foraging rates. Eight feeders were set up within the forested areas of the garden at approximate 300m intervals to ensure minimal human disturbance to chickadee foraging and were implemented in the Fall 2018 and Fall 2019 field seasons. The feeders were equipped with antennas near the feeder opening that were connected to radio frequency identification (RFID) readers that were able to detect the presence of a passive integrated transponder (PIT) tag within the immediate vicinity of the feeder opening. The feeders were thus able to register the foraging visits of chickadees equipped with PIT tags on their leg bands (156 individuals: 74 females and 82 males) and determine the time of the visit, as well as the identity of the individual detected at the feeder. Antenna data was then filtered by implementing a minimum threshold of 5 seconds between each successive detection of the same individual. This was done to avoid overinflation of actual foraging visits that may potentially occur due to multiple detections of the PIT tag within a single foraging event by the same individual. The 5 second threshold was determined based on video recordings made of marked chickadees using the feeders (see Arteaga-Torres et al. 2020 for further details)

Feeders were kept stocked with black-oil sunflower seeds. Because foraging behaviour shows seasonal variation, we standardized the time frame in which foraging data was collected in each of the two years. We did not use data from the 7 days immediately following catching sessions, to minimize the effects of catching on observed foraging rates. Foraging data was used from the 7-8 days following this recovery period; from 27th Nov-3rd December in 2018 and from 13th Nov-20th Nov in 2019. We calculated the total number of foraging visits for each individual chickadee across all the feeders per day, and the inter-visit intervals (IVI) between foraging events.

2.5 DNA extraction and molecular sexing

A Harris Uni-core punch was used to obtain a bloodspot of 1.2mm diameter from the Whatman FTA card of each individual and DNA extraction was carried out using Extracta DBS (Quantabio) extractions reagent following the manufacturer's recommended DNA extraction protocol. Molecular sex determination of individuals was carried out using a modified version of the sexing protocol described by Griffith's et al. (1998) using the P2 (reverse)/P8 (forward) sexing primer system that sequenced the chromobox-helicase-DNA-binding gene (CHD-W and CHD-Z) that generated PCR products of different sequence lengths across the two genes.

The reaction mixture for the PCR included the following reagents: 6.27 μl water, 1.5 μl 10x PCR Buffer (100mM Tris pH 8.8, 1% Triton X-100, 500mM KCl, 1.6mg/ml BSA), 0.9 μl $MgCl_2$ (25mM) , 0.3 μl dNTPs (10mM), 0.48 μl P2 (10 μM) + P8 (2.5 μM) primer mixture, 0.48 μl of 5' fluorescent labelled P8 primer with PET (10 μM), 0.075 μl of Taq polymerase (from Qiagen), 5 μl of DNA template.

The PCR conditions were as follows:

94°C	5 min		
	94°C	30 sec	x 30
	48°C	45 sec	
	72°C	45 sec	
	94°C	30 sec	x 8
	53°C	45 sec	
	72°C	45 sec	
	72°C	10 min	
	4°C	Pause	

Fragment analysis was conducted on the resultant PCR products on the ABI3730 DNA analyzer, with a sample mixture consisting of 4 μl of PCR product, 8 μl of HiDi and 0.25 μl of Liz 500 size standard.

Following fragment analysis, data was imported onto Genemapper to sex individuals according to genotype, with the resolution of two peaks indicating two alleles of varying sequence length

(344bp/400bp) identifying an individual as a heterozygous female, and the resolution of a single peak indicating the presence of two alleles of identical sequence length (344bp/344bp) identifying an individual as a homozygous male.

2.6 Statistical analysis:

We conducted our analyses in two steps. First, we assessed which of the morphological and behavioural traits measured varied across sexes to inform which traits could potential contribute to a sex discrimination function. This was done by running a linear mixed effects model in the R statistical environment (R Development Core Team 2018) using the lmer function from the lme4 package (Bates et al. 2015) with sex (determined molecularly) fitted as a fixed effect, and individual identity and feeder identity (for traits involving the measurement of natural foraging rates) fitted as random effects.

We used the result from the analyses described above to inform which traits we could consider in our linear discriminant function (LDA). We only used traits which showed significant sex-related differences in our discriminant function (see Results, and Table 1). We considered three different linear discriminant functions based on traits that showed sex-specific differences: 1) a linear discriminant function including all morphological traits, 2) a linear discriminant function that only included morphological rates with high repeatability in measurements to minimise the effects of measurement errors and 3) a linear discriminant function that included all morphological traits and one behavioural trait (foraging IVI). Linear discriminant functions were run in R using the lda function from the MASS package (Ripley 2002) using a random sample split of 80% training data relative to 20% test data for each model variant to determine the accuracy of sex assignment for each combination of variables. Given that a randomly generated seed set was used for each model in the linear discriminant analysis, to ensure that the model would still have similar accuracy rates regardless of the composition of the training seed set, a jackknifing analysis was done to determine if accuracy rates would remain similar for the discriminant function despite the changing composition of the training data set. A jackknifing analysis involves predicting the sex of one individual using the remaining individuals as a training data set, followed by cycling through all the individuals in the data set until the sex of

all individuals are predicted and calculating the overall accuracy rate of the combined predictions. A high accuracy rate would indicate that data set is robust such that the discriminant function would have a high ability to correctly assign sex to unknown individuals despite the varying composition of individuals in the training data set. This would lend support to using a linear discriminant function generated from a random training data set and applying it for unknown members in the study population.

3. Results

3.1 Univariate analyses

First, we assessed whether there were sex differences in each of the morphological and behavioural traits we recorded. Each of our morphological measures showed significant sex differences (Table 1), with males being larger compared to females. However, foraging IVI was the only behavioural trait that showed significant sex differences (Table 1); males had shorter IVIs compared to females (Table 1).

3.2 Discriminant functions

Given that our univariate analyses found significant sex differences in all of the morphological traits (Table 1), for our first model, we ran the discriminant function using the average score for all morphological traits for which significant differences had been found between males and females (Table 2). This linear discriminant function showed an overall accuracy of 86% (84.8% Males, 87.2% Females), with a slightly higher accuracy rate demonstrated when the data set was cross-validated with a leave-one-out (jackknifing) analysis that had an overall accuracy of 89.1% (90.9% Males, 87.3% Females). The high accuracy rates obtained from the jackknifing analysis demonstrates the overall robustness of the dataset, supporting the application of this discriminant function equation on unknown individuals from the same test population.

