Genetics of migration timing in bar-tailed godwits

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Zoology

at Massey University, Manawatū, New Zealand.

Ángela María Parody Merino

2018

Dedicado a mis padres;

vosotros me disteis alas para volar.

Preface

In this thesis, I used molecular and genomic techniques to give insight on the genetic elements behind migratory timing behaviour in a very suitable natural population bird system: The New Zealand godwit. This thesis is structured as a series of related and connected manuscripts with exception of the introduction, which provides the terminology, background and founding of my thesis: Each chapter is a stand-alone piece of work, therefore, there will be some unavoidable repetition between them and the introduction. This thesis contains three main chapters (see listed below), which are intended to be published.

- Chapter 2: Parody-Merino A.M., Battley P.F., Verkuil Y. I., Conklin J.R., Prosdocimi F., Potter, M.A., Fidler A.E. Population genetic structure of New Zealand-overwintering bar-tailed godwits (*Limosa lapponica baueri*) in relation to breeding geography and migration timing.
- Chapter 3: Parody-Merino A.M., Battley P.F., Conklin J.R., Fidler A.E. No evidence for an association between *Clock* gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*).
- Chapter 4: Parody-Merino A.M., Battley P. F., Fidler A.E., Prosdocimi F., Lima N., Conklin J.R., Potter M.A., Cox M.P. Polygenic architecture underlies departure time in a long-distance migratory bird.

As part of a Marsden Grant project I developed the research questions addressed in this thesis, as well as most of the work (i.e. experiments, molecular data collection, analysis and writing), which was performed by me with the guidance of supervisors and some of the co-authors. The godwit behavioural data from previous years were collected principally by my supervisor Phil Battley and co-author Jesse Conklin; those from 2013 onwards were collected by Jesse Conklin (Manawatu), Tom Burns and Ian Southey (Miranda) and Peter Langlands (Catlins). I have collaborated in fieldwork and helped in the collection of behavioural data (in the Catlins) whenever I could. Yvonne Verkuil has been an important contributor to Chapter 2 by guiding in analyses and subsequent interpretation of results. Francisco Prosdocimi and Nicholas C. B. Lima were in charge of the bioinformatic process of the godwit genome assembly and annotation, which was crucial for Chapter 2 and Chapter 4. My co-author Murray Cox supervised and guided

me for the analyses (i.e. bioinformatic steps) of Chapter 4, as well as helping resolve computational needs. He also made substantial contributions to the manuscript. All supervisors provided comments and contributed to framing of all the manuscripts.

Synopsis

"A hidden drive at the right time"

Ferdinand Johann Adam von Pernau

Bird migration is one of the most amazing behaviours observed in nature and it fascinates because of its diversity and its – still – unclear reasons and origin. My thesis converges primarily two big areas of research in biology: ecology and genetics/genomics. The ecology part is the bird system (the godwit), which has been extensively studied and whose individual departure date has been proven to be exceptionally consistent. Studies of this kind are not possible without a suitable system, and longitudinal behavioural records of the same individuals are not easy to obtain. The genetics and genomics are, without doubt, the biggest part of this thesis. What can we learn about how DNA encodes biological rhythms in natural populations taking godwits as an example?

There is a huge amount of literature on chronobiology, mainly based on experiments with model organisms or caged individuals. Thanks to these experiments the elements (i.e. proteins and genes) involved in the biological clock have been identified and the characteristics of circadian and circannual rhythms have been described. However, it remains unclear how these elements link to an individual's timing behaviour in nature. Extensive research on the migration timing of bar-tailed godwits (*Limosa lapponica baueri*) in New Zealand gave rise to an intriguing observation: individual's departure dates seemed to be primarily driven by an 'internal signal', and this 'internal signal' was quite consistent across the years, while the population showed a departure-span of approximately a month. This raised questions such as: Are there elements of the internal clock that determines such within species (population) diversity and such intra-individual consistency? In my thesis research I first I assessed the population genetic structure of godwits related to migration time, a step that is necessary when trying to link phenotypes with genotypes. Then I used genomic approaches and integrated behavioural data to understand whether elements related to the 'internal clock' are associated to individuals' migratory departure time in godwits.

I found evidence of slight population genetic structure between northern and southern breeders as well as between earlier and later migrants departing from the stop-

over in Asia, but not between earlier and later migrants from New Zealand. Detailed analyses of migration timing in relation to polymorphisms in *clock* (a gene implicated in other studies as potentially influencing migration timing) found no support for *clock* having any role in godwit migration timing. Analysis of variation in 120 genes associated with the internal clock, photoreception, the physiology the Hypothalamic-Pituitary-Gonadal Axis – which modulates the internal clock –, and fat metabolism and storage indicates that individual migration timing of godwits has a genetic basis to some degree, but differences between individuals seem to be associated with large numbers of genes of small effect rather than a few genes of large influence. Godwit migration timing therefore appears to be a complex trait in which genetic differences between individuals explain some of the variation timing, but a large amount of the variation observed is not explained by the suite of genes studied. It is possible that key genes exist that were not studied, and/or that non-genetic factors may be influencing an individual's decision to migrate on a given day above and beyond the genetic influence. In general, this thesis contributes to the understanding of the nature of behaviours (i.e. genes-behaviour link) in natural populations, specifically in the area of chronobiology.

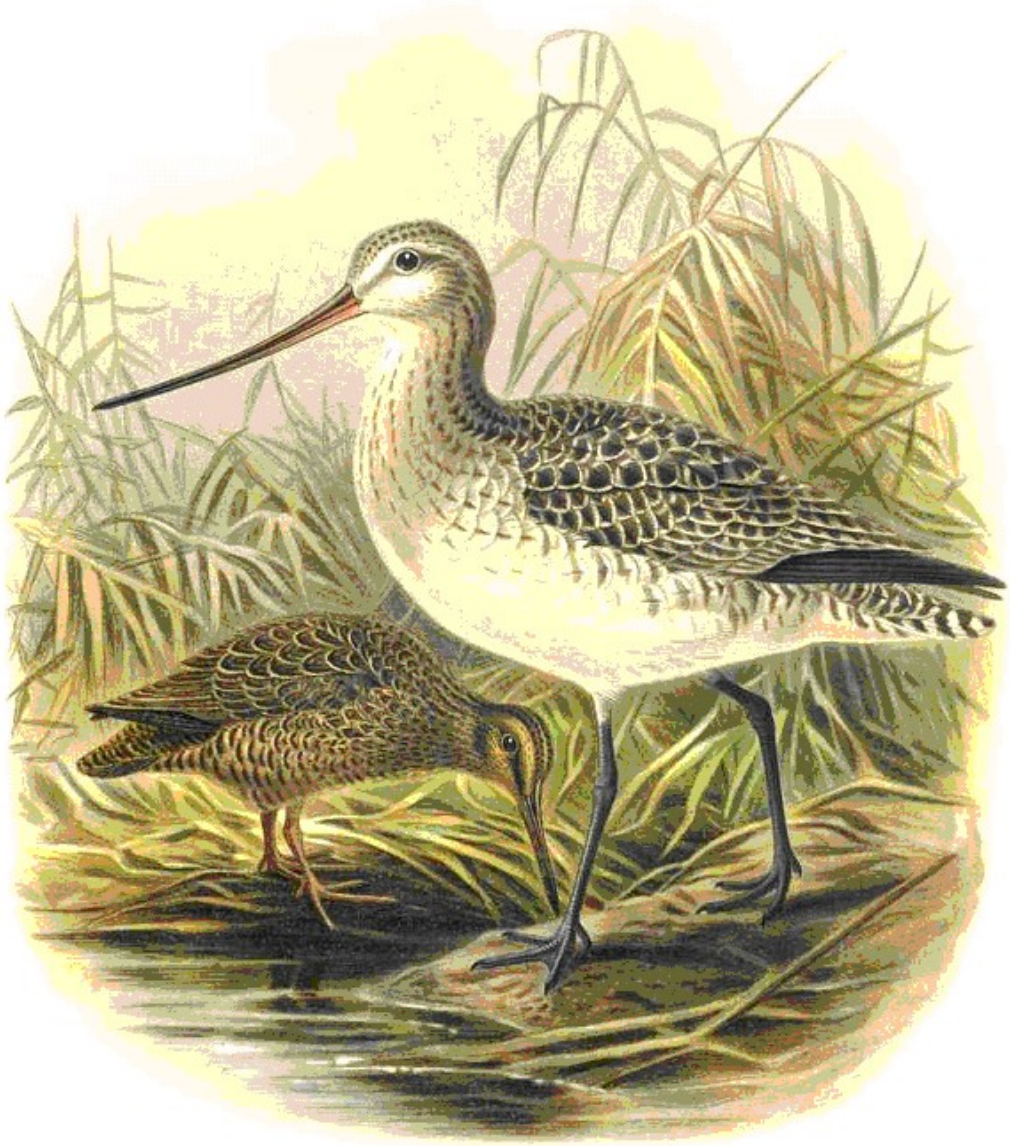
Te Reigna – rarangi mai ra te rangai kuaka/ kia tau hikohiko he pai tu waho

Flocks of godwits are gathering/ moving restlessly on the seaward cliffs

Part of a saying composed by Tumatahina of Te Aupouri.

“Kuaka” is the Māori name for the godwit.

Taken from: http://www.hekuaka.co.nz/



Acknowledgements

First, I want to thank many people whose contribution to my research, because although this thesis is a result of many hours of work on my own, there are a bunch of people that, without their help, I would not have accomplished and fulfilled my goals. My supervisors, Andrew Fidler and Phil Battley, I cannot be more grateful that you trusted on me to carry on this project: I want to thank you for all the help and support to progress, improve and complete this thesis. Andrew: for all what I learnt about the 'magical' world of the molecular laboratory procedures and being there to discuss about my thesis every time I needed. Not least, for always caring about me, as a person, not only as a student; for introducing me in New Zealand for the first time and making me easier to get used to the everything-new life here. Phil: for telling me all about godwits. I am immensely grateful for giving me the opportunity to be in the field, which made me feel like a complete biologist: field and lab (well, and office). Being close to the godwits, observing their movements for hours, surviving the 4-seasons – in 4-minutes – typical from New Zealand when transitioning to the winter, and waiting for some individuals to show me their second – and hidden – leg to check their colour bands. For welcoming me when I had to move to Palmy and for your incredibly useful help and support during the last years of my doctorate. To Murray Potter, for your support and your contributions and good ideas to improve the chapters. I really appreciate your support on providing me the opportunity to take the online course on "bioinformatic tools for genomic data", which helped me tremendously to clarify many of the bioinformatic work done by collaborators/ co-authors, as well as to perform the main analyses in Chapter 4. Enormous thanks to Jesse Conklin, whose invaluable work on godwits' behaviour and ecology has been the foundation of my thesis. And of course, for all your contribution, comments, ideas and explanations that helped me during the progress of my thesis.

Yvonne Verkuil, I am so grateful for that day we met and during our conversation I suddenly I found out you were the author of one important publication of my molecular experimental design. I cannot be more grateful to you for having me guided through the world of population genetic analysis. All the comments and guidance were essential for me to improve this chapter and to learn a lot. Thanks for the invitation to the University of Groningen in 2014, and for all those delicious dinners you invited me for during my stay. To Murray Cox, thanks so much for your valuable guidance and support, which were

essential to accomplish one of the most difficult of the challenges I encountered during my PhD – and also my most wished professional and personal accomplishment : to feel confident performing bioinformatic analyses. There is a long way of learning, but now I feel I am prepared for it. Thanks to the numerous people helped to collect data and samples in the field: David Melville, Adrian Riegen, Rob Schuckard, Peter Langlands, Ian Southey, Tom Burns…etc and a long list of volunteers!

I cannot forget about the people that contributed most to my well-being: My family and friends. This thesis is especially dedicated to my parents, who have always been with me, helping me to every step that I needed to make in order to reach my goals. To my friends, the ones from overseas and the ones I met in New Zealand. They all have been an enormous support as well. In particular, I want to thank Hugo, for all the support and good moments during the initial years of my PhD in which I was based in Nelson (South Island). Thanks to Andrea and David for the numerous enriching and enjoyable conversations we had during lunch (or coffee) breaks; I hope this continues! I want also to thank to friends that made me discover new sports and activities in New Zealand, which gave me some balance during my PhD life: Anika for the African dances, Chifuyu and Amy for Body combat, and Hugo and Alan for those games in Squash. Thanks to those who have accompanied me to camping and tramping adventures around this beautiful country. Last – and not least but most! – to Jorge, who followed me to this adventure. I would never be able to thank you enough for all your unconditional support.

Myself checking banded godwits in a beautiful day at the Catlins Coast (March 2014).

Table of Contents

Chapter 1

General introduction

Setting nets and cannons for shorebird catching. Photo taken at the Manawatu River estuary (November 2013)

.

Time-programed lives

The primary source of energy of life on our planet is the star, and helium fusion reactor, at the centre of our solar system: The Sun. The Earth rotates around both its own axis and around the sun with approximately 365 rotations (days) for every rotation around the sun (years). In addition, the distance between the sun and the Earth changes throughout this translational movement. Moreover, the Earth's axial tilt on a 23˚ angle causes differences in day length across latitudes (Hut *et al.* 2013). As a result, there is predictable variation in solar energy along the year, dependent on latitude. The biological consequence is that plants and animals have to deal with changes in **photoperiod**[1] (or daylength) and energy distribution, which also determines changes in productivity and temperature, among others. As a consequence, animals, especially those living at temperate regions where the annual variation is the greatest, have developed annual movements, called **migrations**, in order to find better resources and to avoid difficult weather. **Seasonal migration** is precisely one of the most used strategies by birds, with species living half of the year in one hemisphere and the other half in the other hemisphere. Seasonal migration strategies are quite variable across species (Pulido 2007a), with some species travelling short distances, some others long-distances, even species that are **opportunistic**. Whatever the strategy, an **internal clock** mechanism aware of the environmental conditions gives advantages to animals as they can anticipate and prepare for the migratory journey (Zhang & Kay 2010). The importance of such an internal clock tells us that it must be printed, somehow, in the genetic code.

Chronobiology: Daily and annual rhythms.

Doubtless from earliest times humans were very aware of the daily and annual changes around them, including in animal behaviour. Chronobiology (from the ancient Greek χρόνος, *chrónos* meaning 'time') is the study of how living organisms measure, anticipate and respond to rhythmic changes in their environment (Dunlap *et al.* 2004). There are two main rhythms that orchestrate most of the physiological and behavioural events in all organisms: Circadian rhythms (from Latin, *circadiem* = 'about a day'), which are events that occur with a **period** of approximately 24 h (Hirota & Fukada 2004) and circannual rhythms (from Latin, circannual = 'about a year' (Visser *et al.* 2010), which

[1] bold terms are defined in the glossary (page 25).

are events that occur with a period of approximately a year (such as godwits' onset of migration).

Circadian rhythms have been the most extensively studied due to the shorter period, which facilitates experimental studies. These experiments focussed on information from the external environment, therefore, allowing study of the nature and properties of the '**free-running rhythms'** (De Coursey 1960; Rajaratnam & Redman 1998; Daan *et al.* 2002; Stephan 2002; Gwinner 2003). The critical observation that different individuals of the same species differed in the length of their 'free running rhythm' periods, when isolated from environment inputs, argued for such rhythms being endogenously generated (Aschoff *et al.* 1971; Biebach *et al.* 1991; Gwinner 1996; Dunlap 1999; Daan 2000; Nisimura & Numata 2001; Goymann *et al.* 2012). This issue was finally resolved by the application of genetics, first to *Drosophila* and then to a wide range of **model organisms**, which showed that **allelic** variants of single **loci** could influence free-running rhythm period **phenotypes** in a classic **Mendelian** manner (Goldbeter 1995; Lewis *et al.* 1997; Price 1997). Such genetic evidence clearly established that the free-running rhythms observed under constant environmental conditions were generated by an '**endogenous clock**' (Gwinner 1996; Berthold 1999; Møller 2001; Pulido *et al.* 2001; Pulido & Berthold 2003; Styrsky *et al.* 2004; Liedvogel *et al.* 2009; Kumar *et al.* 2010; Goymann *et al.* 2012). Nonetheless, the environment, principally light/dark transitions, acted as synchronising signals or ***Zeitgebers*** (termed '*Zeitgeber*' from the German 'time giver') that **entrain** the endogenous clocks to 24 hours rhythms in the external environment (Brandstatter *et al.* 2001b; Brandstatter 2003; Wikelski *et al.* 2008).

Soon, chronobiologists detected that the circadian mechanism was also involved in measuring changes in day length (Dawson 2002; Ikegami & Yoshimura 2012), therefore its elements are relevant in the study of seasonal rhythms. It is still unclear, though – there are theoretical hypotheses –how this connection works, and which are the elements that connect these two clocks of different amplitude, but it is quite certain that they, somehow, 'talk' to each other (Pegoraro *et al.* 2014).