In an effort to improve accuracy and minimise the effects of measurement errors, a second model was run using only morphological features that had high repeatability ($r > 0.5$) in measurements, thus removing the measurements of bill length and bill depth from the model. This second model had an overall accuracy of 89.2% (91.3% males, 87.2% females) (Table 2) which resulted in a slightly higher proportion overall of correctly classified individuals in

comparison to the first model. The cross-validation of the data set with the leave-one-out (jackknifing) analysis showed similar overall accuracy of 88.5% (89.2% males and 87.8% females).

The frequency distribution of discriminant scores based of the second model (Figure 2) shows that female chickadees have a discriminant score range of 73-81 (mean = 77.3, sd = ±1.6) and male chickadees have a discriminant score range of 77-86 (mean = 81, sd = ±1.58). Thus, all individuals with a discriminant score < 77 can be classified as female, and all individuals with a discriminant score > 81 can be classified as male with 100% accuracy. Individuals with discriminant scores between 77 and 81 cannot be assigned as male or female conclusively. However, smaller values within this range are much more likely to be female; only 4% of individuals that had score range of 77 to 78 were male. At the same time, birds at the high end of this range (i.e., with discriminant scores between 80 and 81) are much more likely to be male; 68% of individuals in this range were male.

Lastly, since significant sex differences were also found in foraging IVI (Table 1), a third model was run on a reduced sample size of 155 individuals (82 males, 73 females) for which foraging data was available, to investigate the potential value of incorporating behavioural traits into discriminant function. This linear discriminant function showed an overall accuracy of 90% (93.8% males, 85.7% females) (Table 2), with a slightly higher accuracy demonstrated when the data set was cross-validated with a leave-one-out (jackknifing) analysis that had an overall accuracy of 91.6% (92.7% males, 90.4% females). The high accuracy rates obtained from the cross-validation analysis demonstrates that the data set is still robust despite the smaller sample size necessitated by the inclusion of the behavioural trait (foraging IVI) in this model. Further models were run on the same data set (i.e., reduced sample size data set) to compare accuracy rates when i) only all the morphological traits were used, and ii) when only highly repeatable morphological traits were used (Supplementary Information: Table S1), both of which produced similar accuracy ranges as LD3.

4. Discussion

Here, we investigated the utility of using morphological and behavioural measures to develop a sex discriminant function for chickadees. Discriminant functions built using different combinations of morphological features had similarly high accuracy rates suggesting that both models (LD1 and LD2) would be equally valid in assigning sex to unknown individuals in the population in the future. The similar accuracy rates of the cross-validation analysis of the models despite the differences in the composition of the data supports the idea proposed by Dechaume-Moncharmont et al. (2011) that changes to model validation methods and application of stepwise variable selection do not significantly affect overall model accuracy when sample sizes are large (i.e., N > 200). Surprisingly, even with the comparatively smaller data set used in LD3 (N = 155), accuracy varied minimally across the different models tested (see supplementary information table 1) suggesting that after a certain threshold, further increase in sample sizes did not contribute towards significant increases in model accuracy. Below, we discuss the utility and limitations of our discriminant function for sex determination in chickadees and provide suggestions for the future work in this area.

The high accuracy in ability to discriminate between both sexes using morphological trait measurements alone are not surprising given that significant differences were found between males and females for all morphological traits measured, with males being generally larger than females, in line with the findings of previously published literature (Smith 1997). This study had slightly lower correct sex assignment rate than Desrochers's (1990) study that looked a chickadee population in a similarly Northern range, which achieved an overall accuracy of 93.7% (92% female accuracy, 95% male accuracy). This study had a similar sample size to our study (457 individuals: 195 females and 263 males) with a model than implemented a sample split of approximately 69% training data (143 females and 171 males) relative to 31% test data (52 females, 92 males). However, it must be noted that this study had a skewed sex ratio (43% females: 57% males) with a larger proportion of males in the study population, that skewed even more drastically in the test data set where there were 75% more male individuals tested compared to females.

Variation in sex ratios have been demonstrated to introduce bias into discriminant functions (Brennan et al. 1991, Sáez-Gómez et al. 2017). In highly skewed populations, the ability for a model to correctly assign sex to individuals that belong to the majority sex is high. However, this comes at the cost of greater error rates in assigning sex to individuals that belong to the minority sex, as the model lacks sufficient data about the minority sex to ensure the correct sex assignment of individuals in the minority sex that may present a more intermediate phenotype. Therefore, while there appears to be an overall high accuracy for both male and female sex assignments in the model proposed by Desrochers (1990), it must be considered that there was a significantly smaller proportion of females evaluated in the test data, which could potentially explain why there was such high accuracies for female sex assignments even though the dataset skewed towards males. Thus, it may be possible that implementing that discriminant function in a test data set that skewed more heavily towards the females, the accuracy of the correct sex assignments may reduce slightly.

In contrast, our study (238 females and 231 males) had an approximate 1:1 sex ratio. This sex ratio is in line with what would be expected in a natural population, suggesting that our sampling efforts were not biased. Our test data set had a total of 93 individuals (47 females and 46 males) and also demonstrates a female: male population split that would be similar to what would be expected for our wild study population. This provides support that our model is not influenced by sex ratio biases and is likely to maintain similar accuracy rates for future sex assignments for individuals in the population.

Another aspect to consider with respect to Desrochers's study (Desrochers 1990) is the type of morphological features that were analysed. Similar to our study, Desrochers' most successful discriminant function also only utilised three morphological features: wing length, mass and rectrix length. Given the high accuracies rates attained by his model, some consideration should be given about the possibility of including the measurement of rectrix length in our model as a variable that can potentially improve correct sex assignment rates. However, the inclusion of rectrix length could also introduce potential measurement error into our model due to the seasonal wear and tear of rectrix feathers which would affect length measurements, as well as due to the higher probability of rectrix feathers dropping off, as they tend to be more loosely

attached to the body in prey species (Møller et al. 2006). Desrochers (1990) did not account for the possible effects of seasonal wear and tear on rectrix length in their study, and hence the potential measurement error it would introduce into a model is unknown. Additionally, the improvement in the accuracy of their model when the measurement of rectrix length was also included as opposed to just utilising wing length and mass measurements in the model was only 0.4%, suggesting that the inclusion of rectrix length only caused a minimal increase in correct sex assignments. Lastly, measurements of rectrix feathers are typically done by removing the entire feather from an individual. Removing a rectrix feather from a small prey species like a chickadee could impact their ability to maneuver through their natural environment and hence could potentially pose fitness costs to individuals (McDonald and Griffith 2011). Additionally, birds that are already missing rectrix feathers upon capture may not be able to have additional rectrix feathers removed, and therefore, may not be able to be sexed using a model that requires rectrix data. Therefore, the possibility of including the measurement of rectrix length in our model is one that needs to weigh the potential improvement the measurement could bring to our model against the possible risks of introducing measurement error and negatively impacting study individuals. Given, the minimal improvement in accuracy the inclusion of rectrix length caused in the model used by Desrochers (1990), we would argue the potential costs of including this measurement in our model would likely outweigh any probable increase in correct sex assignment rates that may occur due to the inclusion of an additional morphological feature.