Circannual rhythms are found in many long-lived organisms, and have been best studied in mammals and birds (Bradshaw & Holzapfel 2008; Wikelski *et al.* 2008; Helm *et al.* 2009). Pioneering work in the early to mid- 20th century established that, at least at temperate latitudes, the primary driver of seasonality was the relative length of the light

and dark portions of the 24 h / day summarized under the term 'photoperiod'. William Rowan was the first to prove experimentally the phenomenon of photoperiodism in birds. Rowan kept two migratory bird species outdoors in aviaries and exposed them to longer days in winter using artificial light. He observed that, despite it being winter, the birds developed their gonads prematurely and showed behaviour typical of breeding season (e.g. they sung and showed migratory restlessness – also called with the German word '***Zugunruhe***' –) (Rowan 1932). A long-distance migrant, the dark-eyed junco (*Junco hyemalis*) was held in constant light for 3 years and still exhibited rhythms of *Zugunruhe* (Holberton & Able 1992). Gwinner kept willow warblers (*Phylloscopus trochilus*) under constant and equatorial photoperiods (12h light – 12 h dark, LD 12:12 h) for many years and individuals showed *Zugunruhe* with period lengths of approximately a year (Gwinner 1986).

It is now accepted that day length photoperiod, or better said, the shape of the annual change in photoperiod, is the primary external clue interpreted by the endogenous clock for measuring **seasonality** in most organisms living at temperate latitudes (Leder *et al.* 2006; Dawson 2007; Coppack *et al.* 2008; Cornelius *et al.* 2013; Helm *et al.* 2013). In other words, photoperiod varies according to the time of the year, therefore, the light signal received in the **Suprachiasmatic Nucleus** (SNC) also varies (Schaap *et al.* 2003; Hazlerigg *et al.* 2005; Johnston *et al.* 2005; Ikegami & Yoshimura 2012) and differences as small as in the tropics can be detected by birds (Hau *et al.* 2008; Goymann *et al.* 2012).

How is time measured?

Biological clocks are conceptualised to possess three components: 1) a sensory input from the environment as signal and **synchronizer**; 2) an **oscillator pacemaker** that measure day length; and 3) an output signal (neural or humoral) that regulates physiological and behavioural aspects (Dunlap *et al.* 2004; Kumar *et al.* 2004; Yoshimura 2010). Moreover, these independent clocks interact with each other and work as the **Centralized Clocking System** (CCS) (Brandstatter 2003; Kumar *et al.* 2004; Helm & Davidowitz 2012), controlling peripheral oscillatory organs (Hirota & Fukada 2004; Kumar *et al.* 2010; van der Veen *et al.* 2017). Peripheral organs (i.e. liver, heart, liver, kidney) are also able to produce oscillations (Abraham *et al.* 2003), but they are not self-sustainable, thus they depend on the main oscillatory organs to maintain their rhythms (Reppert & Weaver 2002; Kumar *et al.* 2010). The difference between the master and

peripheral oscillatory organs at the molecular level is unknown, but it is believed that there should be a difference in clock protein levels and/or kinetics and not because of the expression of a unique-master/peripheral element (Zeman & Herichova 2011). They may even use different external cues for synchronization (Akashi & Nishida 2000).

Unexpectedly, birds' clock organization seems to be more complex than that of mammals (Brandstatter *et al.* 2001a; Gwinner & Brandstatter 2001; Abraham *et al.* 2002), which is believed to reflect the common ancestor of mammals, but not birds and reptiles, having passed through some sort of nocturnal phase (Gerkema *et al.* 2013). In birds there are three major self-sustained oscillators (Abraham *et al.* 2003) (Figure 1.1): the pineal gland, the retina, and the SNC (Dawson *et al.* 2001; Bell-Pedersen *et al.* 2005). The pineal

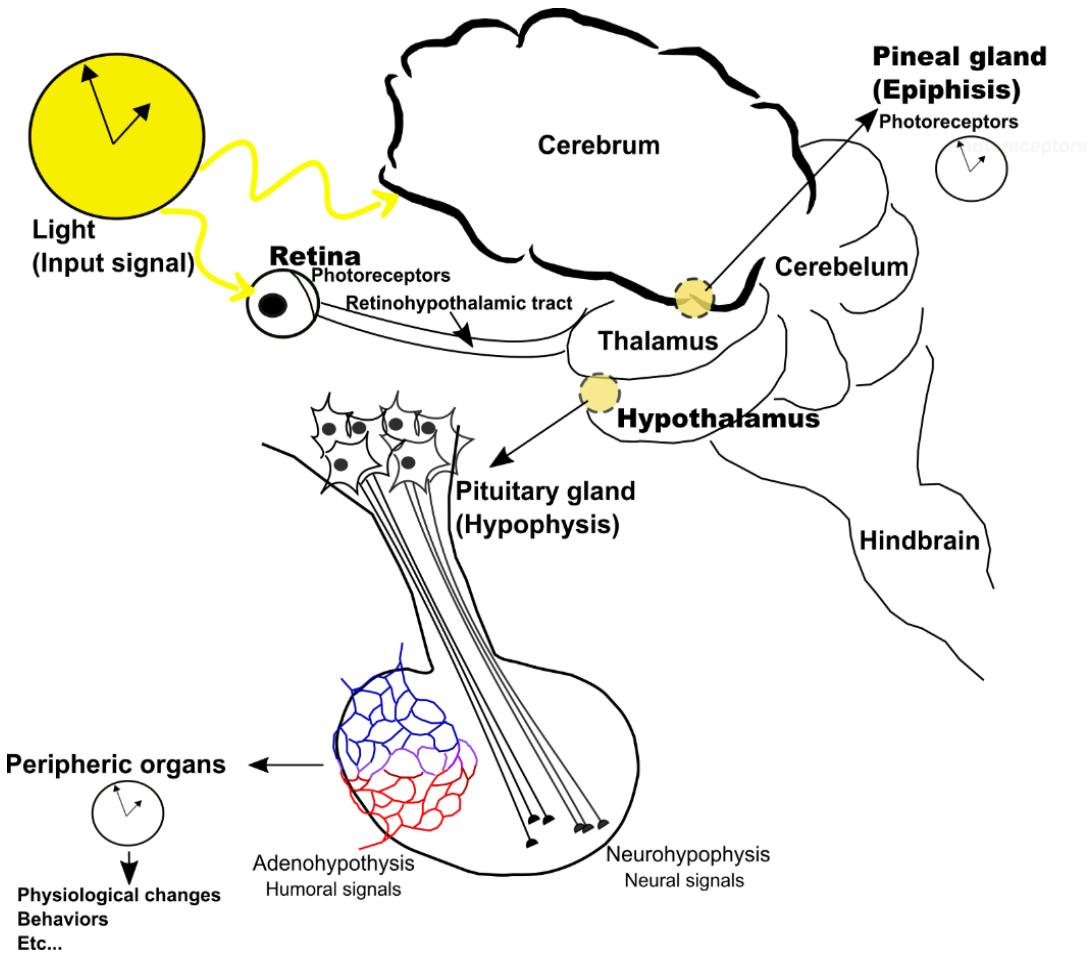


Figure 1.1: Schematic diagram of the general physiological elements involved in migratory timing in birds. This figure was modified from Albrecht 2012.

gland is believed to be the main clock system in birds (Okano & Fukada 2001; Brandstatter 2003; Faluhelyi *et al.* 2004; Nagy & Csernus 2007; Racz *et al.* 2008; Yoshimura 2010). The complex set of interconnected elements consisting of the hypothalamus, the pituitary gland and the adrenal gland is known as the **Hypothalamic-Pituitary-Adrenal axis** (HPA axis) and it is a major part of the neuroendocrine system playing a fundamental role in the physiology of rhythmic processes (Scheiermann *et al.* 2013).

Other important elements are those connecting these organs (i.e. hormones and neurotransmitters). Melatonin is a hormone secreted in the retina and pineal gland and it transmits the light signal (photoperiod information) to the hypothalamic oscillator and maintains the coordination between these elements of the brain (Dawson *et al.* 2001). Interestingly, reductions in the level of nocturnal melatonin appear to be associated with migratory restlessness (or '*Zugunruhe*') (Hau & Gwinner 1994; Gwinner 1996; Gwinner & Helm 2003; Ikegami & Yoshimura 2012). The 'internal clock' can be visualized from a physiological perspective, as described in the previous paragraphs, however the indivisible elements that compose the 'internal clock' are the cells – not the organs or tissues –. Therefore, organisms are 'composed' of millions of 'clocks' that work in synchrony (Herzog 2007). What does a cell 'clock' look like? The combined application of classical and molecular genetics, particularly in the model organisms *Drosophila melanogaster* (Common fruit fly), *Neurospora crassa* (Fungus), and *Mus musculus* (Mouse), facilitated the identification of genes that encode the central elements of the 'Core Circadian Oscillator' (CCO) that forms the heart of the circadian clock across a very wide range of taxa (Goldbeter 1995; Nakashima & Onai 1996; Lewis *et al.* 1997; Zhang *et al.* 2012). In essence, the CCO is considered to consist in its most basic elements as an interconnected transcription/translation loop with elements causing it to oscillate with a period of approximately 24 h (Dunlap 1999; Albrecht 2012) (Figure 1.2). While the CCO elements are all contained within single vertebrate cells (Dunlap 1999; Reppert & Weaver 2002; Hazlerigg & Lincoln 2011; Albrecht 2012; Steinmeyer *et al.* 2012), the overt physiological and behavioural rhythms, along with the measurement of photoperiod, arise from the interaction of multiple 'clock' tissues (Abraham *et al.* 2002). From the perspective of the determination of genetic variants associated to migratory timing behaviour, genetic variation in any of the elements of the CCO might potentially influence departure timing.

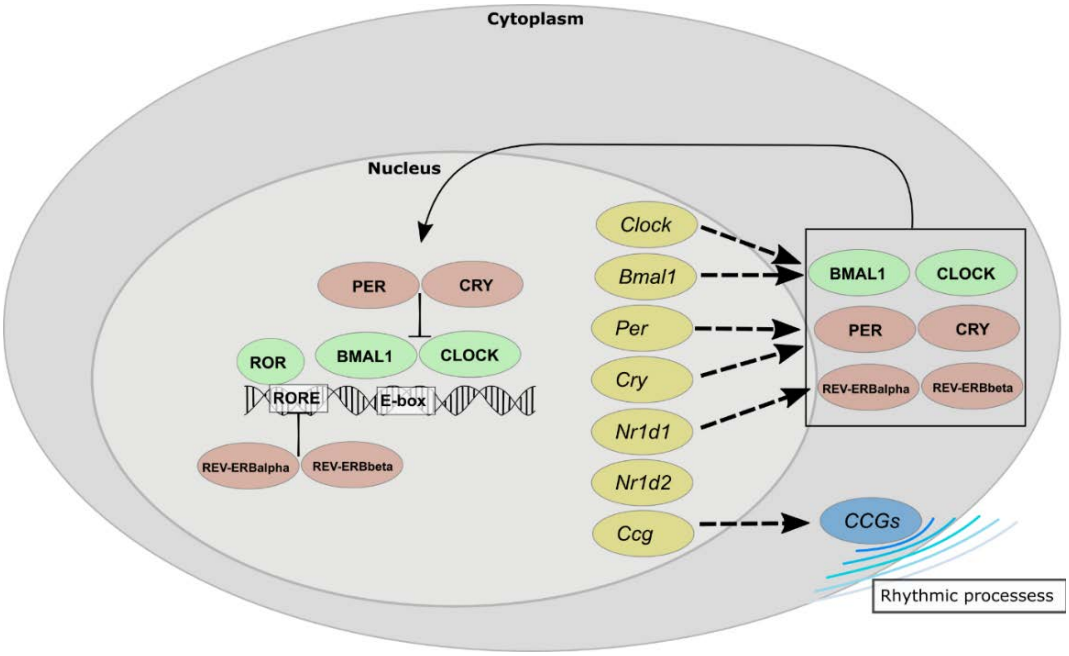


Figure 1.2: Simplified diagram of the Core Circadian Oscillator elements (i.e. *Clock*, *Bmal1*, *Per*) involved in the intracellular transcriptional feedback loops. Modified from (Scheiermann *et al.* 2013).

Understanding the genetics of migration of birds in the Era of genomics.

Most of what we know so far about the genetics of migration timing in birds is based on classic crossbreeding and selection experiments (Berthold & Helbig 1992), **quantitative genetic analyses** (reviewed in Pulido & Berthold (2003) and experiments with captive birds (Eikenaar *et al.* 2014; Schmaljohann *et al.* 2015). At some point, it was quite clear and accepted that migratory traits have a strong genetic component (Pulido 2007a; Liedvogel *et al.* 2011), with **heritabilities**, for example, for timing of autumn migration being between 0.34 and 0.45 in blackcaps (*Sylvia atricapilla*) (Pulido *et al.* 2001). This gave rise to studies trying to give the first insights into the dissection of the genetics of migratory behaviour (i.e. timing) in order to 'give name' to its components. These initial studies suggested **additive genetic variation** of migratory traits, which seems to be inherited in the form of a '**migratory gene package'** (Liedvogel *et al.* 2011; Lundberg *et al.* 2017). In addition, taking into account the large variety of migratory strategies found in nature, scientists were not only interested in identifying the elements of the 'migratory gene package' but also to identify the causal genetic elements and mechanisms that explain the phenotypic diversity observed in the wild. This interest goes

beyond of the purely exciting feeling of being part of the team discovering solid answers about how the common internal clock works, which resulted in a Nobel prize to Jeffrey C. Hall, Michael Rosbash and Michael W. Young in 2017. Discovering the elements governing timing traits, such as migratory timing, would have a huge impact for evolutionists and population geneticists, who are interested in how these elements could be involved in population gene flow characteristics and its potential to drive evolutionary processes such as microevolution or speciation (Husby *et al.* 2015; Tarka *et al.* 2015; Knief *et al.* 2017). Most of the studies addressing these questions use genetic or genomic approaches, principally, **transcriptomic (i.e. Microarray, RNA-sequencing), Candidate Gene Approaches (CGA)** and **trait mapping** (i.e. from **Restriction site Associated DNA markers (RADtag)**, **Genotyping By Sequencing (GBS) or Capture Sequencing (Capture-Seq),** being **Genome Wide Association Studies (GWAS)** one very complete form of analysis.

Gene expression profile studies (or transcriptomic) are a type of approach that try to give insight on the elements determining certain events (i.e. circannual rhythms) by measuring gene expression at different life stages or between species or populations showing different behaviours. For example, based on the fact that we know that during the process of preparation for migration the brain undergoes physiological changes (Pravosudov *et al.* 2006), Jones *et al.* (2008) investigated the genetic expression profile in the brain (in the telencephalon, to be more precise) of a migratory bird, the white-crowned sparrow (*Zonotrichia leucophyris gambelii*). They found that a few genes related to stress and energy demands (i.e. GLUT1 and Heat Shock proteins) showed significantly different expressions profiles when comparing between migration and non-migration periods. Boss *et al.* (2016) compared the gene expression profile in the whole brain of two willow warbler subspecies (i.e. *Phylloscopus trochilus trochilus* and *Phylloscopus trochilus acredula*) at two life-stages: breeding and wintering season. They were expecting more differences in gene expression when comparing the subspecies between seasons than subspecies, and indeed, they found this. A second interesting finding was that these two subspecies, which were known to have different migratory strategies, showed a few differences in their gene expression profiles, and this could give insight on the genetic basis of migratory traits in the wild. One of these genes was *adcyap1*, the gene that encodes for the protein PACAP, which had been previously reported to be associated with restlessness in some bird species (Mueller *et al.* 2011; Peterson *et al.* 2015). This

study also suggested that maybe only a few genes are responsible for the different migration strategies between willow warbler subspecies. Similar findings were obtained in a study in European blackbirds (*Turdus merula*) (Franchini *et al.* 2017), which compared the transcriptomic profile of resident and migrant birds in order to identify possible elements determining migratory behaviour. Although scientists are making great progress in improving the techniques and the experiment designs, which is leading to the identification of potential candidates, there is a lack of understanding about the functionality of those identified candidates (Liedvogel & Lundberg 2014). Consequently, there is still the need for corroboration of the already identified loci associated to migratory traits, as well as new studies on other systems that might identify new candidates. Transcriptomic studies require the sampling of the tissues (or organs) related to the behaviour in question, and consequently the sacrifice of individuals. This is, in most of the cases, ethically unaffordable for most of studies on wild species. This is why, transcriptomic analyses has been used when studying model organisms or with captive organisms. These studies, however, nurture CGA studies, in which prior knowledge of gene function is used to select a gene (or genes) that potentially relate to the trait of study.

The most frequently used approach to identify the elements of the genetic architecture of migratory traits is probably the CGA, and the genes most studied are *clock* and *adcyap1* (Mueller *et al.* 2011; Mettler *et al.* 2015; Bazzi *et al.* 2016c; Contina *et al.* 2018). A variety of systems and approaches have been studied, testing for relationship between these genes and migratory traits. Such as, comparative studies of the genetic make-up of non-migratory versus migratory species or populations (Lugo Ramos *et al.* 2017) and between populations with different migratory strategies (Bazzi *et al.* 2016a). The gene *clock* in particular has been widely studied since a finding that the distribution of the frequencies of a polymorphic coding region – specifically on the number of glutamines (coded as Q) – was associated with a latitudinal cline in the migratory blue tits (*Cyanistes caeruleus*) (Johnsen *et al.* 2007). This finding suggested a potential role of this gene in the adaptation to photoperiodic regimes at different latitudes. This study gave rise to new studies that explored possible links between *clock* and other migratory timing traits (e.g. Bazzi *et al.* 2015; Peterson *et al.* 2015; Saino *et al.* 2015; Contina *et al.* 2018), as well as in non-migratory but timing-related traits (i.e. breeding time, e.g. Liedvogel *et al.* 2009; Bourret & Garant 2015). So far, there is a lack of conclusive agreement about what the role of *clock* is in determining migratory timing traits in birds, and the view

nowadays is that the genetic basis of migratory traits (i.e. timing) probably differs between taxa (Dor *et al.* 2011a; Dor *et al.* 2011b).

There are other migratory species that, through economic interest or ease of field studies, offer more possibilities for in-depth studies of migratory traits. For example, there is an extensive literature – at least larger than that of wild birds – on salmon (Hecht *et al.* 2012; Hess *et al.* 2016) and butterflies (Zhan *et al.* 2014). These studies can give some insight on what to expect from the genetics of migration in birds. Hess *et al.* (2016) used both model- and non-model-based methodologies to look for genes associated to individual's migratory timing phenotypes in salmon, and found evidence of this phenotype being largely explained by a few genes, two of them with significantly larger effects, showing that a combined approach could help the identification of genetic variants associated to traits.

The advent of high-throughput sequencing technology made whole genome assessment (i.e. GWAS) on the sequence accessible also for non-model organisms. This means that with this approach, not only candidate genes are available, but also the possibility to investigate the **genetic architecture** of several traits. There are several studies that show the first attempt to understand the 'big picture' – that is, the genetic architecture – of complex traits in birds, some specifically of migratory traits (e.g. Santure *et al.* 2013; Santure *et al.* 2015; Delmore *et al.* 2016; Lundberg *et al.* 2017). These studies point out that the genetic architecture of complex traits is likely to be governed by many genes (polygenic), as well as their interactions (**epistasis**) (Santure *et al.* 2013). Besides, new studies highlight the importance of including all levels of complexity into the study, such as the integration of environmental factors when studying genetic aspects in wild species (Santure *et al.* 2015; Gienapp *et al.* 2017).

Indeed, in the last years, there has been an increasing awareness of the importance of the integration of environmental factors when trying to identify the elements underlying migratory traits (Gienapp *et al.* 2017). This is because the behaviour is often modified by environmental factors (i.e. so that the expression of the internal clock is actually not directly measured) (Bauer *et al.* 2008; Helm *et al.* 2017). For example, Gienapp *et al.* (2017) tested the association of a number of genetic variants against egg-laying time in great tits (*Parus major*). Even though they did not find any significant result, the variance explained by the fixed effects in the model significantly increased

when genotype by environment interactions were included after adjusting for the increased model complexity. In this case, temperature is known to be the main environmental factor affecting breeding time and therefore, influencing egg-laying date. However, it is not always the case that one knows clearly which is/are the environmental factors that need to be taken into account. For example, for migratory timing the environmental factor triggering the departure for migration is believed to be the photoperiod, but which specific aspect of the photoperiod (i.e. daylength threshold, rate of change in daylength) is actually the driver of migratory time is not clear. Nevertheless, migratory timing behaviour has sufficient scientific support for having a strong genetic basis, therefore, not including environmental factors should not impede the detection of at least large-effect genes as shown in studies on other traits that did not take into account environmental factors (Delmore *et al.* 2016; Hess *et al.* 2016; Lundberg *et al.* 2017).

Unfortunately, genomic studies in non-model organisms, in which genetic resources (i.e. reference genome) are usually not available, represent a big obstacle and a limiting factor for new studies on the genetic architecture underlying complex traits in natural populations (Slate *et al.* 2010). Fortunately, this is changing with new genomes of non-model bird species being generated (e.g. Zhang *et al.* (2014), https://b10k.genomics.cn/). There are 94 model and non-model bird genomes public in the National Center for Biotechnology Information (NCBI) to date (February 2018), with the majority having been generated in the last 5 years. Another limiting factor for studying the genetics of migratory traits is the difficulty that goes with studying annual behaviours in the wild. The recording and characterization of behaviours in the wild are not straightforward, but is a task that requires an investment of time and money over multiple years, and continuously, if possible. Until now, there is not a single study in which well-stablished repeatable and solid measurements of departure date for migration over several years have been used to investigate the genetics underlying phenotypic variation in a wild species.

In this Era of fast technological and computational advances in which there is more and more interest in understanding biological clocks in the wild (Schwartz *et al.* 2017), there is a common horizon that is making these two fields, ecology and genomics, converge in one, and this thesis is an example of this.

Approaches to the study of genetics of migration.

Genomic approaches

There are two approaches/methodologies that are being extensively used as first step for identifying loci linked to **traits** (behaviour, morphology):

- Candidate Gene Approach (CGA) focuses on the most promising genes based on bibliographic research (i.e. theoretically and experimentally reported links between genes/proteins and the phenotype of interest) (Fitzpatrick *et al.* 2005). This approach has proven efficiency when the candidates have a well-described function linked to the phenotype, either trait or behaviour in natural populations (e.g. Fitzpatrick *et al.* 2005; Fidler *et al.* 2007; Ekblom & Galindo 2011; Mueller *et al.* 2013a; Poelstra *et al.* 2013).
- Genome-Wide Association Study (GWAS) is a powerful approach that allows to explore differences not only at **candidate gene regions** but also at undescribed/unknown regions of genomes. In this approach, several genomes are examined for genetic variants and compared between individuals with different phenotypes (i.e. chronotypes) using different metrics (i.e. ***Fst***). This approach can serve to prioritize the most promising SNPs associated to the trait in question, overall in cases where a limited number of genomes are available (Kwon & Goate 2000; Tabor *et al.* 2002; Zhu & Zhao 2007; Ekblom & Galindo 2011).

CGA, compared to a GWAS, has two advantages: lower cost and being focused on a set of putatively the most promising genes associated to the trait, which has been described as a great advantage for studies with limited number of samples (i.e. commonly in studies on natural populations). Both approaches need a preliminary assessment of the potential **confounder** by **population genetic** background. This is because having discovered a genetic variant associated to the trait in question under a scenario of population structure, this association may be a false positive, or in other words, this putative association might be a product of underlying population genetic background confounding. An *a priori* assessment on the population genetic background would serve to control this potential issue. Another important issue/characteristic associated with these kind of approaches is multiple testing. This is simply a statistical issue due to the number of independent tests that large-scale GWAS involves (Sham & Purcell 2014). Bonferroni

correction (a type of **Family Wise Error Rate**, FWER) is a too conservative approach to deal with this issue (i.e. this works better for less large numbers of independent tests), consequently, other approaches were developed, such as Benjamin-Hochberg correction or **False Discovery Rate** (FDR) (Benjamini & Hochberg 1995), which, in the context of genomics, is the most used.

Linking genes to behaviour.

The idea of this section is, first, to explain under which hypothesis the link of genes with behaviours can be made, and second, to summarize briefly the basic principles on which genomic analyses are based.

Variations in coding regions can potentially produce different proteins (i.e. non-synonymous/silent SNPs), which might lead to differences in its functionality, as documented mostly in humans and model organisms (Holmes *et al.* 2001; Hejjas *et al.* 2007; Våge *et al.* 2010; Murphy *et al.* 2014; Zhao *et al.* 2014). These phenotypic variations could also be a consequence of differential regulation of protein transcription (Whitfield *et al.* 2003; Murphy *et al.* 2014; Pfenning *et al.* 2014; Benowitz *et al.* 2017). Therefore, SNPs placed in coding and **flanking regions** of candidate genes are interesting targets. **Linkage Disequilibrium (LD)** (i.e. indicating 'blocks' of the genome that are physically close and, consequently, with reduced recombination rates within it, or, in other words, they are inherit as a 'block') is an important factor to keep in mind when doing association studies. Levels of LD are not easy to estimate and, moreover, differ along the genome (Stapley *et al.* 2010a; Kawakami *et al.* 2014) (i.e. between chromosomes and as a function of genetic distance), across taxa (Gaut & Long 2003; Backstrom *et al.* 2006; Li & Merila 2010) and even amongst populations (Goddard *et al.* 2000; Reich *et al.* 2001; Sawyer *et al.* 2005; Li & Merila 2010). A general estimate of LD in birds is 10–100 kb (Edwards & Dillon 2004; Li & Merila 2010; Stapley *et al.* 2010a), although see Poelstra *et al.* (2014) which shows lower estimates of LD of around 100 bp in European crows (*Corvus corone*).

Genomic analyses are based on **statistical models**, which are tested against a null model (i.e. non-existence of that model) for each of the candidate SNPs. There are three genetic models which are the most widely used for genomic analyses: additive, dominant and recessive (much less used: over-dominant model) (Figure 1.3). These models also

assume the distribution of effects of the loci tested for association, generally: 1) Gaussian or normal distribution, which assumes that all variants have an effect on the phenotype but with effect sizes normally distributed (**polygenic** modelling); or 2) not normal distributed where a relatively small proportion of all variants have an effect on the phenotype distribution (**mono/oligogenic**). These two models are used by the Linear Mixed Model (LMM) and Bayesian Variable Selection Regression Model (BVSR) which are the most common models implemented in software or pipelines for genomic analyses (i.e. PLINK, TASSEL).

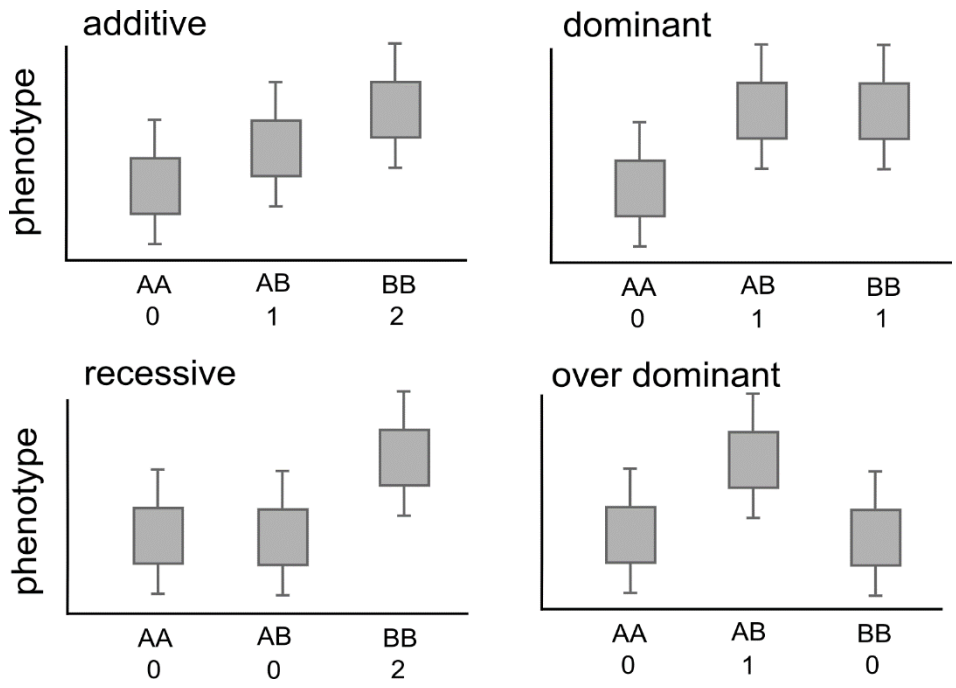


Figure 1.3: Genetic models used in genotype-phenotype association studies. Effects of the genotypes (AA, AB, BB) on a phenotype are assumed to be distributed as for an additive, dominant, recessive or over-dominant model. Numbers (0, 1, 2) correspond to modelled genotypes (Tsepilov *et al.* 2015).

Bar-tailed godwits: An excellent candidate model organism.

Bar-tailed godwits (Figure 1.4), *Limosa lapponica*, are a migratory shorebird with remarkable capacity for long-distance flight. The Alaskan-breeding subspecies *baueri* (godwits hereafter) is famous for making the longest non-stop endurance migratory flight in the animal kingdom, >11 500 km from Alaska to New Zealand (Gill *et al.* 2009). After 'wintering' in the southern summer, adults then make another massive flight, 10,000 km north to staging grounds in Asia, where they refuel before migrating back to Alaska to

breed (Battley *et al.* 2012). Unsurprisingly given the length of these flights, they also a reach extremely high fat loads (30–55%) before departing on these flights (Battley & Piersma 2005).

Figure 1.4: A fat male godwit ready for migration and carrying colour bands and a geolocator. Photo by Phil Battley.

Godwits are also notable for the consistency of their migration timing. While the population departs over a 5–6-week period, individual birds tend to migrate in the same week or even the same few days, year after year (Battley 2006; Conklin *et al.* 2013). Consequently, New Zealand godwits have one of the highest repeatabilities of migration timing documented (Conklin *et al.* 2013). The reason for this pattern of large population variation but low individual variation in timing relates largely to geographic variation on the Alaskan breeding grounds (Conklin *et al.* 2010) – birds breeding in southern Alaska tend to migrate early in the season, while birds breeding in northern Alaska tend to migrate much later –, a pattern that becomes even stronger when birds leave their staging grounds in Asia to fly to Alaska (Figure 1.5).

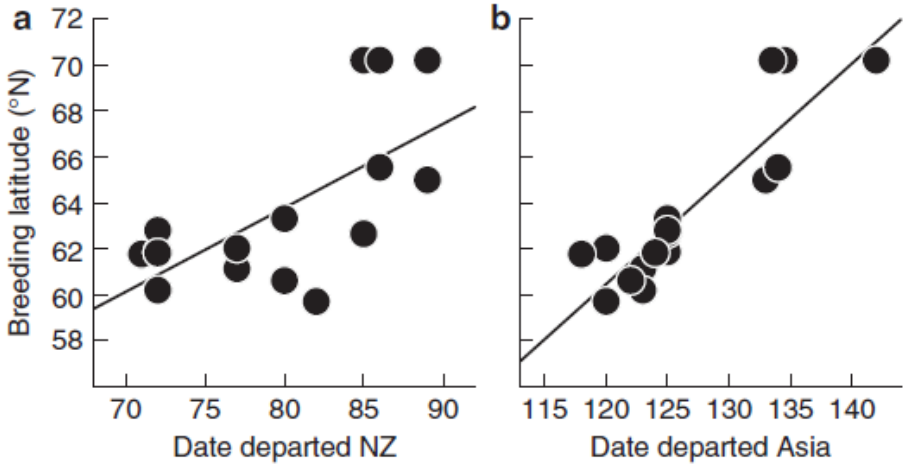


Figure 1.5: The relationship between migration timing (a) from New Zealand, (b) from Asia, and breeding latitude as shown in Conklin et al. (2010).

Juvenile godwits only four-months old perform their first migration flight southwards crossing the Pacific Ocean to New Zealand or Australia. Once they arrive on the non-breeding grounds in New Zealand, they spend between two and four years exploring around these islands (Battley 2007; Battley *et al.* 2011), before returning to the breeding sites. As previously described, photoperiod (daylength) likely plays an important role in the godwits' annual cycle since it is believed that it is the main spatial and temporal indicator for natural populations (Gwinner 1996, 2003). Under this scenario, it is interesting to see how godwits are exposed to drastic changes of photoperiod due to the annual migration journey between their breeding sites in Alaska (northern hemisphere) and New Zealand (southern hemisphere) (Figure 1.6).