The model also revealed clear cut off points in the discriminant score to assign sex to males (>81) and females (<77). However, 31.1% of individuals within our population fell within the intermediate score range (78 ≤ 81), with 41.8% of individuals falling within an intermediate score range of 77 ≤ 81 if more conservative thresholds were established to account for the slight possibility of smaller male individuals occurring within the score range of 77 to 78. This clearly shows that a large proportion of individuals within our study population present an intermediate phenotype and consequently, there will be a high degree of uncertainty in assigning sex to individuals within this range.

An obvious solution is to selectively implement alternative sex confirmation techniques such as DNA analysis only for the individuals that fall within this intermediate range to ensure correct sex assignment. However, a caveat to consider with this solution is to ensure that we do not accidentally introduce confounding effects into our study population through a biased field processing protocol. For examples, if we were to only collect blood samples from individuals with an intermediate phenotype for DNA analysis, this would mean that only smaller male individuals and larger female individuals in the population would be subjected to additional handling in the field. This additional field processing may potentially induce physiological and behavioural changes that are only experienced by a biased sample of individuals within our study population. Therefore, if the effects of additional handling in the field were to manifest itself as altered foraging patterns (e.g., a longer latency to resume feeding at the feeder after a potentially more stressful experience from blood sampling) and we were to only subject a biased sample within our population (e.g., smaller males and larger females) to this process, this could create confounding effects on the foraging patterns we observe. We may then risk drawing an incorrect conclusion between the association of body size and the latency to resume feeding after capture, when the pattern that is resolved is a direct consequence of differential field processing and not an effect of body size. Therefore, the way to eliminate any confounding effects that may be caused by differential field processing would be to enforce a field protocol where a random sub-sample of the population is subjected to blood sampling (e.g., one in every three individuals captured has blood taken irrespective of the discriminant score). After such a random sampling protocol, it would then be possible to verify if there are truly any confounding effects that result due to the additional field processing and if no such effects are found, a biased sampling protocol can then be implemented subsequently.

This study also attempted to utilise behavioural measurements to develop our discriminant function by using both natural behavioural traits, as well as traits measured through standardised assays. In the case of our study population, there were only significant differences found in one of the four traits measured. Two of the studied traits were measured through standardised assays; specifically they were an exploration score generated through a novel environment test (Stuber et al. 2013), as well as an aggression score generated through a

handling aggression test (Dubuc-Messier et al. 2016). Standardised assays are behavioural assessments that are simple to execute in a uniform manner and hence serve as an efficient tool to generate large scale data on behavioural variation within a study population. There has been much debate about the biological significance of measuring these abstract traits by exposing organisms to artificial stimuli and extrapolating them to behaviours observed in the natural environment (Carter et al. 2012, Beckmann and Biro 2013, Carter et al. 2013). Given the disjunct between the contrived nature of a standardised assay and the natural cues an organism would encounter in its habitat, we would not *a priori* expect organisms to develop variation in a biologically significant manner in response to a standardised assay. Thus, in the case of the two standardised assays that we utilised in the study, it is perhaps not surprising that we found no significant difference between male and female chickadees in either the cage test or the handling aggression test.

However, while standardised assays may have inherent limitations in being able to capture biologically meaningful traits, several studies have found that traits measured through standardised assays have been correlated to measurements of natural behaviours (Verbeek et al. 1994, Marchetti and Drent 2000, Nicolaus et al. 2015). Proponents for standardised assays argue that while these assays may not perfectly measure the trait that is perceived as being assessed (i.e., movement in the novel environment is related to navigation of the natural environment, or handling aggression is related to aggression exhibited in the wild), the coarseness of the measurements obtained from standardised assays means that they could potentially capture some variation related to a natural behaviour of interest. This could explain why other studies have found significant sex differences in traits measured through standardised assays (Videlier et al. 2015, Simeonovska-Nikolova 2016, Tuliozi et al. 2018). Therefore, the potential of traits being measured through standardised assays being implemented as a proxy to gather sex specific information about individuals within our population should not be disregarded. Implementing nuanced standardised assays that are still quick and easy to execute, offer the possibility for us to resolve sex-specific differences through the measurement of behavioural traits in our population.

We also assessed a natural behaviour, foraging, using two different metrics: foraging rate (daily visit rate to feeders), and the inter-visit interval between foraging events at the feeder (seconds), in this study. Of the two metrics of foraging behaviour, only foraging IVI was found to have a small, significant differences between the sexes. Given that foraging rate and IVI represent two alternative metrics of the same data, the finding that one had significant sex differences while the other did not is somewhat surprising. However, the observed effect was small, and the larger sample size for IVI provided greater statistical power for detecting a small effect.

Our model that included foraging IVI as a trait along with all five measured morphological features (LD3) produced an accuracy rate of 90% that was comparable with that output of the first two models (LD1 and LD2) despite its smaller sample size. While the output of the model seemingly suggests that the inclusion of the behavioural trait results in high accuracy, comparison of the same dataset when only all of the morphological traits were used in the model revealed an overall higher accuracy of 93% (100% males, 85.7% females) (See supplementary information Table S1), and when only highly repeatable morphological traits were used in a model , a similar accuracy rate of 90% (100% males, 78.57% females) was generated (See supplementary information Table S1). This is consistent with the earlier observed pattern of LD1 and LD2 having similar model accuracy outputs despite having different variables, providing further support for the idea the number of variables included in the model become less important with regards to model accuracy past a certain threshold in sample size. It is apparent that the inclusion of the behavioural trait in the model provided no notable increase in the accuracy of the model. This was perhaps not an unforeseen result given that that the sex differences in foraging IVI had a relatively smaller effect size of 0.358 (Table 1). Given that that the difference between male and female foraging IVI was small, it is understandable that the inclusion of the trait did not increase the resolution of the model.

Black-capped chickadees have linear dominance hierarchies where males are more highly ranked than females (Smith 1997), and earlier work suggests that more dominant individuals would have more ready access to foraging sites than subordinates (Glase 1973). Previous studies (Desrochers 1989), including one conducted in our study population (Arteaga-Torres et

al. 2020) have also reported sex-related differences in foraging patterns in chickadees . Thus, we had initially predicted differential foraging patterns emerging between the more dominant males and less dominant females. However, there was only a small, albeit significant, difference between the sexes in foraging IVI and our data did not reveal any differences in foraging rate between males and females. This result is consistent with other studies that also did not find any significant effects of sex or dominance rank on foraging rates in chickadees (Van Buskirk and Smith 1989, Ficken et al. 1990). This demonstrates that sex-specific patterns in behaviour are subject to variation and may be both population specific and context dependent. In the case of our study population, within the same study we show that two measures of foraging derived from the same data can show differences in statistical significance and strength of association with sex; with foraging IVI presenting statistically significant differences between sexes with an effect size of 0.35, in comparison to foraging rate which did not have statistically significant differences, with a smaller effect size of 0.25. This result further highlights the importance for researchers to not simply assume which behaviours would resolve sex-related differences. Researchers should validate their assumptions by determining if they can observe considerable sex effects on behavioural traits within in their study population, as well as by testing to see if the use of the traits introduces any noise into their model, before making decisions on which behavioural measurements would prove most useful for integrating into a discriminant function.

A caveat to consider when making decisions on which behavioural traits to include in a discriminant function is to determine whether any potential confounding effects on future behavioural analyses could emerge by using a specific behavioural trait to develop a discriminant function. For example, it would not be appropriate to assess sex-related differences in foraging behaviour if foraging behaviour itself was used to determine sex. Therefore, the type of behavioural data that is included in discriminant functions should be critically evaluated such that the chosen behaviour is easy to measure in a consistent manner over time, and that it does not introduce confounding effects in future behavioural analyses conducted in the same population. In this aspect, behavioural measurements from standardised assays may prove particularly valuable, because if these artificially construed

measurements were used for sex assignment, they would be independent of measures of the natural behaviours in a population.

Another opportunity to improve the accuracy of future discriminant functions would be to use other phenotypic measures such as vocalizations or plumage characteristics. Vocalisations have previously been used to improve correct sex assignment rates in a population of Ma'oma'o (*Gymnomyza samoensis*) (Stirnemann et al. 2015). While songs have been traditionally regarded as a sex-specific behaviour exhibited by male songbirds, recent research across several songbird species (Garamszegi et al. 2006, Odom et al. 2014), including chickadees (Hahn et al. 2013), have revealed that female birds are also capable of songs. Chickadees in particular have been shown to be able to identify both the sex (Hahn et al. 2015) and the identity of specific individuals through their songs (Montenegro et al. 2020), offering the possibility of incorporating acoustic analyses to discriminate between males and females in our model. Additionally, black-capped chickadees display sexually dichromatic plumage characteristics (Mennill et al. 2003), with a canonical discriminant function being able to discriminate between males and females with 90% accuracy based on reflectance scores generated from the mantle and cheeks, as well as from the measurements of bib area and cap area. Thus, the analysis of reflectance scores observed in various plumage characteristics through the use of portable field spectrophotometry devices, as well as measurements of the size of prominent plumage characteristics such as the bib (Otter and Ratcliffe 1999, Mennill et al. 2003) are options to explore in future field seasons.

4.1 Conclusions and future directions

In this study, we developed a discriminant function that was able to assign sex to individuals in our population with high accuracy (89%) using three morphometric measurements: body mass, wing length, and tarsus length. We demonstrated that inclusion of additional traits in the discriminant function did not translate to significant increase in accuracy and that the decision to include more traits in a discriminant function must balanced against the increased investment of time and effort that comes with measuring additional traits with high precision, with the gains in discriminant function performance.

We recommend that discriminant functions should be developed and applied in population specific and context dependent manner that account for variation in traits that may exist at the between and within population level for both morphometric and behavioural traits. Historically, while discriminant functions may have been developed based on one population and have been extrapolated to other populations, this does not account for spatial, temporal and context specific variation that have been shown to influence how various phenotypic traits present across populations and within populations. For example, the discriminant function developed by Desrochers (1990) based on a central Albertan chickadee population have been applied in chickadee populations in Ontario (van Oort and Otter 2005) and British Columbia (Christie et al. 2004a). However, chickadees from northern populations have been suggested to be somewhat larger than their southern counterparts (Smith 1997). Consequently, applying a discriminant function developed based on a population of larger northern birds within a population of smaller southern birds is likely to introduce bias into the model and thus diminish accuracy rates. Additionally, applying a discriminant function indiscriminately in a population without first validating that that sex-related variation exists in the analysed traits poses the risk of nullifying the validity of the function in that population entirely.

Even on smaller geographical scales, we would recommend caution in applying the same discriminant function across multiple populations as even on a smaller microhabitat level such as rural and urban areas, different morphotypes could present themselves due to differences in the environmental context (Hutton and McGraw 2016). Temporal variation could be another potential source of error that could diminish accuracy rates, both on a smaller scale (e.g., seasonal effects) as well as on a larger scale due to population level changes over the years due to varying environmental conditions, depending on the traits included in the discriminant function. Thus, in the cases of long-term application of a discriminant function to the same study population, we suggest periodic cross-validation of the model with other sex-confirmation techniques to determine if model accuracies remain consistent with time. Finally, we emphasize the need to be selective in the choice of traits being measured for discriminant analysis. This involves verifying the presence of class-related variation in the chosen trait, determining the ease of measuring the trait with high precision and lastly, in the case of more

complex measurements such as behavioural traits, identifying potential risk of confounding effects in future analyses should the chosen trait be used.

5. References

Anholt, R. R. H., P. O'Grady, M. F. Wolfner, and S. T. Harbison. 2020. Evolution of reproductive behavior. Genetics **214**:49-73.

Arteaga-Torres, J. D., J. J. Wijmenga, and K. J. Mathot. 2020. Visual cues of predation risk outweigh acoustic cues: a field experiment in black-capped chickadees. Proceedings of the Royal Society B: Biological Sciences **287**:20202002.

Badyaev, A. V., and G. E. Hill. 2003. Avian sexual dichromatism in relation to phylogeny and ecology. Annual Review of Ecology, Evolution and Systematics **34**:27 - 49.

Barrett, R., M. Peterz, R. Furness, and J. Durinck. 1989. The variability of biometric measurements. Ringing & Migration **10**:13-16.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software **67**:48.

Beckmann, C., and P. A. Biro. 2013. On the validity of a single (boldness) assay in personality research. Ethology **119**:937-947.

Bell, A. M., S. J. Hankison, and K. L. Laskowski. 2009. The repeatability of behaviour: a meta-analysis. Animal Behaviour **77**:771-783.

Brennan, L. A., J. B. Buchanan, C. T. Schick, and S. G. Herman. 1991. Estimating sex ratios with discriminant function analysis: the influence of probability cutpoints and sample size. Journal of Field Ornithology:357-366.