Because godwits live in open habitats, it has proven possible to determine migration dates of individuals in New Zealand simply by recording the daily presence of colour-marked birds (see Figure 1.4). For the timing of migration overseas, geolocators (light-sensitive loggers that may also record wetness) can allow other flights to be documented remotely (Conklin *et al.* 2010; Conklin *et al.* 2013; Battley & Conklin 2017). Godwits therefore have a number of factors that make them an ideal candidate for studying the control of migration timing: 1) Blood samples are available of many individuals with behavioural data; 2) Phenotype quantification is vital for the successful of genotype-phenotype association studies, and godwits have a very well-characterized

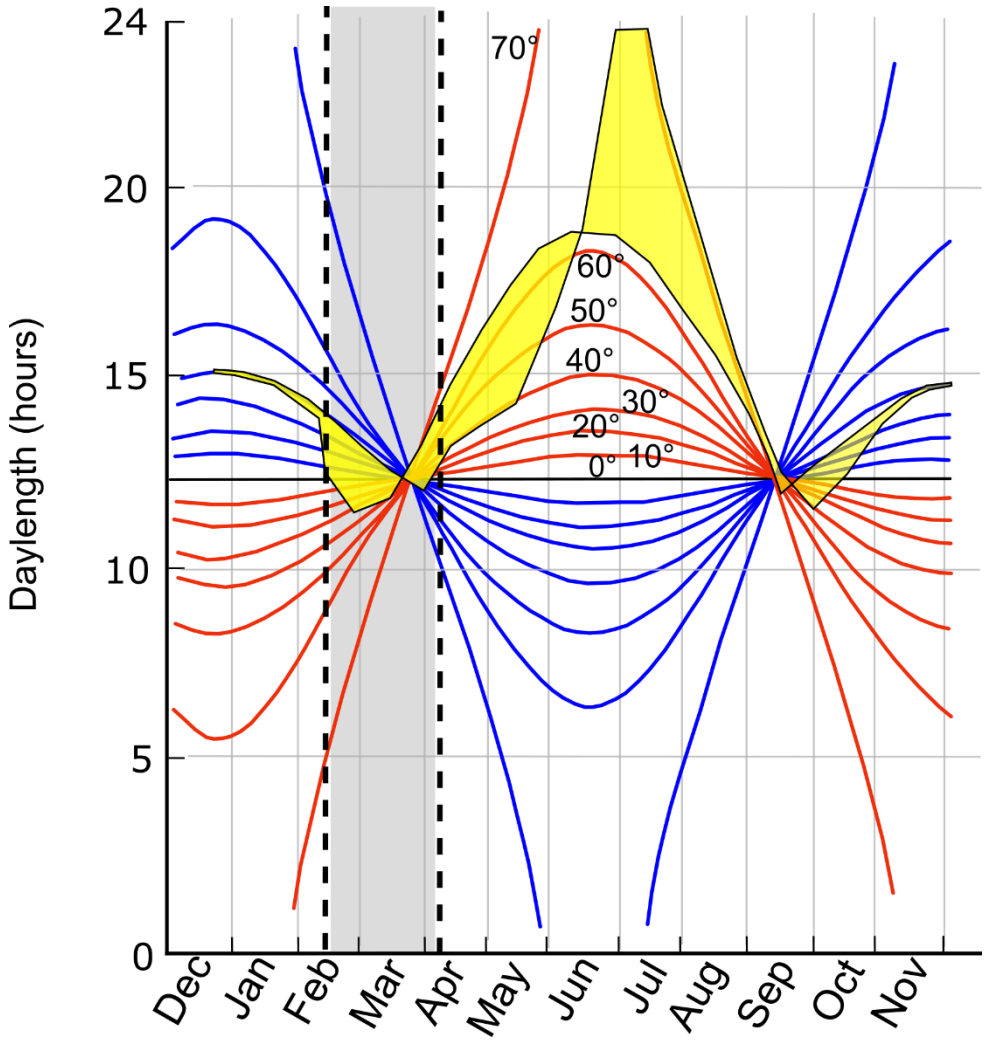


Figure 1.6: Daylengths at different latitudes in the northern hemisphere (red lines) and in the southern hemisphere (blue lines). Daylengths New Zealand godwits are exposed to along their annual cycle is indicated in yellow (varying with non-breeding latitude). Grey zone within dashed lines indicates the period in which godwits initiate their migration to the breeding sites.

individual behavioural data record of up to 10 years; 3) Individual behaviour has a solid and proven repeatability (Conklin *et al.* 2013) of individual migratory time; 4) Godwits are very faithful to their over-wintering sites (Conklin *et al.* 2011), therefore facilitating the study of individual's migratory behaviour through geolocators, which gave additional precision to some migratory departure date records; 5) sufficient funds were obtained to generate a reference genome.

Key questions and chapters overview

Before this thesis was initiated, individual departure times for migration of godwits had been extensively studied over three years (Conklin 2011) giving rise to an intriguing question: Is the onset of godwit migration governed by a specific genetic element of the 'endogenous clock'? This question could not be addressed without a reference genome and a well-established knowledge on the genetic structure of New Zealand godwits. That is how, in short, this thesis was structured so that Chapter 2 deals with the population structure assessment, Chapter 3 deals with a specific genetic polymorphism, and Chapter 4 addresses questions regarding genetics of migratory departure time at a broader scale and with larger sample of genes and chronophenotypes.

In summary, in Chapter 2 I assessed the genetic structure of the New Zealand population of godwits, using a set of microsatellites (neutral markers) in order to be able to answer the following questions:

- Is there any sign of population structure within New Zealand bar-tailed godwits?
- Are different individual departure dates, from the overwintering sites in New Zealand and from the stop-over in Asia, associated to a potential population genetic structuring?
- Does genetic background – if any – need to be taken into account for the genotype-phenotype association study in godwits performed in Chapter 4?

In Chapter 3 I focused on one particular polymorphic genetic locus (the *Clock*polyQ), about which there is active debate about whether it is – or is not – actually involved in determining chronotypes. In this chapter I addressed this specific question: Can *Clock*polyQ allele frequencies predict migratory departure date of godwits? I also provide an extensive overview of all the studies to date on *Clock*polyQ, particularly on birds' timing-related traits.

Chapter 4 provides the core analyses of my thesis. This chapter deals with the phenotype-genotype association study in which I used a sequenced and assembled genome of the godwit (done by two bioinformaticians collaborators: Francisco Prosdocimi, Nicholas Lima), partial sequencing of 19 additional genomes and 140 candidate genes regions based on bibliography, and tested for association between 3,412

SNPs placed within candidate gene regions and the individual departure date of migration. Specifically, I address these questions:

- What is the genetic architecture of migratory departure date in godwits?
- Are there SNPs (genes) that explain a significant proportion of the chronophenotypes?

Chapter 5 is a synthesis of the main findings of this thesis, its contributions to better understand the genetic basis of migratory behaviour in birds, as well as suggestions and unanswered questions that remain for future studies.

Overview of the techniques and technologies used in this thesis.

Genetic techniques

- DNA fragment analysis by **capillary electrophoresis** (Figure 1.7): Specific to microsatellites **tailed-primers** labelled with **fluorescent dyes** are used for **amplification** using a **thermal cycler**. There are four dyes which emit light

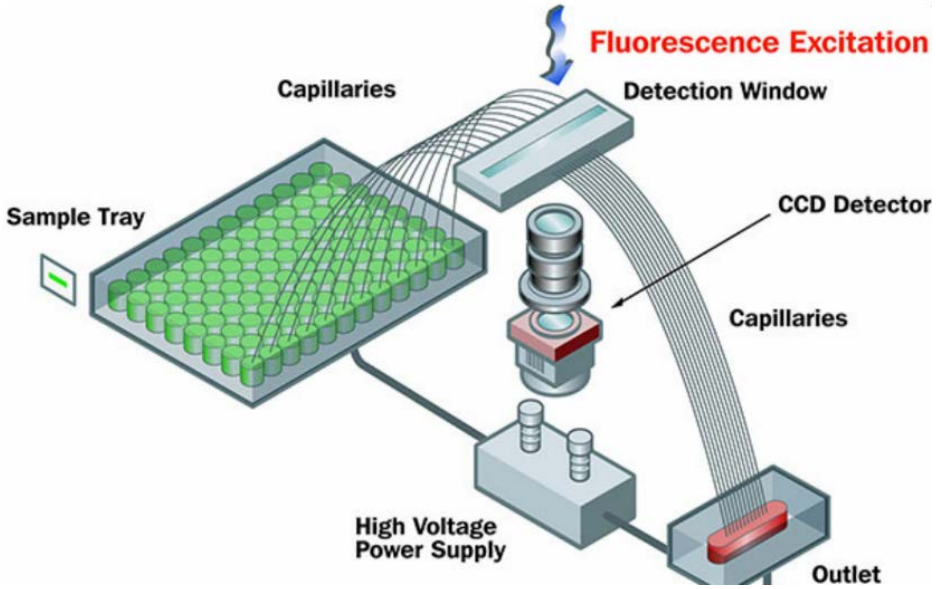


Figure 1.7: Schematic representation of DNA fragment analysis technology. Taken from http://www.labgene.ch/gene-mutations-screening/425-fragment-analyzer-crispr.html.

at different wavelengths when excited by a laser (i.e. 6-FAM™ for blue, VIC® for green or NED™ for yellow, NED™ for red) allowing microsatellites to be pooled by 4 for the actual capillary electrophoresis step. This capillary – or Genetic Analyzer instrument – consists of a very thin channel filled with a matrix (gel) and exposed to a current of high voltage. When the sample is yield (together with a size standard, LIZ®), the fluorescently labelled fragments of DNA (i.e. microsatellites and size standard fragments) are separated by size and are directed to a laser beam. The order in which microsatellites fragments cross the laser beam with respect to the size standards serves to estimate its sizes. The fluorescent dye is detected by a camera and recorded by a computer. The same technique was used for the

genotyping clock poly-Q, with the sole variation that we designed *a priori* standard sizes specific to the poly-Q **polymorphisms** present in godwits.

A **cloning** procedure and **sequencing** (**second-generation sequencing**), as well as a DNA fragment size genotyping were used in Chapter 3.

- Cloning/ bacteria transformation by heat shock (Figure 1.8): Polymorphic fragments of interest (i.e. ***Clock* poly-Q** fragments) are first identified using

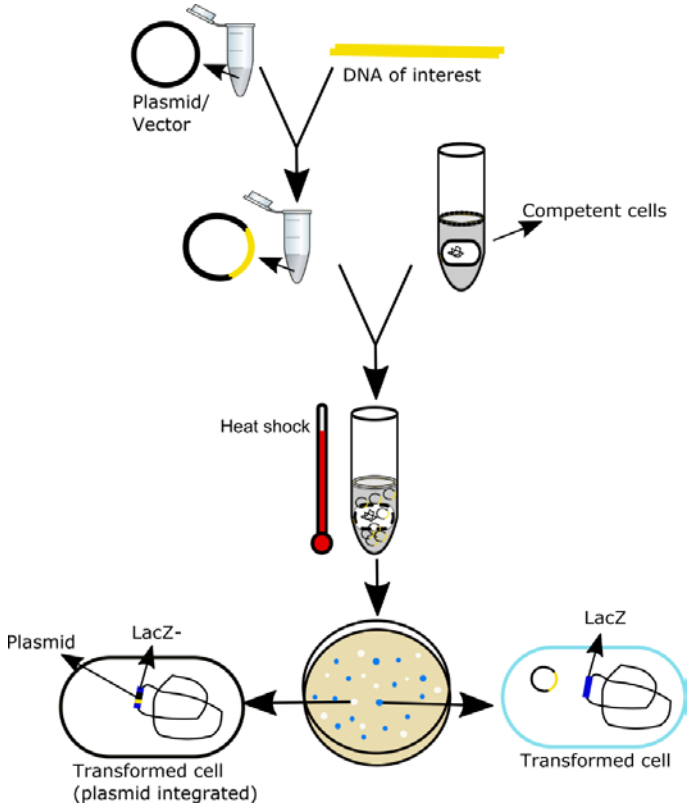


Figure 1.8: Schematic representation of cloning or bacteria transformation technique.

DNA fragment analyses and subsequently amplified by thermal cycle and extracted/ recovered by cutting the specific bands from the **electrophoresis gel**. Amplified DNA need to be ligated into a plasmid (or vector), which is achieved by a thermal procedure. This plasmid contains a gene/protein that allows for resistance to a specific antibiotic. Transformation of competent cells (i.e. *Escherichia coli*) occurs by 'heat chock'. Briefly, the rapid change of temperature allows the entry of the plasmid, which carries the DNA fragment of interest and the antibiotic-resistant gene, via pores inside the cells. These cells are designed to integrate foreign fragments of DNA with its own

DNA. Specifically, the foreign DNA is inserted in a way that disrupts a gene (β-galactosidase) that encodes for a protein whose product has an intense blue colour. This is used to select successfully transformed cells (white colonies) which are then isolated prior to be sequenced.

- Second-generation sequencing (Heather & Chain 2016) (Figure 1.9): Refers to a number of sequencing technologies (i.e. Illumina/Solexa, 454 sequencing, Ion Torrent, SOLiD sequencing) that revolutionised the field of genomics primarily for making sequencing much more affordable and faster than the previous technology (Sanger-sequencing) (van Dijk *et al.* 2014). Illumina/Solexa sequencing is probably the most used (and it was used for this project). Briefly, this technology consists of three steps: amplification, sequencing and analysis. DNA is first fragmented randomly and **adaptors** are

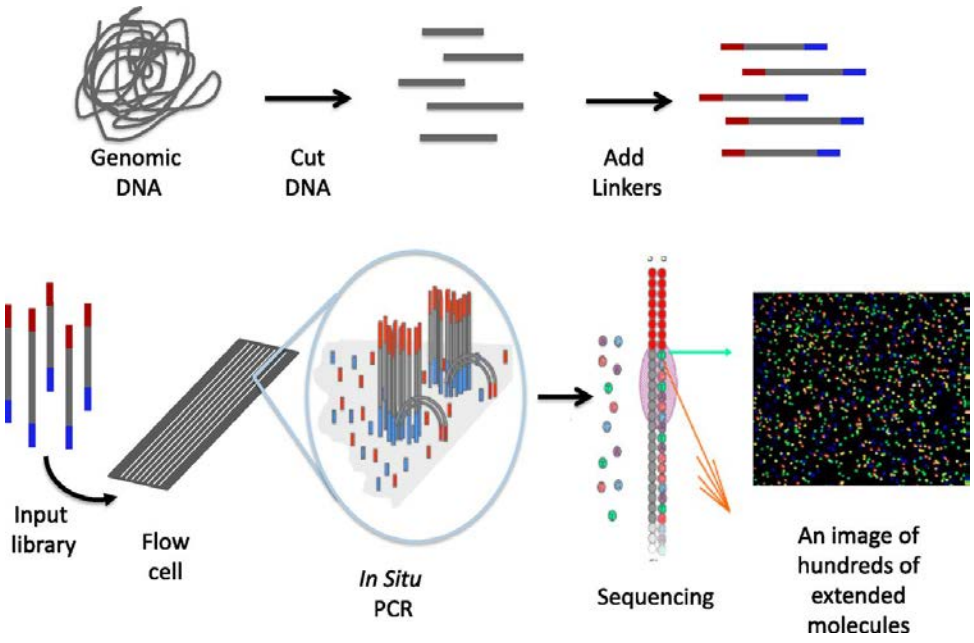


Figure 1.9: Schematic explanation of second generation sequencing technology.

added. Then this DNA is amplified by PCR (thermal cycle). A DNA sample is loaded into a chip in which each DNA molecule attaches, via its adaptor, to a primer which is attached to the base of the chip. Then the sequencing step begins: nucleotides and **polymerases** are added and amplification occurs. The 'magic' of these technologies reside in the nucleotides. They have a block that forces the polymerase to add only one nucleotide at a time, as well as a fluorescent tag which gives the signal of which nucleotide the polymerase has

just added. There is a camera that records all this fluorescent information and the computer decodes these fluorescent colours into nucleotides.

Finally, for Chapter 4, ***de novo* genome sequencing** (**library** preparation + next generation sequencing) and **Single Nucleotide Polymorphism** Capture Sequencing (SNPs Capture-Seq) were the molecular techniques used.

- Whole-genome sequencing (Figure 1.10): For a whole-genome sequencing it is necessary first to build a library. For the purpose of sequencing a *de novo* genome both **paired-end** and **mate paired-end** libraries are needed. A paired-end library consists in two reads from which we know the distance between them, therefore it facilitates the **assembly** of the genome. Typically, the distance between paired-end reads does not exceed 500 **pair bases** (bp). A mate paired-end library is the same but the distance between the reads is much

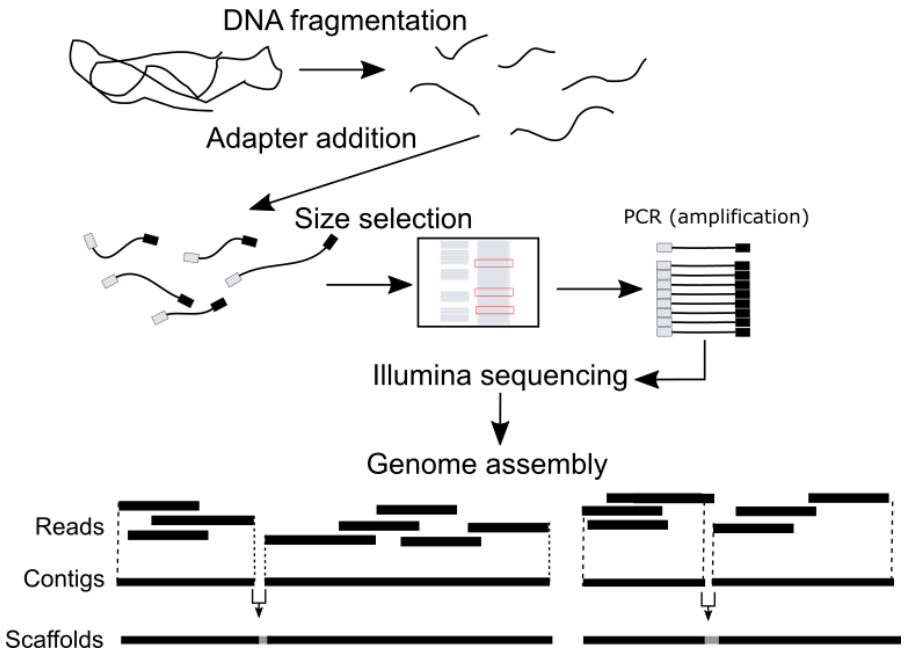


Figure 1.10: Schematic explanation of the steps for genome generation. Modified from (van Dijk *et al.* 2014).

longer, therefore generally the sequence between these two contains gaps, but again, knowing the distance between these reads helps for the assembly of the genome. The result of sequencing these libraries are reads, which are then assembled into larger sequences (**contigs**) using bioinformatic pipelines based on algorithms that, in short, look for overlapping zones between reads. These contigs are then assembled into larger sequences (**scaffolds**). Once a first

version of a genome is produced, more sequencing efforts can be done in order to improve the quality, the assembly and the sequence length of the genome by preparing more paired-end libraries and contracting more lanes of Illumina.

- SNP genotyping/capture-sequencing (Figure 1.11): This technology allows to genotype target SNPs. First steps consist in **probe** design and synthesis, library preparation (i.e. samples are **barcoded**) and **enrichment** of the targeting regions. The beauty of this technology lies on the customized and barcoded probes which are specifically designed to target thousands loci allowing for SNPs characterization and its correspondent identification to the sample it comes from. Finally, targeting regions flanked by the probes are sequenced and targeted SNPs (alleles) are identified per individual.