Brommer, J. E., and E. Kluen. 2012. Exploring the genetics of nestling personality traits in a wild passerine bird: testing the phenotypic gambit. Ecol Evol **2**:3032-3044.

Brown, M. B., and C. R. Brown. 2009. Blood sampling reduces annual survival in cliff swallows (*Petrochelidon pyrrhonota*). The Auk **126**:853-861, 859.

Carter, A., H. Marshall, R. Heinsohn, and G. Cowlishaw. 2012. How not to measure boldness: novel object and antipredator responses are not the same in wild baboons. Animal Behaviour **84**:603–609.

Carter, A. J., W. E. Feeney, H. H. Marshall, G. Cowlishaw, and R. Heinsohn. 2013. Animal personality: what are behavioural ecologists measuring? Biol Rev Camb Philos Soc **88**:465-475.

Christie, P. J., D. J. Mennill, and L. M. Ratcliffe. 2004a. Chickadee song structure is individually distinctive over long broadcast distances. Behaviour **141**:101-124.

Christie, P. J., D. J. Mennill, and L. M. Ratcliffe. 2004b. Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. Behavioral Ecology and Sociobiology **55**:341-348.

Chu, C. Y. C., and R. D. Lee. 2012. Sexual dimorphism and sexual selection: a unified economic analysis. Theor Popul Biol **82**:355-363.

Clutton-Brock, T. H. 1989. Review lecture: mammalian mating systems. Proceedings of the Royal Society of London. Series B, Biological Sciences **236**:339-372.

Darwin, C. 1859. The origin of species by means of natural selection John Murray, London.

Darwin, C. 1871. The descent of man, and selection in relation to sex. London: Murray **415**.

Dechaume-Moncharmont, F.-X., K. Monceau, and F. Cézilly. 2011. Sexing birds using discriminant function analysis: a critical appraisal. The Auk **128**:78-86.

Delestrade, A. 2001. Sexual size dimorphism and positive assortative mating in alpine choughs (*Pyrrhocorax graculus*). The Auk **118**:553-556.

Desrochers, A. 1989. Sex, dominance, and microhabitat use in wintering black-capped chickadees: a field experiment. Ecology **70**:636-645.

Desrochers, A. 1990. Sex determination of black-capped chickadees with a discriminant analysis. Journal of Field Ornithology **61**:79-84.
Dingemanse, N. J., C. Both, P. J. Drent, K. Van Oers, and A. J. Van Noordwijk. 2002. Repeatability and heritability of exploratory behaviour in great tits from the wild. Animal Behaviour **64**:929-938.
Dubiec, A., and M. Zagalska-Neubauer. 2006. Molecular techniques for sex identification in birds. Biological Letters **43**:3-12.
Dubuc-Messier, G., D. Réale, P. Perret, and A. Charmantier. 2016. Environmental heterogeneity and population differences in blue tits personality traits. Behavioral Ecology **28**:448-459.
Ellrich, H., V. Salewski, and W. Fiedler. 2010. Morphological sexing of passerines: not valid over larger geographical scales. Journal of Ornithology **151**:449-458.
Fairbairn, D. J. 1997. Allometry for sexual size dimorphism: Pattern and process in the coevolution of body size in males and females. Annual Review of Ecology, Evolution and Systematics **28**:659 - 687.
Fairbairn, D. J., W. U. Blanckenhorn, and T. Székely. 2007. Sex, size & gender roles: evolutionary studies of sexual size dimorphism. Oxford Univ. Press, Oxford.
Fairbairn, D. J., and R. F. Preziosi. 1994. Sexual selection and the evolution of allometry for sexual size dimorphism in the water strider, Aquarius remigis. American Naturalist **144**:101 - 118.
Ficken, M. S., C. M. Weise, and J. W. Popp. 1990. Dominance rank and resource access in winter flocks of black-capped chickadees. The Wilson Bulletin **102**:623-633.
Fisher, R. A. 1915. The evolution of sexual preference. The Eugenics Review **7**:184.
Fisher, R. A. 1930. The genetical theory of natural selection. Oxford Clarendon Press.
Flecknell, P. 2002. Replacement, reduction and refinement. Altex **19**:73-78.
Friard, O., and M. Gamba. 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods in Ecology and Evolution **7**:1325-1330.
Garamszegi, L. Z., D. Z. Pavlova, M. Eens, and A. P. Møller. 2006. The evolution of song in female birds in Europe. Behavioral Ecology **18**:86-96.
Gebhardt, K. J., and L. P. Waits. 2008. High Error Rates for Avian Molecular Sex Identification Primer Sets Applied to Molted Feathers. Journal of Field Ornithology **79**:286-292.
Gill, S., and M. Vonhof. 2006. Sexing monochromatic birds in the field: Cryptic sexual size dimorphism in Buff-breasted wrens (Thryothorus leucotis). Ornitologia Neotropical **17**:409-418.
Gill, S. A. 2003. Timing and duration of egg laying in duetting Buff-breasted Wrens. Journal of Field Ornithology **74**:31-36.
Glase, J. C. 1973. Ecology of social organization in the Black-capped Chickadee. Living Bird **12**:235-267.
Gow, E. A., and K. L. Wiebe. 2014. Males migrate farther than females in a differential migrant: an examination of the fasting endurance hypothesis. Royal Society open science **1**:140346-140346.
Griffiths, R., M. Double, K. Griffiths, and R. Dawson. 1998. A DNA test to sex most birds. Molecular Ecology **7**:1071-1075.
Hahn, A. H., J. Hoang, N. McMillan, K. Campbell, J. Congdon, and C. B. Sturdy. 2015. Biological salience influences performance and acoustic mechanisms for the discrimination of male and female songs. Animal Behaviour **104**:213-228.
Hahn, A. H., A. Krysler, and C. B. Sturdy. 2013. Female song in black-capped chickadees (Poecile atricapillus): Acoustic song features that contain individual identity information and sex differences. Behavioural Processes **98**:98-105.
Hedrick, A. V., and E. J. Temeles. 1989. The evolution of sexual dimorphism in animals: Hypotheses and tests. Trends in Ecology & Evolution **4**:136 - 138.
Holberton, R. L. 1993. An endogenous basis for differential migration in the dark-eyed junco. The Condor **95**:580-587.

Huang, P., K. Kerman, K. E. Sieving, and C. M. St. Mary. 2016. Evaluating the novel-environment test for measurement of exploration by bird species. Journal of Ethology **34**:45-51.
Hutton, P., and K. J. McGraw. 2016. Urban–rural differences in eye, bill, and skull allometry in house finches (*Haemorhous mexicanus*). Integrative and Comparative Biology **56**:1215-1224.
Huynen, L., C. D. Millar, and D. M. Lambert. 2002. A DNA test to sex ratite birds. Mol Ecol **11**:851-856.
Indykiewicz, P., P. Minias, J. Kowalski, and P. Podlaszczuk. 2019. Shortcomings of discriminant functions: a case study of sex identification in the black-headed gull. Ardeola **66**:361-372, 312.
James, J. C., N. Kang, K. J. Ahmad, and K. Wang Luan. 1994. Discriminant Analysis of Morphometric Characters as a Means of Sexing Mynas Journal of Field Ornithology **65**:1-7.
Kodric-Brown, A., and J. H. Brown. 1987. Anisogamy, sexual selection, and the evolution and maintenance of sex. Evolutionary Ecology **1**:95-105.
Marchetti, C., and P. J. Drent. 2000. Individual differences in the use of social information in foraging by captive great tits. Animal Behaviour **60**:131-140.
McDonald, P. G., and S. C. Griffith. 2011. To pluck or not to pluck: the hidden ethical and scientific costs of relying on feathers as a primary source of DNA. Journal of Avian Biology **42**:197-203.
Mennill, D. J., S. M. Doucet, R. Montgomerie, and L. M. Ratcliffe. 2003. Achromatic color variation in black-capped chickadees, Poecile atricapilla: black and white signals of sex and rank. Behavioral Ecology and Sociobiology **53**:350-357.
Møller, A. P., J. T. Nielsen, and J. Erritzøe. 2006. Losing the last feather: feather loss as an antipredator adaptation in birds. Behavioral Ecology **17**:1046-1056.
Montenegro, C., W. D. Service, E. N. Scully, S. K. Mischler, K. A. Campbell, and C. B. Sturdy. 2020. Black-capped Chickadees (*Poecile atricapillus*) can identify individual females by their fee-bee songs. The Auk **137**.
Morinha, F., J. A. Cabral, and E. Bastos. 2012. Molecular sexing of birds: A comparative review of polymerase chain reaction (PCR)-based methods. Theriogenology **78**:703-714.
Muriel, R., E. Casado, D. Schmidt, C. P. Calabuig, and M. Ferrer. 2010. Morphometric sex determination of young Ospreys Pandion haliaetus using discriminant analysis. Bird Study **57**:336-343.
Murphy, M. T. 2007. A cautionary tale: cryptic sexual size dimorphism in a socially monogamous passerine. The Auk **124**:515-525.
Nicolaus, M., K. J. Mathot, Y. G. Araya-Ajoy, A. Mutzel, J. J. Wijmenga, B. Kempenaers, and N. J. Dingemanse. 2015. Does coping style predict optimization? An experimental test in a wild passerine bird. Proceedings of the Royal Society B: Biological Sciences **282**:20142405.
Nowakowski, J. J. 2000. Long-term variability of wing length in a population of the Reed Warbler Acrocephalus scirpaceus. Acta Ornithologica **35**:173-182.
O'Dell, D., M. Carlo, A. Kimmitt, E. Bikowski, K. Morris, and A. Dolby. 2014. A comparison of techniques measuring stress in birds. Virginia journal of science **65**:133-149.
O'Donald, P. 1980. Genetic models of sexual selection. Cambridge University Press.
Odom, K. J., M. L. Hall, K. Riebel, K. E. Omland, and N. E. Langmore. 2014. Female song is widespread and ancestral in songbirds. Nature Communications **5**:3379.
Otter, K., and L. Ratcliffe. 1999. Relationship of bib size to age and sex in the black-capped chickadee Journal of Field Ornithology **70**:567-577.
Parker, G. A. 1970. Sperm competition and its evolutionary consequences in the insects. Biological Reviews **45**:525-567.
Parker, G. A. 2011. The origin and maintenance of two sexes (anisogamy), and their gamete sizes by gamete competition.
Perktas, U., and A. Gosler. 2010. Measurement error revisited: its importance for the analysis of size and shape of birds. Acta Ornithologica **45**:161-172.

Price, T. D. 1984. The evolution of sexual dimorphism in Darwin's finches. American Naturalist **123**:500 - 518.
R Development Core Team. 2018. R: A language and environment for statistical computing. See http://www.R-project.org, Vienna, Austria.
Réale, D., N. Dingemanse, A. Kazem, and J. Wright. 2010. Evolutionary and ecological approaches to the study of personality. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **365**:3937-3946.
Réale, D., S. M. Reader, D. Sol, P. T. McDougall, and N. J. Dingemanse. 2007. Integrating animal temperament within ecology and evolution. Biological Reviews **82**:291-318.
Ripley, W. N. V. a. B. D. 2002. Modern applied statistics with S. Fourth edition. Springer, New York.
Rovatsos, M., and L. Kratochvíl. 2017. Molecular sexing applicable in 4000 species of lizards and snakes? From dream to real possibility. Methods in Ecology and Evolution **8**:902-906.
Sáez-Gómez, P., C. Camacho, A. Álvarez, N. Varo, C. Ramo, and A. J. Green. 2017. Sexual dimorphism and sex determination in Black-necked Grebes Podiceps nigricollis: on the need to account for geographical variation in morphology and sex ratio. Bird Study **64**:502-512.
Salewski, V., W. M. Hochachka, and W. Fiedler. 2010. Global warming and Bergmann's rule: do central European passerines adjust their body size to rising temperatures? Oecologia **162**:247-260.
Schuett, W., T. Tregenza, and S. R. X. Dall. 2010. Sexual selection and animal personality. Biological Reviews **85**:217-246.
Shaw, C. N., P. J. Wilson, and B. N. White. 2003. A reliable molecular method of gender determination for mammals. Journal of Mammalogy **84**:123-128.
Shine, R. 1990a. Ecological causes for the evolution of sexual dimorphism: a review of the evidence. The Quarterly review of biology **64**:419-461.
Shine, R. 1990b. Proximate determinants of sexual differences in adult body size. American Naturalist **135**:278 - 283.
Sih, A., K. J. Mathot, M. Moiron, P. O. Montiglio, M. Wolf, and N. J. Dingemanse. 2015. Animal personality and state-behaviour feedbacks: a review and guide for empiricists. Trends in Ecology & Evolution **30**:50-60.
Sikora, A., and A. Dubiec. 2007. Sex identification of jack snipe (*Lymnocryptes minimus)* by discriminant analysis of morphometric measurements. Ardea **95**:125-133.
Simeonovska-Nikolova, D. 2016. Sex differences in novel object recognition with similar objects in mound-building mouse, Mus spicilegus (Mammalia: Rodentia). Comptes rendus de l'Académie bulgare des sciences: sciences mathématiques et naturelles **69**.
Smith, S. M. 1997. Black-Capped chickadee. Stackpole Books.
Steinmeyer, C., H. Schielzeth, J. Mueller, and B. Kempenaers. 2010. Variation in sleep behaviour in free-living blue tits, Cyanistes caeruleus: Effects of sex, age and environment. Animal Behaviour **80**:853-864.
Stirnemann, R. L., M. A. Potter, D. Butler, and E. O. Minot. 2015. Acoustic differences enable sex discrimination in Ma'oma'o (*Gymnomyza samoensis*), a species with high sexual morphological overlap. The Wilson Journal of Ornithology **127**:376-386, 311.
Stuber, E. F., Y. G. Araya-Ajoy, K. J. Mathot, A. Mutzel, M. Nicolaus, J. J. Wijmenga, J. C. Mueller, and N. J. Dingemanse. 2013. Slow explorers take less risk: a problem of sampling bias in ecological studies. Behavioral Ecology **24**:1092-1098.
Stuber, E. F., K. J. Mathot, B. Kempenaers, N. J. Dingemanse, and J. C. Mueller. 2015. Sex-specific association between sleep and basal metabolic rate in great tits. Animal Behaviour **109**:15-22.
Thanou, E., S. Giokas, V. Goutner, V. Liordos, and S. Fraguedakis-Tsolis. 2013. Efficiency and accuracy of PCR-based sex determination methods in the European *Phalacrocoracidae*. Annales Zoologici Fennici **50**:52-63, 12.