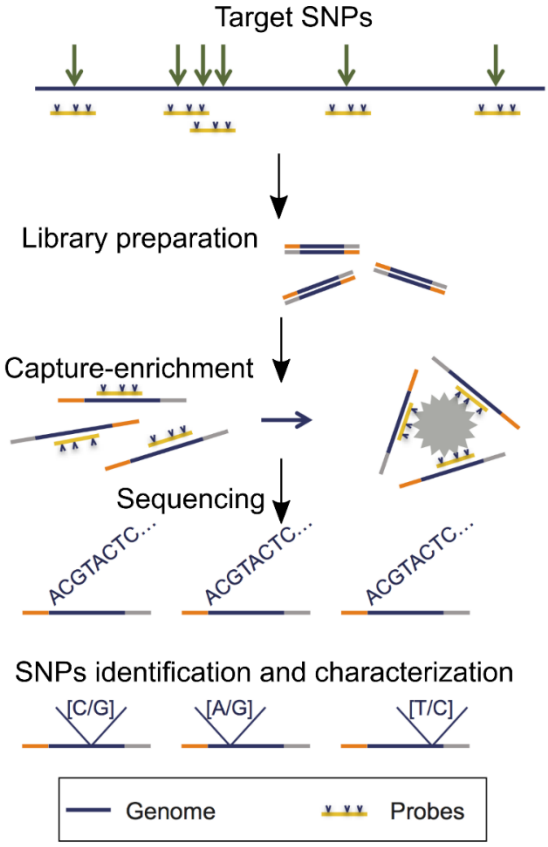


Figure 1.11: Schematic representation and steps of Capture-sequencing technology. Modified from a figure taken from RAPiD Genomics website.

Glossary

Adaptor • In genetic engineering is a short, chemically synthesized, single-stranded or double-stranded oligonucleotide that can be ligated to the ends of other DNA or RNA molecules (Wikipedia).

Additive genetic variance • The component of genetic variance due to the additive effects of alleles segregating in the population (Risch 2000).

Allele • One version of a gene at a given locus.

Amplification • The multiplication of DNA fragments typically achieved by thermal cycling (PCR).

Barcode • Short-sequence identifier for individuals labelling (barcoding) of sequencing libraries (Ekblom & Wolf 2014).

Base pair (bp) • Two nitrogen-containing basis pair together between double-stranded DNA. They form the building blocks of the DNA double helix (Wikipedia).

Candidate gene • Genes to be known to be involved in a pathway affecting a phenotype in some organisms and therefore are used in genotype-phenotype association studies of other organisms (species or populations) (Stapley *et al.* 2010b).

Candidate Gene Approach (CGA) • Candidate genes influence the expression of one focal phenotype in one or several (model) organisms and are thus potentially influencing a similar phenotype in other organisms. Candidate trait loci are evaluated based on sequence differences (SNPs, insertions or deletions) between individuals with different phenotypes (Liedvogel *et al.* 2011).

Capillary electrophoresis • Electrokinetic separation method performed in submillimetre diameter capillaries and micro- and nanofluidic channels (Wikipedia).

Capture Sequencing (Capture-Seq) • Method of genotyping consisting of capturing and sequencing specific, targeted regions of the genome. Taken from RAPiD genomics website (http://www.rapid-genomics.com/technology/).

Centralized Clocking System • Refers to the ensemble of areas in the brain that centralize the control of biological rhythms of an organism.

Clock poly-Q or ClockpolyQ • Region of the gene *clock* situated which transcribe for the carboxyl-terminal polyglutamine repeat motif, which regulates the transcription-activating potential of the protein CLOCK (Darlington *et al.* 1998).

Cloning or Transformation • Experimental method in molecular biology that are used to assemble recombinant DNA molecules and to direct their replication within host organisms. The use of the word cloning refers to the fact that the method involves the replication of one molecule to produce population of cells with identical DNA molecules (Wikipedia).

Confounder • Variable or factor that influences both the dependent variable and independent variable causing spurious associations (Wikipedia).

Contig • A contiguous linear stretch of DNA or RNA consensus sequence constructed from a number of smaller, partially overlapping sequence fragments (reads) (Ekblom & Wolf 2014).

***De novo* genome sequencing** • The generation, for the first time, of the genome sequence of an organism.

Electrophoresis gel • A method to analyse the size of DNA fragments based on an electric field and a gel (matrix).

Endogenous clock • A rhythm capable of self-sustained oscillations, generated by living organisms without need for external rhythmic input (Schwartz *et al.* 2017).

Enrichment • The targeted sequencing of specific regions of the genome using next-generation sequencing (i.e. SNPs, exons) (Stapley *et al.* 2010b).

Entrainment • The process of synchronization of the clock's oscillation to the *Zeitgeber* (usually the environmental day-night cycle) by resetting clock speed and/or phase (Schwartz *et al.* 2017).

Epistasis • Genetic variance due to non-additive effect of alleles at distinct loci (Risch 2000). The effect of one gene (or locus) is dependent on the presence of more modifier genes (Wikipedia).

False Discovery Rate (FDR) • The proportion of incorrectly rejected hypotheses relative to the total number of rejected hypotheses (Benjamini & Hochberg 1995).

Familywise Error Rate (FWER) • The probability of making one or more false discoveries (type I errors) among all the hypotheses tested.

Flanking region • DNA sequences extending on either side (i.e. 3' or 5') of a specific locus or gene. Promoters (i.e. regions of DNA that initiates transcriptions of a particular gene of about 100-1000 base pairs long) are usually located in the flanking region, as well as enhancer and other regulatory elements.

Fluorescent dye • Non-protein molecules that absorb light and re-emit a longer wavelength often used to label biomolecules (i.e. DNA).

Free running rhythms • The endogenous rhythm exhibited under constant conditions, characterized by its period length and amplitude (Schwartz *et al.* 2017).

***Fst* or fixation index** • Measure of population differentiation ('genetic distance') based on differences in allele frequencies. There are several algorithms for calculating *Fst* and the most used is that by (Weir & Cockerham 1984). Weir and Cockerham defined *Fst* as the ratio of the variance between populations to the total variance in the ancestral population.

Genetic architecture • A description of the number, effect size and allele frequencies of loci that contribute to the quantitative genetic variation, as well as their mode of action (i.e. additive, dominant, recessive), the way they interact with one another (**epistasis**) and the way their effect depends on environmental conditions (gene by environment interaction) (Slate *et al.* 2010). OR The number, frequency, effect size, dominance relationship and interactions of genetic variants that affect a trait in populations of a species (Bendesky & Bargmann 2011).

Genome assembly • The alignment and merging of fragments of DNA (i.e. reads, contigs and scaffolds) in order to reconstruct the genome of an organism (Wikipedia).

Genome-Wide Association Study (GWAS) • A complete form of analysis of mapped sequence reads (e.g. from RADtag, GBS, Capture-seq) or whole genomes to establish associations between genotype and phenotype. The more individuals included in the analyses the more powerful is this approach. Modified from (Liedvogel *et al.* 2011).

Genotypes • Part of the DNA sequence of an organism genome which could determine a specific phenotypic characteristic of that organism (i.e. physiological, physical or behavioural).

Genotyping • The process of determining differences on the genetic make-up (genotype) of an individual through molecular tools (i.e. DNA fragment analyses, sequencing) (Wikipedia).

Genotyping By Sequencing (GBS) • Method to discover SNPs in order to perform genotyping studies, such as GWAS (Wikipedia).

Geolocator • Electronic device that records ambient light levels over time, from which position (latitude and longitude) can be estimated. It may also record conductivity (wetness), allowing periods of flight to be determined.

Heritability • The proportion of phenotypic variation within populations explained by additive genetic variation (Liedvogel *et al.* 2011). Can be locus specific or for all loci combined (Risch 2000).

Hypothalamic-Pituitary-Adrenal (HPA) axis • The HPA axis consist of a complex set of input and feedback mechanisms between the hypothalamus, the pituitary gland and the adrenal gland. It is a major part of the neuroendocrine system (Scheiermann *et al.* 2013).

Internal clock • Internal representation of time given by the phase of endogenous rhythms, It can determine an organism's response to an environmental factor (Schwartz *et al.* 2017).

Library (genomic library) • Collection of DNA (or RNA) fragments modified in a way that is appropriate for downstream analyses, such as high-throughput sequencing (Ekblom & Wolf 2014).

Linkage Disequilibrium (LD) • Two alleles at different loci that occur together within an individual more often than would be predicted by random chance. Also called population allelic association (Risch 2000).

Locus (pl. loci) • Latin term meaning 'place'. It refers to a fixed position on a chromosome or genome.

Mate paired-end library • Collection of fragments of DNA of known and selected length (2–5 kb long) that generate mate pair-end reads (or Long-Insert Paired-End Reads,

LIPER). LIPERs are useful for genome assembly by informing about the distance between pair of DNA fragments (Ekblom & Wolf 2014). The combination of pair-end libraries and mate paired-end libraries allows to get larger contigs and scaffolds, therefore improving the accuracy of genome assemblies.

Mendelian (trait) • A monogenic trait, that is a one-to-one correspondence between genotype and phenotype. Modified from Risch (2000).

Model organism • In vivo models used in science to understand particular biological phenomena.

Mono/oligogenic • When the trait in question is determined by one or a few genes.

Microarray • NA Type of analysis techniques used in interpreting the data generated from experiments on DNA, RNA and protein microarrays, which allow researchers to investigate the expression state of a large number of genes – in some cases an organism's entire genome – in a single experiment (Wikipedia).

Microsatellite • DNA variant due to tandem repetition of a short DNA sequence (usually two to four nucleotides) (Risch 2000).

(Animal) Migration • The long-distance relocation of individuals' animals, which usually involves a substantial number of individuals and is often in response to a seasonal or life-history cue. Taken from https://www.nature.com/subjects/animal-migration.

Migratory gene package • Set of genes controlling physiological adaptations necessary for a successful migratory journey, such as timing of migration, body fat deposition and moult as well as the genes that encode migratory routes. Taken from (Liedvogel *et al.* 2011).

Next-generation-sequencing (NGS) • Highly parallel DNA sequencing where hundreds of thousands or millions of reads (sequences) are produced in one run (Stapley *et al.* 2010b).

Opportunistic • Organisms that can take rapidly advantage of favourable conditions and adapt to a diverse range of circumstances.

Oscillator • A system capable of producing regular, periodic fluctuation of an output around a mean (Schwartz *et al.* 2017).

Pacemaker • A function entity capable of self-sustaining oscillations. Taken from (Dunlap *et al.* 2004).

Pair-end library • Collection of fragments of DNA of known and selected length (200-800 bp long) that generate pair-end reads (or Short-Insert Paired-End Reads, SIPER). SIPERs are useful for genome assembly by informing about the distance between pair of DNA fragments (Ekblom & Wolf 2014). The combination of pair-end libraries and mate paired-end libraries allows getting larger contigs and scaffolds, therefore improving the accuracy of genome assemblies.

Period • Time after which a defined phase of the rhythm re-occurs; i.e. time taken for a full cycle (Schwartz *et al.* 2017).

Phenotype • The observable characteristics in an individual resulting from the expression of genes and possibly affected by environmental factors. Modified from NCI dictionary of genetic terms website).

Photoperiod or daylength • The period of time during which an organism receives illumination from the sun each day.

Polymerase • Enzyme that links nucleotides together into a strand using another strand as a template.

Polymorphism • More than one variant of a particular DNA sequence or loci.

Polygenic • When the trait in question is determined by a cumulative effect of many genes.

Population genetics • The study of the genetic composition of populations**.**

Primer • A short strand of DNA (~ 18–22 bp) that serves as a starting point for DNA synthesis by DNA polymerases.

Probe • Fragment of DNA (or RNA) which is labelled in some way and which is used to hybridize with a target nucleic acid of interest.

Quantitative genetic analyses • A branch of population genetics that deals with phenotypes that vary continuously (such as height or mass) as opposed to discretely identifiable phenotypes and gene-products (such as eye-colour, or the presence of a particular biochemical) (Wikipedia).

Read • Short base-pair sequence inferred from the DNA/RNA template by sequencing (Ekblom & Wolf 2014).

Restriction site Associated DNA (RADtag) • Genetic method that could be considered as "intermediate" between marker and whole genome focus. By digesting the genome with a restriction nuclease and attaching a series of adapters to the resulting DNA fragment, large number of genetic variation such as SNPs can be readily identified with next-generation sequencing technologies. Modified from https://www.floragenex.com/rad-seq/).

Retinohypothalamic tract • A photic input pathway that connects photosensitive retinal ganglion cells directly to the suprachiasmatic nuclei of the hypothalamus (Scheiermann *et al.* 2013).

RNA-sequencing • Highly accurate and efficient approach for transcriptome profiling and characterisation using next-generation sequencing technology (Liedvogel *et al.* 2011).

Scaffold • Sequence of pair bases reconstructed from one or more contigs using read-pair information (Ekblom & Wolf 2014). Unsequenced regions between pairs of contigs and between scaffolds are represented as runs of N's in the final genomic assembly.

Seasonal migration • Type of migration characterize by periodic movement of a population from one region or climate to another in accordance with the yearly cycle.

Seasonality • The presence of variations that occur according to the seasons.

Sequencing • The procedure of determining the order of base pairs in a section of DNA.

Suprachiasmatic nucleus • A small region in the hypothalamus of the brain consisting of roughly 10,000- 20,000 neurons located in a small region of the hypothalamus above the optic chiasm, from where they receive environmental light input through the **retinohypothalamic tract**. Also known as the master clock in mammals (Zhang & Kay 2010; Scheiermann *et al.* 2013).

Synchronizer • This term is equivalent to *Zeitgeber*: A periodic external signal capable of entraining a biological rhythm.

Single Nucleotide Polymorphism (SNP) • A single variable nucleotide in the DNA sequence. SNPs are usually bi-allelic, but multi-allelic SNPs do exist. They are the most

common form of variation in the genome and are used extensively to study genetic differences between individuals and populations (van Dijk *et al.* 2014).

Statistical model • Mathematical representation of reality (i.e. relationship between two variables, for example: genotypes and phenotypes) which embodies a set of assumptions concerning this reality.

Tailed-primers • Primers that contain a non-complementary sequence at their 5' end for a diversity of purposes (i.e. improve quality of PCR products to obtain better accuracy of DNA fragment sizes).

Thermal cycler or Polymerase Chain Reaction (PCR) • A method that produces millions of copies of a short segment of DNA.

Trait • Feature (i.e. morphological or behavioural) of an organism. OR A distinct variant of a phenotypic characteristic of an organism; it may be either inherited or determined environmentally, but typically occurs as a combination of the two (Lawrence 2005).

(Quantitative) Trait (Loci) mapping studies • Approaches in which genetic loci related to complex traits are identified through the statistical analysis of complex traits (Barton & Keightley 2002).

Transcriptomic technologies • These are techniques used to study an organism's transcriptome, that is, the sum of all of its RNA transcripts (Wikipedia).

Zeitgeber • A German term that means 'time-giver'. A periodic external signal capable of entraining a biological rhythm. In circadian context light is the predominant (Schwartz *et al.* 2017).

Zugunruhe • German term that means 'restlessness'. Expression of migratory activity showed by caged migratory individuals which reflects a physiological disposition to migrate (Berthold, 1996).

Chapter 2

Population genetic structure of New Zealand-overwintering bar-tailed godwits (*Limosa lapponica baueri*) in relation to breeding geography and migration timing

Rob Schuckard (on the left) and Gillian Vaughan checking the position of godwits in front of the hidden cannon nets at the Catlins Coast (November 2014).

Ángela M. Parody-Merino, Phil F. Battley, Yvonne I. Verkuil, Jesse R. Conklin, Francisco Prosdocimi, Murray A. Potter, Andrew E. Fidler

Abstract

A cline exists in the morphology of Alaskan-breeding bar-tailed godwits (*Limosa lapponica baueri*), with northern breeders being smaller and darker-plumaged than southern breeders. Differences in the timing of the spring thaw mean that northern breeding grounds become available later than southern breeding grounds, a difference that is reflected in the timing of northward migration of birds from the non-breeding grounds in New Zealand (N.Z.) and after staging in Asia. In the context of investigating the molecular basis of individual migration timing, we tested whether there was underlying neutral genetic structure between northern-breeding and southern-breeding godwits, and between early-migrating and late-migrating birds. Analysing 27 polymorphic microsatellite loci in 36 geolocator-tracked birds, we found evidence of weak genetic structure (F_{ST} = 0.011) between individuals breeding in southern Alaska on the Yukon-Kuskokwim Delta and those breeding further north (Seward Peninsula and northwards). As the timing of migration from Asia is highly correlated with breeding latitude, there was a similar level of structure between early and later migrants from Asia (F_{ST} = 0.013–0.016). In contrast, no detectable structure was present across 156 birds with known migration phenology in N.Z. (F_{ST} = 0.002, not significant). Greater overlap in migration timing of northern and southern breeders in migration timing from N.Z. than from Asia is apparently sufficient to dilute the already slight population structure present.

Introduction

Genetic structuring of natural populations may arise from phylogenetic and demographic histories and/or natural selection but, whatever its origin, such structuring must be maintained by barriers to gene flow (Slatkin 1985). Geographical features are the most often invoked barriers to gene flow, however differences in the timing of reproduction (i.e. allochrony) can also hinder gene flow, even in sympatric populations (Hendry *et al.* 1999; Rolshausen *et al.* 2009; Møller *et al.* 2011; Liedvogel & Lundberg 2014; Fudickar *et al.* 2016; Taylor & Friesen 2017). Differences in reproduction-associated timing in the order of weeks or even days can constitute an allochrony-based barrier to gene flow (Hendry *et al.* 1999; Miyatake 2002; Tauber *et al.* 2003; Fuchikawa *et al.* 2010; Møller *et al.* 2011; Rund *et al.* 2012; Taylor & Friesen 2017). In annually migrating animals, migratory allochrony may interact with geography to limit gene flow,

with an animal's location at the time of breeding/mating being influenced by the timing mechanisms governing its migratory schedule (Hendry & Day 2005).

Bar-tailed godwits (*Limosa lapponica*) provide a striking example of the interaction between spring migratory departure date and ultimate breeding location (Conklin *et al.* 2010). Godwits of the sub-species *baueri* breed in western Alaska in the boreal summer, then migrate non-stop across the Pacific Ocean to non-breeding grounds in Australia and New Zealand (N.Z.) (Gill *et al.* 2009; Battley *et al.* 2012). In the austral autumn (i.e. boreal spring) godwits return to Alaska with a refuelling stop of 5–7 weeks in eastern Asia (Gill *et al.* 2009; Battley *et al.* 2012) (Figure 2.1a). Their breeding grounds extend from the Yukon-Kuskokwim Delta in the south (~59˚N) to the North Slope of the Brooks Range (71˚N). Body size and extent of breeding plumage vary across this range, with southern breeders being larger and paler than northern breeders (Conklin *et al.* 2011).

Bar-tailed godwits leave N.Z. from late February–early April, but individual birds consistently leave over a much tighter time-frame, typically within the same week each year (Battley 2006; Conklin & Battley 2012; Conklin *et al.* 2013). The explanation for this pattern is that optimal arrival times in Alaska vary with breeding latitude (the spring is roughly 3.5 weeks later in the north), and therefore southern breeders migrate and arrive in Alaska much earlier than northern breeders (Conklin *et al.* 2010). This ultimate timing difference on the breeding grounds is anticipated by earlier departures of southern birds from Asia and even from N.Z. (two months before the breeding season; Conklin *et al.* 2010). Thus, the timing of departure of individual birds in the southern hemisphere largely reflects geographic variation fully 12,000 km away in the northern hemisphere, 18,000 km as the northward-flying godwit travels (Battley *et al.* 2012). As arrival synchrony on breeding grounds can be an important determinant of selection and retention of mates (Gunnarsson *et al.* 2004) the timing of migration of individual birds has the potential to affect gene flow by limiting the formation of new mates and therefore contributing to a less genetic admixture in the population (D'Urban Jackson *et al.* 2017).

In this study, under our expectation of no or little genetic structure in the N.Z. godwit population (Rönkä *et al.* 2008; Conklin *et al.* 2016), we test whether there is population genetic structure across the Alaskan breeding range, using birds tracked via geolocators from their non-breeding grounds in N.Z. We then ask whether any population structure present in Alaska is also associated with differences in migration timing in Asia and N.Z., given (a) the relationships between migration timing and breeding latitude

(Conklin *et al.* 2010) and (b) the consistency of migration schedules between individual godwits at all stages of the migration (Conklin *et al.* 2013). We also discuss whether there is any evidence for allochrony *per se* operating independent of geographical variation in migration timing.

Materials and methods

Blood sample collection

Godwits (n = 223) in N.Z. were trapped by cannon net or mist net from 2004–16 at two sites in the North Island (Firth of Thames, 37.17˚S 175.32˚E, n = 95; Manawatu River estuary, 40.47˚S, 175.22˚E, n = 109) and at one site on the South Island (Catlins Coast, 46.48˚S, 169.70˚E, n = 19) (Figure 2.1b). Blood samples (ca. 50.0 µl) were collected from the metatarsal or brachial vein using microhaematocrit capillary glass

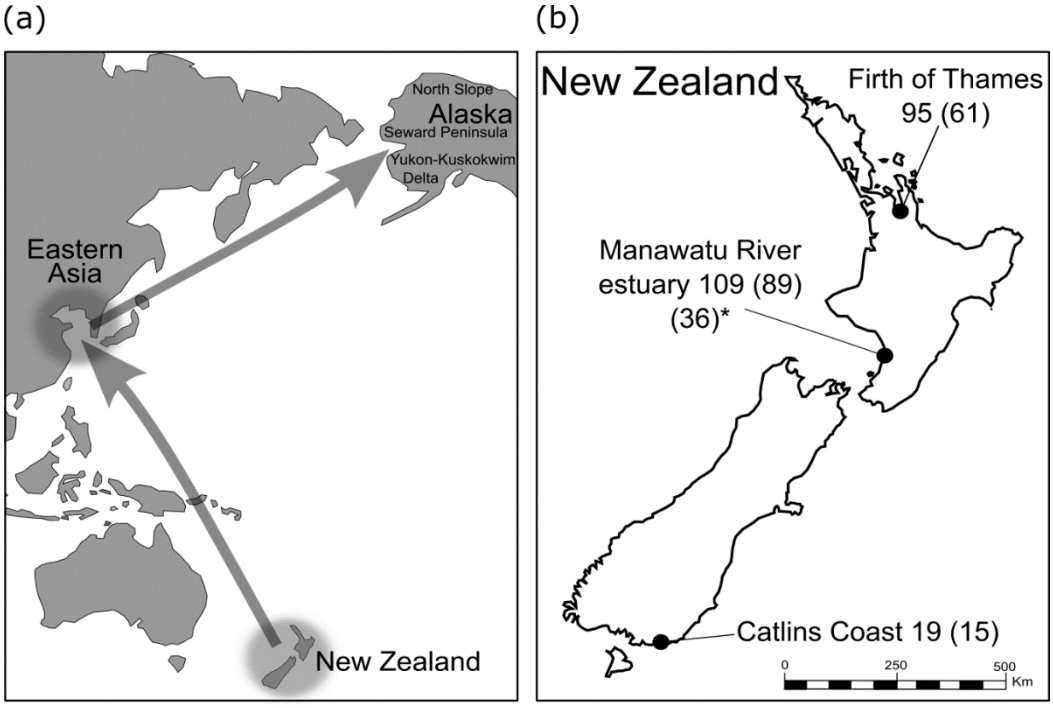


Figure 2.1: (a) General migratory route of bar-tailed godwits from over-wintering sites in New Zealand to the stopover in eastern Asia and on to the breeding grounds in Alaska (which extend from the ~59˚N on the Yukon-Kuskokwim Delta to 71˚N on the North Slope. (b) Locations of the three study sites in N.Z. with sample sizes for birds blood sampled, the subset with known migratory departure dates (in parentheses), and for the Manawatu River estuary, the number of birds tracked by geolocation ()*.

tubes and preserved in either 96% (v/v) ethanol or Queen's Lysis buffer (QLB) (10 mM Tris, 10 mM NaCl,10 mM EDTA, 1.0% (w/v) *n*-lauroylsarcosine, pH 8.0) (Seutin *et al.* 1991) before long-term storage at ambient temperature. Godwit capturing and blood sampling were carried out under both N.Z. Department of Conservation (DOC) permits and Animal Ethics Committee approval from Massey University (#07/163, #12/90) and the University of Otago (#66/03).

Microsatellite genotyping

Genomic DNA was extracted from blood samples using a commercial kit following the manufacturer's protocol (Quick-gDNA™ MiniPrep, cat. no. D3050, Zymo Research, Orange, U.S.A.). PCR primer pairs (n = 41) for amplifying candidate microsatellite sequences were obtained from two sources: 12 from a previous publication (Verkuil *et al.* 2009) and 29 identified in a bar-tailed godwit genome sequence (GenBank Acc. No. LXVZ00000000.1). Fluorescent labelling of PCR-amplified sequences was achieved following the procedure of Schuelke (2000) whereby a fluorescently labelled M13(-21) universal sequence (18 bp) primer (5'- TGTAAAACGACGGCCAGT -3') was included in the reaction mix along with the locus specific primer pair, one of which also included the generic primer sequence. The M13(-21) universal primers were 5' labelled with 6-FAM™ (blue), VIC® (green), NED™ (yellow), PET™ (red) allowing discriminating colour labelling of microsatellite amplicons and pooling of up to four different microsatellite loci amplicons for fragment length estimation. Reverse primers included the 5' 'PIGtail' sequence (5'- GTTT -3') to promote the addition of single 3' A's to amplicons thereby reducing amplicon length heterogeneity (Brownstein *et al.* 1996). Template genomic DNA (2.0 µl) aliquots were dried on the base of 0.2 ml PCR tubes (70˚C, 10 min.). PCR reaction cocktails (10.0 µl) were added consisting of 1x MyTaq™ HS Mix (cat. no. BIO25045, Bioline, London, U.K.), locus-specific forward primer (0.2 µM), locus-specific reverse primer (0.6 µM), and one of the four alternative florescent dye-labelled M13 universal primers (0.6 µM). Thermocycling conditions for all primer pairs were: 95˚C/2 min., 1 cycle; 95˚C/30 sec., 52˚C/20 sec., ramping at 0.2˚C/sec to 72˚C, 72˚C/30 sec., 20 cycles; 95˚C/30 sec., 54˚C/30sec., ramping at 0.2 ˚C/sec to 72˚C, 72˚C/30 sec., 22 cycles; 72˚C/5 min., 1 cycle; 60˚C/30 min. 1 cycle; 15˚C/hold. Amplicon lengths were estimated on an *ABI3730* DNA Analyzer with the GeneScan™-500 LIZ™ Size Standard (Massey Genome Service, Massey University, N.Z.). PEAK SCANNER™ v.2.0 (Life Technologies, Carlsbad, U.S.A.) was used to

estimate amplicon fragment sizes with the software TANDEM used to round the fragment sizes to valid integer values (Matschiner & Salzburger 2009).

Microsatellites analyses

We conducted initial analyses on the full set of 223 godwit blood samples. The software package GENALEX v.6.5 was used to check for samples having identical genotypes and for calculating a range of standard population genetic parameters (i.e. allele frequency, alleles/marker, expected (H_e) and observed (H_o) heterozygosities) (Peakall & Smouse 2012). MICRO-CHECKER v.2.2.3 (van Oosterhout *et al.* 2006) was used to detect possible genotype scoring errors and to detect possible null alleles, as indicated by an apparent excess of homozygotes. Linkage disequilibrium (LD) analysis was performed using ARLEQUIN v.3.5.1.3 (Excoffier & Lischer 2010) using 10,000 permutations, 10 initial conditions and an interval of confidence (IC) of 0.05. ARLEQUIN v.3.5.1.3 was used to detect possible deviations from Hardy-Weinberg equilibrium (HWE). Each HWE analysis consisted of 1,000,000 iterations of the Markov chain Monte Carlo (MCMC) and 1,000,000 dememorization steps. Sequential Bonferroni corrections ($P < 0.05$) for multiple comparisons were applied to both HWE and LD results to minimize type I statistical errors (Rice 1989).

Nine of the initial 41 primer pairs did not reliably generate amplicons or did not provide clearly interpretable electropherogram peaks and so were discarded. Of the remaining 32 microsatellite loci, one (LIM22; Verkuil *et al.* (2009)) was monomorphic across all the genomes tested while two (LIM10, LIM12a; Verkuil *et al.* (2009)) were found to be in apparent LD; the less polymorphic locus (LIM10) was removed from further analyses. Summary statistics for the 30 microsatellite loci are listed in Appendix 1, Table 1.1 with details of the primers used to amplify the loci provided in Appendix 1, Table 1.2. Heterozygous females were detected for all 30 microsatellite loci indicating that all are located on autosomal chromosomes.

Numbers of detected alleles (*Na*) at the microsatellite loci ranged from 2 (LIM25) to 14 (LIM26) with a mean of 6.4 and median of 5 (Appendix 1, Table 1.1). No identical genotypes were found amongst the 165 blood sample genotyped, confirming that there were no duplicated samples**.** Loci LIM26, C267, and C998 were identified as possibly having null alleles and their genotype frequencies differed significantly from HWE, *P* (after Bonferroni correction) < 0.05, (Appendix 1, Table 1.1). Microsatellite loci whose genotype frequencies do not conform to HWE may contain important population genetic

structuring information so all subsequent population genetic analyses were carried out using both the complete set of microsatellite loci (n = 30) and a subset of loci (n = 27) from which LIM26, C267, and C998 genotypes were removed. Since results from both microsatellites sets were the same (see Appendix 1, Table 1.3 and Appendix 1, Figures 1.1– 1.9), we present results from the 27 microsatellites loci that conformed to HWE.

Determination of northern and southern-breeders

The light sensors on the geolocators enable us to assign general breeding location (i.e. latitude) in Alaska (i.e. Yukon-Kuskokwim Delta, Seward Peninsula or North Slope). The boundary we used was between the Yukon-Kuskokwim Delta and the Seward Peninsula (64°N) (see Appendix 1, Figure 1.10), principally because this is the latitude threshold where light sensors are able to register nights during the breeding seasons (Conklin 2011). For instance, some ground-truthing in Alaska with units placed out on the ground indicated that in Nome (64˚N) there was no evidence of night time in mid-summer (unpublished data). That latitude also separates two distinct geographic areas (Appendix 1, Figure 1.10). We classified birds that showed some dark during the night-times as breeding on the Yukon-Kuskokwim Delta and denoted as 'southern breeders' and those having essentially continuous daylight breeding on the Seward Peninsula or North Slope and denoted as 'northern breeders' (Appendix 1, Table 1.4). Since there are significant differences in size and extent of breeding plumage between northern and southern breeders as described in Conklin *et al.* (2011), we took into account these biometrics together with light-levels from geolocators to classify individuals into their putative Alaskan breeding groups. We refer to these groups as 'southern breeders' (n = 19) and 'northern breeders' (n = 17).

Determination of individual chronophenotypes and chronophenotype-based groups

We refer to *chronophenotype* as the mean departure date of an individual obtained by calculating the mean departure date from all the years available (1–9 departure dates per bird for the N.Z. dataset; 1 or 2 departure dates for the Asian dataset) (Appendix 1, Table 1.4). Migration departure dates were assigned by designating 1 January as day 1. Thirty-six individuals were tracked from the Manawatu River estuary using geolocators (models used: British Antarctic Survey MK14, Biotrack MK4093 and Migrate Technology Intigeo-C65K; tracking was done in 2008, 2009, 2013 and 2014) allowing their departure date from N.Z. and the Asian stopover along the northward migration route

to be determined. The geolocators recorded conductivity (i.e. marine wetness) every 10 minutes to 4 hours depending on the brand and model; ground-truthing by direct observation confirmed that the conductivity signal changed unequivocally when individuals embarked on a migratory flight from N.Z. (Battley & Conklin 2017). These conductivity data were used to determine the departure date from Asia. Migration departure dates of 165 banded bar-tailed godwits were determined at three N.Z. over-wintering sites, over the years 2004–16 by daily observation of individually-marked birds (Conklin *et al.* 2013) (Figure 2.1b). At two study sites (Manawatu River estuary and Catlins Coast) the local population was small and detailed observations could be made. At the Manawatu River estuary (population ~200) birds were generally observed in departing flocks or departure deduced from daily roll-calls and flock counts (n = 89 birds, 2008–16) (Conklin *et al.* 2013). Note that 36 individuals from the Manawatu estuary carried geolocators therefore their observed departure dates were verified with that indicated by the geolocators' information. At the Catlins site (population ~350), departure dates were deduced from daily roll-calls and flock counts (n = 15 birds, 2013–16). At the Firth of Thames, which has a much larger population (>3000 individuals), last records of birds seen multiple times were taken to represent migration dates (n = 61 birds, mostly 2004–06 and 2014–16) (Battley 2006). Note that calculated repeatability values for migration dates were virtually identical at the Firth of Thames and Manawatu River estuary (R = 0.83–0.84) despite the methodological differences in determining departure date (Battley 2006; Conklin *et al.* 2013).

The chronophenotypes form a continuum that presents a challenge for detecting population genetic differentiation between discrete groupings. If any genetic structure associated to different timing schedules exists, it would be most obvious between individuals placed at the extreme of the chronophenotype continuum. We used the correlation between godwit migratory departure date from N.Z. and Asia and ultimate breeding latitude reported by Conklin *et al.* (2010) to establish reasonable/sensible divides to form discrete chronophenotypes groups. Across the chronophenotype continuum two days of span in migration timing correspond to separation in three breeding areas: Yukon-Kuskokwim Delta, Seward Peninsula and the North Slope (see Appendix 1, Figure 1.10). The 36 godwits tracked by geolocators were assigned to putative groups on the basis of their chronophenotype from Asia (Appendix 1, Figure 1.10). We used departure date to form *a priori* groups at two dates: day 128 (8 May) and day 135 (15 May). We refer to these groups as '< day 128' (n = 16) vs. '> day 128' (n =

20) and '< day 135' (n = 25) vs. '> day 135' (n = 11) (Figure 2.2b). For the N.Z. dataset a timing boundary at day 81 (22 March) was used to group birds into 'earlier' and 'later' chronophenotypes. Nine individuals with chronophenotypes at day 81 were not used for these analyses, so the sample size used on these analyses was 156 rather than 165. We refer to these groups as '< day 81' (n = 111) and '> day 81' (n = 45) (Figure 2.2a).