Tuliozi, B., G. Fracasso, H. Hoi, and M. Griggio. 2018. House sparrows' (*Passer domesticus*) behaviour in a novel environment is modulated by social context and familiarity in a sex-specific manner. Frontiers in Zoology **15**:16.
Van Buskirk, J., and D. C. Smith. 1989. Individual variation in winter foraging of black-capped chickadees. Behavioral Ecology and Sociobiology **24**:257-263.
van Oort, H., and K. A. Otter. 2005. Natal nutrition and the habitat distributions of male and female black-capped chickadees. Canadian Journal of Zoology **83**:1495-1501.
Verbeek, M. E., P. J. Drent, and P. R. Wiepkema. 1994. Consistent individual differences in early exploratory behaviour of male great tits. Animal Behaviour **48**:1113-1121.
Videlier, M., R. Cornette, C. Bonneaud, and A. Herrel. 2015. Sexual differences in exploration behavior in *Xenopus tropicalis*? The Journal of experimental biology **218**.
Volodin, I. A., E. V. Volodina, A. V. Klenova, and V. A. Matrosova. 2015. Gender identification using acoustic analysis in birds without external sexual dimorphism. Avian Research **6**:20.
Voss, M., D. Shutler, and J. Werner. 2010. A hard look at blood sampling of birds. The Auk **127**:704-708, 705.
Wachtmeister, C.-A., and M. Enquist. 2000. The evolution of courtship rituals in monogamous species. Behavioral Ecology **11**:405-410.
Witte, K., and E. Curio. 1999. Sexes of a monomorphic species differ in preference for mates with a novel trait. Behavioral Ecology **10**:15-21.
Wolf, M., and F. J. Weissing. 2012. Animal personalities: consequences for ecology and evolution. Trends in Ecology & Evolution **27**:452-461.
Yezerinac, S. M., S. C. Lougheed, and P. Handford. 1992. Measurement error and morphometric studies: statistical power and observer experience. Systematic Biology **41**:471-482.
Yokoi, S., S. Ansai, M. Kinoshita, K. Naruse, Y. Kamei, L. J. Young, T. Okuyama, and H. Takeuchi. 2016. Mate-guarding behavior enhances male reproductive success via familiarization with mating partners in medaka fish. Frontiers in Zoology **13**:21.
Zefania, S., E. Razafimahatratra, P. Faria, M. Bruford, P. Long, and T. Székely. 2010. Cryptic sexual size dimorphism in Malagasy plovers *Charadrius* spp. OSTRICH **81**:173-178.

Table 1**:** Comparison between male and female Black-caped chickadees for all A) morphometric and B) behavioural measurements. Repeatability for each trait was calculated by group-level variance for each trait by the sum of group level variance and residual variance. The definition of repeatability varies slightly depending on the context of the trait being measured. In traits for which we would expect there to be no variation in the measurement, repeatability solely represents the extent of measurement error in the trait. In traits for which we would expect variation in the measurements over time/context, repeatability not only accounts for measurement error but also accounts for variation at the among-individual level. Differences in sample sizes across traits occurred because we were unable to obtain data for certain traits for some individuals either due to individuals escaping prior to the trait being measured or data entry issues.

	Males		Females		Cohen's D	T-test	P-value	Repeatability of measurement
	Average ± SD	N individuals	Average ± SD	N individuals				
A) Morphological traits								
Mass (g)	11.61 ± 0.52	231	10.77± 0.52	237	1.63	18.91	<0.00001	0.52
Wing length (mm)	69.46 ± 1.51	228	66.42 ± 1.53	236	2.01	21.92	<0.00001	0.57
Tarsus length (mm)	16.50 ± 0.44	230	15.99 ± 0.48	237	1.13	12.56	<0.00001	0.81
Bill length (mm)	9.52 ± 0.33	230	9.38 ± 0.33	237	0.45	5.09	<0.00001	0.35
Bill depth (mm)	4.19 ± 0.17	230	4.1 ± 0.19	237	0.49	4.77	<0.00001	0.30

B) Behavioural traits								
Cage test score	109.84 ± 29.77	232	106.95 ± 30.81	236	0.095	1.86	0.063	0.38
Handling aggression	1.16 ± 1.14	231	1.06 ± 1.10	237	0.089	1.16	0.25	0.34
Foraging rate[1] (visits/day)	215.09 ± 136.17	82	183.47 ± 111.57	74	0.25	1.67	0.098	0.80
Foraging inter-visit interval (IVI in seconds)[1] (log-transformed)	1.91 ± 0.59	82	1.93 ± 0.59	73	0.36	-2.24	0.027	0.03

1. Foraging rate and inter-visit interval (IVI) are two different ways of presenting the same data and are therefore not independent. We considered both variables because we did not know *a priori* which would have greater power to detect sex-specific effects, if they were present. IVI retains each interval between successive visits by an individual as a measure of foraging, resulting multiple measures per day compared to foraging rate, which collapses all foraging visits into a single data point per day. Thus, IVI has both a larger sample size (potentially greater statistical power), but also greater within-individual variance (potentially reduced power) compared with foraging rate.