Population structure analyses

To test for genetic structure within the Alaska, Asia and N.Z. datasets and their corresponding *a priori* groups (i.e. 'southern' vs. 'northern' breeders, 'earlier' vs. 'later' departers from Asia and N.Z.) we used three approaches: 1) Bayesian clustering in STRUCTURE v.2.3.4 (Hubisz *et al.* 2009), using the admixture and prior locations models for $K = 1–5$; length of burn-in: 1,000,000; MCMC: 500,000; number of iterations per run = 15 (Hubisz *et al.* 2009). STRUCTURE HARVESTER (Earl & Vonholdt 2012) was used to merge results from the 15 replicates; 2) *K*-means clustering, which utilizes Bayesian Information Criteria (BIC), was used to detect genetic clusters based on the microsatellite genotypes (R package ADEGENET, function *find.clusters* (Jombart 2008) retaining 100% of the Principal Components (PCs) and $K = 1–5$; and 3) AMOVA- F_{ST} in ARLEQUIN v.3.5.1.3, with 10,000 permutations and using a bootstrap approach (20,000 bootstraps). The best number of clusters for STRUCTURE was decided using the "elbow criterion", that is, looking at the *L(K)* graph where the number of clusters tested is higher or reaching a plateau (i.e. where adding another cluster does not give much better modelling of the data). For *K*-means, the Bayesian Information Criterion (BIC) value was used to determine the best number of clusters. Using STRUCTURE with uneven samples sizes might underestimate the number of clusters (Puechmaille 2016; Wang 2017). As sample sizes for some of our analyses were uneven (see Figure 2.2), we resampled the larger group to standardize the sample sizes and repeated analyses to test whether the imbalance in samples affected the results. This involved subsampling 45 individuals from the 111 in the N.Z. dataset, < day 81, and 11 of 25 for the Asian dataset, < day 135. Subsampling and analyses were repeated 10 times and with the same parameters as described before. Results were identical to those in the original analyses, with support for a single cluster in Asia and either one or possibly two clusters in N.Z. (Appendix 1, Table 1.5). Bonferroni correction ($P < 0.05$) was used to correct p-values for multiple testing (Rice 1989) when needed.

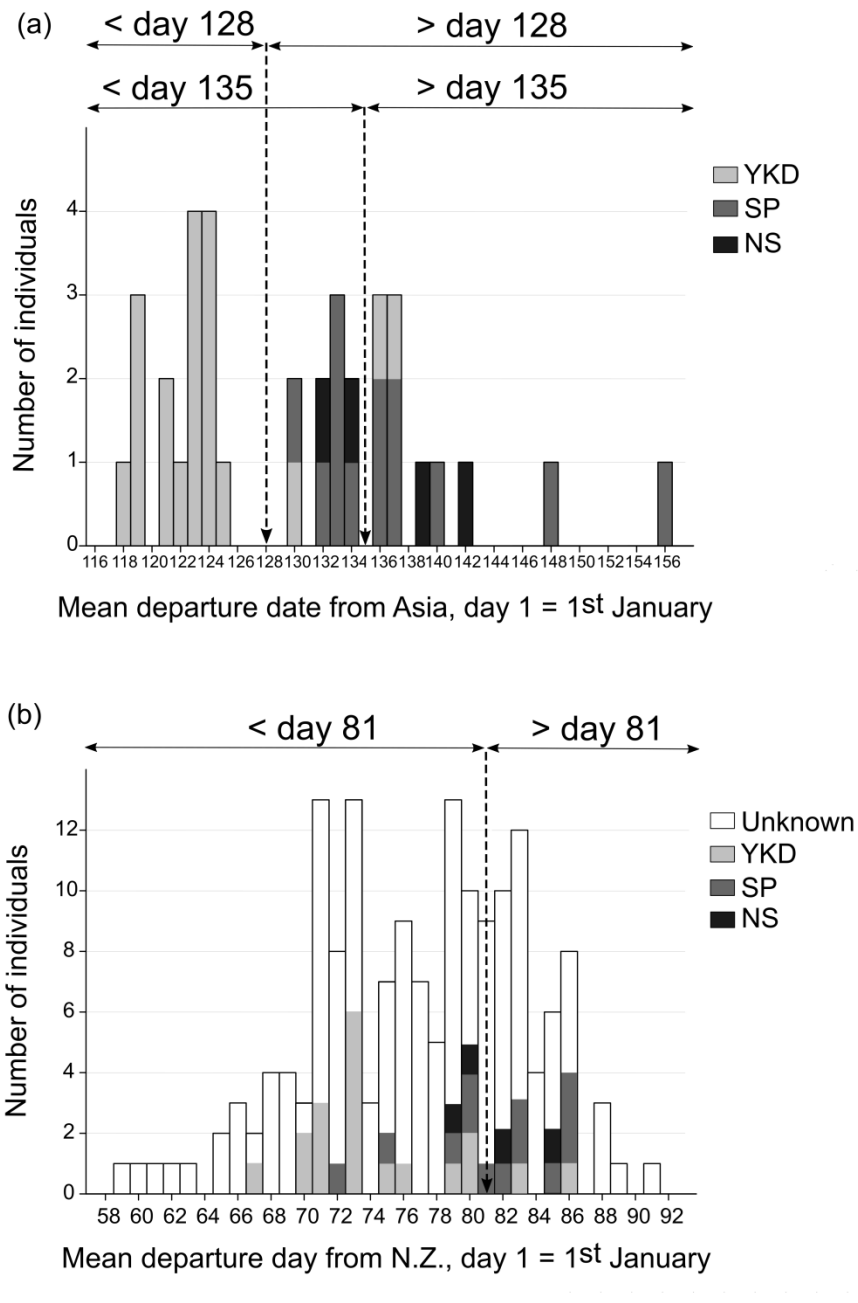


Figure 2.2: Distribution of godwit individuals along two migratory chronophenotype continua. (a) Departures from eastern Asia (n = 36). Chronophenotype was measured as the departure date (if 1 year available) or mean departure date (if 2 years available) from Asia (see Appendix 1, Table 1.4). (b) Departures from N.Z. (n = 165). Chronophenotype was measured as the mean departure date across 1–9 years for individuals at three sites across the over-wintering range in N.Z. (see Appendix 1, Table 1.4). Arrows indicates earlier vs later chronophenotype groups for comparisons defined based on Conklin *et al.* (2010) correlations (see Appendix 1, Figure 1.10). Shading on the bars represents the classification of birds into southern (Yukon-Kuskokwim) and northern (Seward Peninsula and North Slope) based on geolocators light levels and biometrics.

Results

Analysis of the population differentiation within Alaska indicated that there was slight genetic structure between northern and southern birds. Specifically, while clustering methods suggested birds belonged to a single population (Figure 2.3a; Appendix 1, Figure 1.1a and 1.2a), comparison of northern (Seward Peninsula and North Slope) and southern (Yukon-Kuskokwim Delta) breeders showed a small, yet significant genetic difference (F_{ST} = 0.011, P = 0.03; Table 2.1), although it became non-significant after Bonferroni correction (P = 0.08). Equivalent analyses using migration timing from Asia found essentially the same result – birds showed slight differentiation between early and late migrants (using divides of both day 128 and day 135; F_{ST} = 0.013, P = 0.02 and F_{ST} = 0.016, P = 0.01 respectively, the latter remaining significant after Bonferroni correction), though clustering methods supported there being a single cluster (Table 2.1; Figures 2.3b and 3.3c; Appendix 1, Figure 1.3a, 1.3c, 1.4a and 1.4c). In contrast, in N.Z. there was non-significant structure associated with migration timing (F_{ST} = 0.002, P = 0.09, P (Bonferroni) = 0.18), even though clustering methods supported there being either one or two clusters (Table 2.1; Figure 2.3d; Appendix 1, Figure 1.5a and 1.6a).

The geolocator-tracked birds were classified as northern or southern breeders on the basis of light levels and biometrics and we were able to then evaluate how much overlap there was between those groups in migration timing from Asia and N.Z. (Figure 2.2). In the Asia dataset, three southern breeders were classified as 'late' if day 128 was taken as the boundary, and two of those remained classified as 'late' even if the later cut-off of day 135 was used (see Figure 2.2a). However, these represent just 3/36 birds, meaning that the Asia analysis based on migration timing largely approximates the Alaska analysis based on geography. In the N.Z. dataset, there was greater overlap in migration timing between northern and southern birds (see Figure 2.2b), with seven northern birds being 'early' and two southern birds being 'late'. If the 36 geolocator-tracked birds that formed the basis of the significant Asia and Alaska analyses are reanalysed by their N.Z. migration timing, no genetic difference is detectable (F_{ST} = 0.006, P = 0.16, P (Bonferroni) = 0.18) (Table 2.1; Figure 2.3d; Appendix 1, Figure 1.7a and 1.8a).

Table 2.1 Summary of population genetic structure results between: 1) southern (i.e. Yukon-Kuskokwim Delta) and northern breeders (i.e. Seward Peninsula and North Slope) (Alaska_geoloc.); and 2) 'earlier' and 'later' chronophenotype-based groups for the Asian (Asia_chronop.) and N.Z. (N.Z._chronop.) datasets using *a priori* defined timing divides at days 128 and 135 for the Asian and at day 81 for the N.Z. datasets (see text for more details). Comparisons were made using three methods: AMOVA-*F*st, Bayesian and *K*-means clustering and using 27 microsatellites. The same summary table with results using the 30 microsatellites (i.e. including three not in HWE) is shown in Appendix 1, Table 1.3. (*) denotes significant *Fst* at $P < 0.05$ without Bonferroni correction; ** denotes significant *F*st at $P < 0.05$ after Bonferroni correction. *F*st values with Interval of Confidence (IC) = 95% are underlined. Numbers in italics are the numbers of individuals per group in the comparisons. *Abbreviation*: n, number of microsatellites; YKD, Yukon-Kuskokwim Delta; SP, Seward Peninsula; NS, North Slope.

Dataset	Divides	AMOVA- F_{ST}	Bayesian (*K*)	*K*-means (*K*)
Asia_geoloc. **n = 36**	YKD vs. SP+NS *19/17*	0.011(*)	1	1
Asia_chronop. **n = 36**	Day 128 *16/20*	0.013(*)	1	1
	Day 135 *25/11*	0.016**	1	1
N.Z._chronop. **n = 165**	Day 81 *111/45*[1]	0.002	1 (2)	2 (1)
N.Z._chronop. **n = 36**	Day 81 *24/11*[2]	0.006	1	1

[1] Nine individuals and [2] one individual were excluded for having their chronophenotype (mean departure date) = day 81.

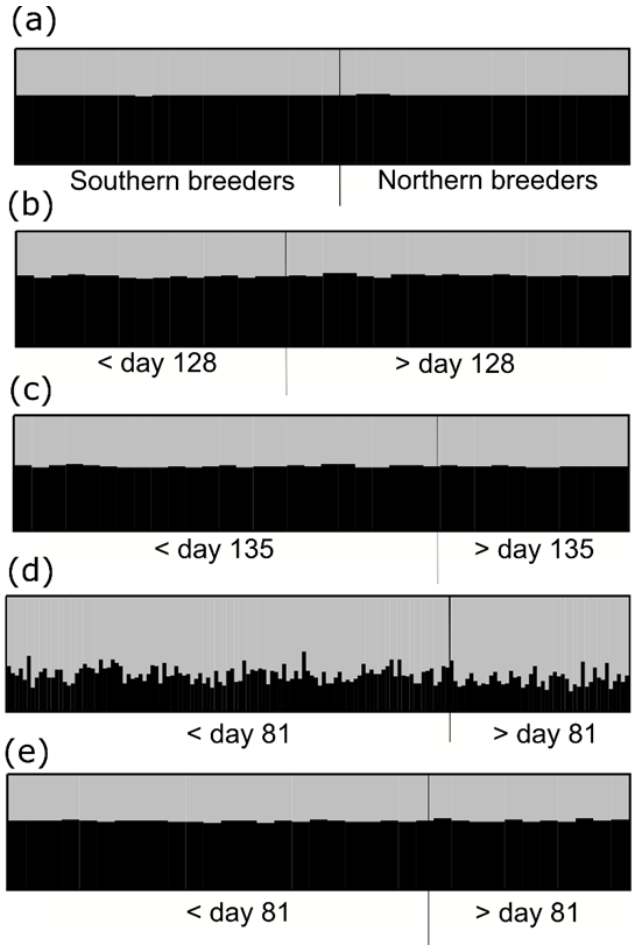


Figure 2.3: Bar-plots generated by STRUCTURE using the subset of 27 microsatellites in the different datasets (i.e. the Alaskan, Asian and N.Z. datasets). (a) Using the Alaskan breeding regions derived from light-levels of geolocators and biometrics (see text for more details) as prior location. Groups are: 19 birds 'southern breeders' (Yukon-Kuskokwim Delta) and 17 birds 'northern breeders' (Seward Peninsula and North Slope). (b) Using the Asian dataset (n = 36) classified in *a priori* 'earlier' and 'later' departers groups based on their chronophenotype derived from 1 or 2 recorded departure dates from eastern Asia (stop-over) and using two timing divides at day 128 and (c) at day 135. (d) Using godwit population with departure date for migration from N.Z. (n = 165) and *a priori* groups based on their chronophenotype and a date divide at day 81 with 111 birds '< day 81' and 45 birds '> day 81'. Note that nine individuals of the initial 165 godwits in the sample were excluded for this analysis as their chronophenotype fell at day 81. (d) Using the godwit dataset with geolocators and chronophenotypes from N.Z. Analyses were made with *a priori* groups based on a date divide at day 81 with 24 birds '< day 81' and 11 birds '> day 81'. One individual out of the initial 36 individuals in the sample was excluded for this analysis as their chronophenotype fell at day 81. Bar-plots of the same analyses but using the 30 microsatellites (i.e. three of them not conforming HWE) are shown in Appendix 1, Figure 1.9. Bars indicate the admixture proportion of individuals [0;1] assuming $K = 2$. When using chronophenotypes (i.e. (b) (c) (d) and (e)), individuals are organized from the earliest to the latest departers from left to right. Day 1 = 1 January.

Discussion

Godwit population genetic structure in Alaska

This is the first study of genetic variation across Alaskan-breeding bar-tailed godwits, and even though our sample size was small (being drawn from birds tracked by geolocators from N.Z.), we found support for a low level of genetic differentiation between northern and southern breeders. It is difficult to imagine that 80–150 km distance between the two regions that separates the boundary used in this study (i.e. the Yukon-Kuskokwim Delta and the Seward Peninsula (64°N)) represents a barrier to a trans-hemispheric migrant. However, godwits likely have high site-fidelity (Kruk *et al.* 1998), which if reinforced by any discontinuities in the breeding habitat might explain the apparent isolation-by-distance and therefore the existence of a morphological cline. The morphology of godwits varies across the Alaska breeding range, though not strictly in a continuous fashion (Conklin *et al.* 2011). Birds of both sexes are on average largest on the Yukon-Kuskokwim Delta, intermediate on the Seward Peninsula and smallest on the North Slope. In terms of breeding plumage, males are more extensively moulted from the north to the south (i.e. darkest on the North Slope and palest on the Yukon-Kuskokwim Delta). Among females, those from the Seward Peninsula show the greatest extent of prebreeding moult. Thus, using geographic boundaries that separate these three regions matches the morphological variation present in the taxon.

Godwit migration departure date and potential evolution by allochrony

Conklin *et al.* (2010) proposed that latitude associated differences in the timing of the spring thaw across the bar-tailed godwit's Alaskan breeding range were the selective force driving individual godwit migration timing in N.Z. and Asia. Here we show that the slight population genetic structure in relation to migration timing in Asia is of the same magnitude as shown between northern- and southern-breeding birds in Alaska. It would appear, therefore, that the population genetic structuring detected in relation to migration timing may be, in a sense, a secondary effect of a cline in godwit breeding location/latitude that is itself correlated with migration departure time. Indeed, we cannot separate how much of the genetic structure is explained by allochrony (separating by time) and how much by geographic isolation, but the two could plausibly operate in conjunction in this system.

For allochrony to operate independent of geography there would have to be sufficient variation in timing within a breeding region to allow assortative mating by timing. It is not impossible that such variation exists, as arrival dates do vary across the Yukon-Kuskokwim Delta (Conklin *et al.* 2010) and mate retention in Icelandic black-tailed godwits seems to rely on synchronous arrivals (Gunnarsson *et al.* 2004), though the low level of genetic differentiation between regions or timing groups implies that any residual variation attributable to allochrony must be slight and would likely be undetectable with our current sample size. Hence, genetically influenced migratory chronophenotypes, in combination with inter-annual philopatry and mate-fidelity, are likely to support genetic differentiation between bar-tailed godwits migrating on different schedules, but whether there is a role for allochrony *per se* as opposed to geography is unclear (Hendry & Day 2005; Casagrande *et al.* 2006; Friesen *et al.* 2007; Ketterson *et al.* 2015).