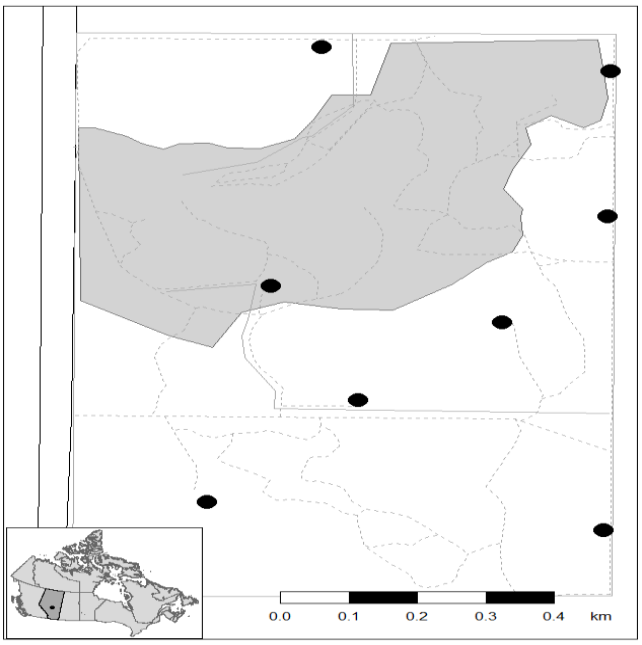

Figure 1: Location of the eight feeders within the study site at the University of Alberta Botanical Gardens in Devon, Alberta, Canada during the Fall 2018 and Fall 2019 field season. The grey region within the map indicates the bounds of the public areas of the garden which visitors were able to access, and the white region indicates the forested areas which members of the public were not able to access. The small map at the bottom left of the figure represents the map of Canada with the approximate location of the study site. Each feeder was placed equidistantly at approximate 300m intervals (as determined from GPS coordinates) within the forested areas. Feeders were kept stocked with black-oil sunflower seeds prior to, throughout and immediately after the catching period to ensure regular usage of the feeders by chickadees.

Table 2: Percentage of individuals correctly classified using different discriminant functions and comparison with percentage of individuals correctly scored when dataset was cross-validated with a leave-one-out analysis.

		Linear Discriminant Model			**Leave-one-out analysis**		
Discriminant function:	**Sample size**	**Total accuracy**	**% of males accurately scored**	**% of females accurately scored**	**Total accuracy**	**% of males accurately scored**	**% of females accurately scored**
LD1: all morphological traits LD1= 0.4636 (Body mass) +1.0526 (Wing length) + 0.33659 (Tarsus length) – 0.03455(Bill length) + 0.16482 (Bill depth)	238 F 231 M	86	84.8	87.2	89.1	90.9	87.3
LD2: Only highly repeatable morphological traits LD2= 0.55886 (Body mass) + 1.0064 (Wing length) + 0.28042 (Tarsus length)	238 F 231 M	89.2	91.3	87.2	88.5	89.2	87.8

LD3: All morphological traits + IVI LD3= -0.198016295 (Foraging IVI) + 0.667880475 (Body mass) + 1.070798829 (Wing length) + 0.273376365 (Tarsus length) – 0.002242264 (Bill length) + 0.103580241 (Bill depth)	73 F 82 M	90	93.8	85.7	91.6	92.7	90.4

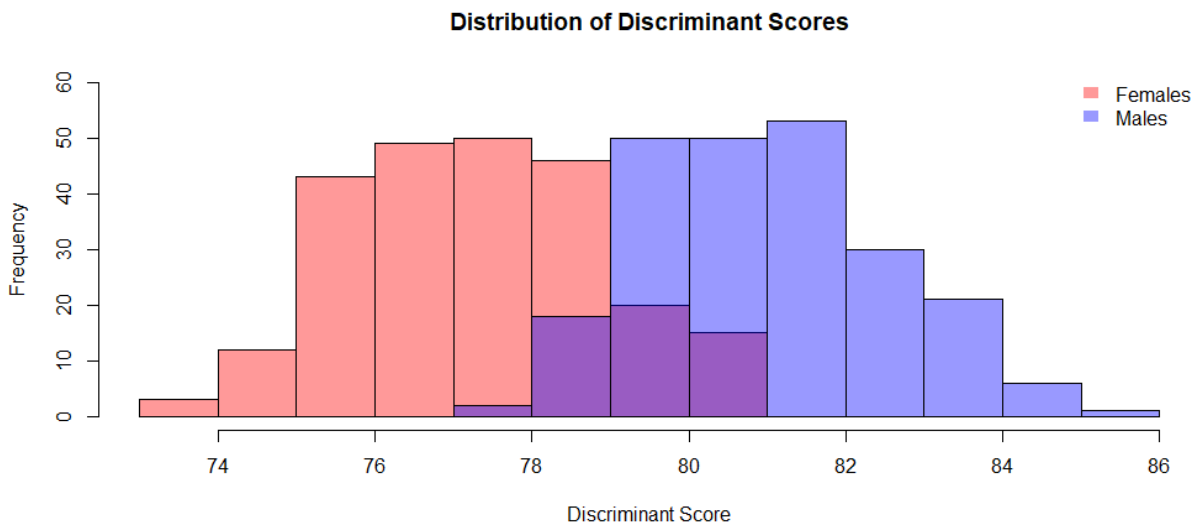

Figure 2: Distribution of discriminant scores for male and female chickadees calculated based on LD2 for the total study population consisting of 469 individuals (238 females and 231 males). Note, these discriminant scores were not centered and standardized to facilitate application of discriminant scoring in the field.

6. Appendix

Table S 1: Percentage of individuals correctly classified and comparison with percentage of individuals correctly scored when dataset was cross-validated with a leave-one-out analysis, when using a model with all morphological traits, or all highly repeatable morphological traits, on the same data set as was used in LD3 (Presented in Table 2).

		Linear Discriminant Model			**Leave-one-out analysis**		
Discriminant function	**Sample Size**	**Total accuracy**	**% of males accurately scored**	**% of females accurately scored**	**Total accuracy**	**% of males accurately scored**	**% of females accurately scored**
Model with all morphological traits LD= 0.640334899 (Body mass) + 1.074185542 (Wing length) + 0.326115894 (Tarsus length) - 0.007857426 (Bill length) + 0.099479770 (Bill depth)	73 F, 82 M	93.3	100	85.7	91	93	89
Model with only highly repeatable morphological traits	73 F, 82 M	90	100	78.6	89.7	90.2	89

LD= 0.6748764 (Body mass) + 1.0606799 (Wing length) + 0.3237492 (Tarsus length)							