Clustering methods versus Fst

We obtained evidence of subtle genetic structure from the AMOVA- F_{ST} analyses but not from the clustering methods, which may be due to differences in how these methods operate. STRUCTURE and *K*-means methods, both follow a Bayesian approach and could not find clusters in the Asia dataset. They seem to be more sensitive to the number of individuals by limiting the possibility to find the 'correct' initial cluster centres. This means that with small sample sizes – like the eastern Asia dataset (n = 36) – it is more difficult for these algorithms to find clusters. With larger sample sizes, as was the case when using the N.Z. dataset (n = 156), we did indeed have some suggestion of more than one cluster being present from the Bayesian and *K*-means analyses. In contrast, AMOVA-$F_{ST,}$ which calculates genetic differentiation between populations as the reduction in heterozygosity of each subpopulation relative to the total heterozygosity assuming random mating, does not seem to be influenced by sample size. We can see this as we performed three analyses using the same 36 individuals (i.e. using breeding location in Alaska, using chronophenotypes from Asia and from N.Z.), however we obtained significant F_{ST} values in two of these analyses and one non-significant result (see Table 2.1). Indeed, in the godwit context (i.e. with a subtle population genetic structure, ~ *Fst* < 0.05) a sample size of 36 individuals and 27 polymorphic microsatellites seem to be enough to produce good estimates of AMOVA-*Fst* (Willing *et al.* 2012).

Conclusion

We found evidence of a low level of genetic differentiation between northern-breeding and southern-breeding bar-tailed godwits in Alaska. This matches differences between earlier and later migrants in Asia, which we show is a result of greater overlap of migration timing of northern and southern birds in N.Z. than in Asia. Whether allochrony (variation of individuals' migratory schedules/time) plays a role in driving this population genetic structure independent of any geographic structure remains unclear.

Building on quantitative genetic studies, the application of genomic technologies is now offering the possibility of characterizing functionally significant variation in key loci at the nucleotide level – taking either a whole genome (Genome Wide Association Studies, GWAS) or candidate gene association (CGA) approach (Liedvogel *et al.* 2011; Chakarov *et al.* 2013; Ruegg *et al.* 2014; Delmore *et al.* 2015; Saino *et al.* 2015). Although a larger panel of markers (i.e. SNPs) would be needed to definitively conclude a lack of population structure, an important implication from this study is that that there is little if any underlying population structure associated with migrating timing for N.Z. that could confound any significant functional associations found in investigations of molecular correlates of migration timing. Genetic structure of the level of what we showed for the Asian dataset in the current study (i.e. as subtle as $Fst = \sim 0.01$) could lead to false positive results – as well as to false negatives – and therefore would need to be accounted when performing phenotype-genotype association studies (Price *et al.* 2010; Qin *et al.* 2010). Nevertheless, bar tailed-godwits provide a promising context in which to investigate associations between migration timing and allelic variation in specific genes. Such research will likely contribute to our understanding of the mechanisms underlying within-population behavioural migratory timing variation as well as the underlying organism-level photoperiodic responses.

Chapter 3

No evidence for an association between *Clock* gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*).

After cannons were fired, captured birds were rapidly extracted and distributed in holding boxes by the volunteers and ornithologists. Photo taken at the Catlins Coast (November 2014).

Ángela M. Parody-Merino, Phil F. Battley, Jesse R. Conklin, Andrew E. Fidler

Abstract

Appropriate timing of seasonal migration is crucial for individuals that breed and over-winter in widely-separated sites. The bar-tailed godwit (*Limosa lapponica baueri*) is a well-documented example amongst long-distance migratory species in which individuals show strong inter-annual consistency in their migration timing despite the population migrating across 5–6 weeks. Recent studies have suggested that variation in a region of the gene *Clock* encoding a polyglutamine repeat (*Clk*polyQcds) is associated with migratory timing. Here we describe allelic variation of the *Clk*polyQcds within the godwit population over-wintering in New Zealand (N.Z.) and investigate whether polymorphism in this region is associated with migration departure date from N.Z. or from eastern Asia (stopover). We genotyped *Clk*polyQcds in 135 individuals with known N.Z. departure dates, 32 of which were also tracked, using geolocators, through their Asian stopover. We analyzed variation in *Clk*polyQcds in relation to mean departure date on migration (chronophenotype) from N.Z. and Asia and in relation to predefined chronophenotype discrete categories ('earlier' vs. 'later' departers). Six *Clock* alleles (Q_7–Q_{12}) were detected and substantial variation was observed between individuals (heterozygosity of 0.79). We found no association between *Clk*polyQcds polymorphism and migration departure timing from either N.Z. or Asia, whether the sexes were analyzed separately or pooled. A slight suggestion of structure in *Clk*polyQcds composition between birds with 'earlier' versus 'later' chronophenotype could not be separated from underlying neutral genetic variation. We conclude that allelic variation at *Clock* plays no clear role in the timing of migration of bar-tailed godwits.

Introduction

Long-distance migratory birds are presented with the annually-repeating challenge of ensuring that a crucial life-history event, migratory departure from their over-wintering sites, is timed so that subsequent reproduction-related life history events (e.g. courtship, mating and nesting) occur at appropriate times at far-distant locations (Verboven & Visser 1998; Both & Visser 2001; Both *et al.* 2006; Nussey *et al.* 2007). Because individuals have to schedule their departure from the over-wintering sites, which could be thousands of kilometers away from the breeding sites (Both & Visser 2001; Jenni & Kery 2003; Both *et al.* 2006; Both *et al.* 2010; Kölzsch *et al.* 2015), predictability

depends strongly on a genetically influenced timing mechanism (Berthold & Querner 1981).

Avian migration timing is thought to be determined by inputs from an internal/intrinsic circannual clock interacting with responses to extrinsic, seasonally-varying *Zeitgebers*, which at temperate latitudes are principally the seasonally changing photoperiod (Both & Visser 2001; Jenni & Kery 2003; Helm & Gwinner 2005; Both *et al.* 2006; Helm *et al.* 2009; Both *et al.* 2010; Helm *et al.* 2013; Kölzsch *et al.* 2015; Majumdar *et al.* 2015). The genetic and physiological mechanisms generating internal circannual clocks remain largely unknown (Rani & Kumar 2013; Evans *et al.* 2015; Majumdar *et al.* 2015). In contrast, some genetic and physiological elements of the avian internal circadian clock used to measure seasonally changing photoperiod can be deduced from their evolutionary and functional relationships with similar (i.e. homologous) elements in more experimentally accessible model organisms (Panda *et al.* 2002; Hardin 2005; Helfer *et al.* 2006).

Bar-tailed godwits that over-winter in Australasia (*Limosa lapponica baueri*, hereafter 'godwits') are long-distance migratory shorebirds that present an extreme example of the need to integrate the timing of initiation of migration with timing of breeding far away (Conklin *et al.* 2010). The annual migration schedule of godwits that over-winter in New Zealand (N.Z.) consists of three non-stop flights over open ocean: departing N.Z. in austral late summer/early autumn godwits fly ~10,000 km to feeding grounds in coastal east Asia, followed by a stop-over of 4–7 weeks and then a flight of ~7,000 km to western and northern Alaska to breed, departing in the boreal autumn, both returning adults and fledglings complete a >11,700 km flight directly across the Pacific Ocean to N.Z. (Gill *et al.* 2009; Battley *et al.* 2012). Given the short temporal window for successful breeding in Alaska (Holmes 1971; Smith *et al.* 2010) it is expected that initiation of all three migratory flight sections is under strong selection pressure with strong penalties against inappropriate timing (Drent *et al.* 2003; Conklin *et al.* 2013; Visser *et al.* 2015). Nonetheless rather surprisingly it has been found that migrating godwits depart N.Z. over an approximately 5-week period in austral late summer/early autumn (i.e. late February – early April) but that individual godwits typically depart within the same week when observed over multiple years (Battley 2006; Conklin & Battley 2012; Conklin *et al.* 2013). Similar inter-individual variation in departure times from the Asian stop-over sites grounds has also been reported albeit from a smaller data-set (Conklin *et al.* 2013). Correlations between departure times from N.Z. or Asia and

ultimate breeding latitude suggest that the inter-individual variation in migration dates is correlated with variation in the spring thaw across the Alaskan breeding range, with more northern-breeding godwits both migrating and breeding later than more southern-breeding birds (Conklin *et al.* 2010). Thus, godwits appear to offer an excellent opportunity to associate genetic variation with variation in migration departure time.

The core circadian oscillator (CCO) is generated by a number of genes/proteins that work together form an oscillatory transcription/translation feedback loop (Hastings 2000; Bell-Pedersen *et al.* 2005; Albrecht 2012; Cassone 2014; Hurley *et al.* 2016). One critical element of the CCO is the gene *Clock* (*Circadian Locomotor Output Cycles Kaput*) encoding one half of a heterodimeric transcription factor CLOCK/BMAL1, a transcription-activating complex that regulates the expression of several CCO genes (e.g. Period, Cryptochrome) (Zhang & Kay 2010; Cassone 2014) in addition to some 'output' genes (Chong *et al.* 2000; Ripperger *et al.* 2000; Yoshitane & Fukada 2009; Reischl & Kramer 2011; Rey *et al.* 2011).

An evolutionary conserved feature of CLOCK protein orthologues is a polyglutamine (poly-Q) repeat sequence located towards the C-terminus. Allelic variation in the length of the poly-Q coding sequence has been reported in many taxa along with some evidence that such variation may have functional consequences at the behavioural level (Fidler & Gwinner 2003; Johnsen *et al.* 2007; Liedvogel *et al.* 2009). Therefore, variation in the glutamine repeat length could be a source of quantitative variation of a phenotype (i.e. behaviour) within a population (Darlington *et al.* 1998; Saleem *et al.* 2001; Resuehr *et al.* 2007; Hands *et al.* 2008).

Attempts to associate variation in *Clk*polyQ genotypic variation with breeding or migration-related traits (i.e. migration timing and distance, breeding timing and latitude) have given inconsistent results, with some studies finding support for an association (Johnsen *et al.* 2007; Liedvogel *et al.* 2009; Caprioli *et al.* 2012; Bourret & Garant 2015) and others not (Liedvogel & Sheldon 2010; Dor *et al.* 2011b; Chakarov *et al.* 2013; Kuhn *et al.* 2013; Peterson *et al.* 2015; Bazzi *et al.* 2016c). Analogous studies in fish and flies support the idea that *Clock* allelic variation is associated with latitudinal adaptations (Costa *et al.* 1992; Leder *et al.* 2006; O'Malley & Banks 2008). Although behaviours are typically complex traits with continuous distributions of phenotypic values and presumably polygenic control (Tschirren & Bensch 2010; Liedvogel & Lundberg 2014), genetic polymorphisms can be directly linked with behavioural variation in natural

populations (Easton *et al.* 2003; Fidler *et al.* 2007; Korsten *et al.* 2010; Tschirren & Bensch 2010; Caprioli *et al.* 2012; Mueller *et al.* 2013a; Wetzel *et al.* 2015).

In this study we investigated whether variation in godwit *Clock* gene poly-Q (*Clk*polyQcds) genotype is associated with variation in individual godwit migration departure timing from both N.Z. (over-wintering) and Asian (stopover) sites.

Materials and methods

Collection and storage of godwit blood samples

Godwits were captured by cannon-net or mist-net at two sites in the North Island (Firth of Thames, 37.17˚S 175.32˚E, n = 13; Manawatu River estuary, 40.47˚S, 175.22˚E, n = 68) and at one site on the South Island (Catlins Coast, 46.48˚S, 169.70˚E, n = 54) of N.Z. (Chapter 2, page 33). Blood sampling and storage were done as described in Chapter 2. All sampling were carried out under both N.Z. Department of Conservation (DOC) permits and Animal Ethics Committee approval from Massey University (#07/163, # 12/90) and the University of Otago (#66/03).

Determination of individual chronophenotypes

Determination of chronophenotypes is explained in detail in Chapter 2. Briefly, departure dates of godwits leaving N.Z. were recorded based on daily observations of individually-marked godwits in the three sites previously detailed over the years 2004–16. An individual's chronophenotype is defined as the mean departure date from all the years available. As avian migratory departure timing at temperate latitudes is strongly influenced by photoperiod (Aschoff 1960; Gwinner 1996; Gwinner & Helm 2003) and the three over-wintering sites were distributed across 9º of latitude, we compared predicted photoperiod differences across the sites over the N.Z. migration departure period. As the variation in photoperiod was minimal between sites (maximum of 34 minutes; Appendix 2, Table 2.1) we used Gregorian dates for describing departure dates. We also performed the same analyses using photoperiod (results not shown); results were consistent with those from dates.

Godwit population chronophenotype grouping

The rationale and methodology of this process is described in detail in Chapter 2. Briefly, the correlation between godwit migratory departure date from both N.Z. and Asia and ultimate breeding latitude reported by Conklin *et al.* (2010) was used to group

individuals in discrete chronophenotype groups representing 'earlier' vs. 'later' departers of our datasets. Specifically, godwits departing N.Z. (n = 135) were placed in two groups corresponding to those with chronophenotype (i.e. mean departure date) '< day 81' (n = 88) and with chronophenotype '> day 81' (n = 39). Eight individuals were discarded because their chronophenotype fell on day 81 (Appendix 2, Table 2.2). Godwits departing Asia (n = 32) were grouped corresponding to two chronophenotype divides at day 128, '< day 128' (n = 15) and '> day 128' (n = 17); and at day 135, '< day 135' (n = 23) and '> day 135' (n = 9) (Appendix 2, Table 2.2).

DNA based sex determination

Godwit sex was determined from gDNA samples using methodology of Fridolfsson & Ellegren (1999) which relies on an intron within the Chromo Helicase DNA-binding (CHD) gene differing in length between the CHD alleles carried on the Z (CHD-Z, ~0.6 kb) and the W (CHD-W, ~0.5 kb) chromosomes. Thus male bird gDNA (ZZ) amplifies a single band and female gDNA (ZW) either two bands or, in some cases as ours, a single band shorter than that of males (Fridolfsson & Ellegren 1999). Godwit gDNA was used as the template for the PCR using the primer pair: 2550F (5'–GTTACTGATTCGTCTACGAGA – 3') and 2718R (5'–ATTGAAATGATCCAGTGCTTG – 3') (Fridolfsson & Ellegren 1999). The PCR reaction mix consisted of: 1 x MyTaq^{TM} HS Mix (Cat. No. BIO25045, Bioline, London, U.K.), forward primer (2550F) (0.2 µM), reverse primer (2718R) (0.2 µM), 0.5 µg/µl non-acetylated bovine serum albumin (BSA; Cat. No. B8667, Sigma Aldrich, St. Louis, U.S.A.) with a final volume of 10.0 µl. Thermocycling parameters consisted of: 94˚C/2 min; 94˚C/ 2 min; 60˚C/40 sec., ramping +1˚C/sec. to 72˚C, 72˚C/40 sec., 10 cycles; 94˚C/30 sec, 50˚C/30 sec, 72˚C/35 sec, 30 cycles; 72˚C/5 min.; hold at 4˚C. Amplification products were separated by agarose gel (2.5% (w/v)) electrophoresis and visualized by staining in ethidium bromide before photographing on a UV transilluminator (ChemiDoc^{TM} MP, Imaging System, BioRad, Hercules, U.S.A.). Samples of gDNA generating a single amplicon ~0.5 kb were interpreted as male and a larger amplicon ~0.6 kb as female. Our N.Z. dataset comprises 70 females and 65 males, and our Asian dataset comprises 17 females and 15 males.

A t-test was used to look at differences in migratory departure time between sexes in the N.Z. and Asian datasets. Briefly, we could detect that female godwits departed on their migratory flights earlier than did males both from N.Z. (female mean departure date

= 76, SD = 6.43; male mean departure date = 78 for males, SD = 6.34, $t_{132.5} = -2.22$, $P < 0.05$; Appendix 2, Figure 2.1), and even more so from Asia (female mean departure date = 125; male mean departure date = 135 for males, $t_{25} = -3.44$, $P = 0.002$; Appendix 2, Figure 2.1b). This difference may be a sampling artifact, as there is no difference in N.Z. departure between sexes across all birds for which we have departure dates (P.F.B., unpubl. data). Therefore, there is no *a priori* reason to think that the putative association studied here would appear just in one of the sexes since the *Clock* gene is autosomal. However, natural and sexual selection can affect allele/genotype frequencies differently in males and females (Kissner *et al.* 2003; Ellegren & Parsch 2007; Spottiswoode & Saino 2010; Saino *et al.* 2013; Bazzi *et al.* 2016c), therefore we used assessed the potential association between *Clk*polyQcds and migration departure time not only in the pooled population but also for each sex.

Genotyping of the ClkpolyQcds locus

The godwit *Clk*polyQcds genotyping procedure followed Johnsen *et al.* (2007) in first amplifying, cloning into plasmids and sequencing a range of *Clk*polyQcds alleles that then provided templates to generate size standards for subsequent precise determination of *Clk*polyQcds allele lengths/genotypes. Godwit *Clk*polyQcds sequences were amplified using the generic avian *Clk*polyQcds primers: forward Gen*Clk*For: 5' – TTTTCTCAAGGTCAGCAGCTTGT – 3' and Gen*Clk*Rev reverse: 5' – CTGTAGGAACTGTTGYGGKTGCTG – 3' (Johnsen *et al.* 2007) with reaction conditions 1x BioMix (BioLine, London, U.K.), 0.8 µM both forward and reverse primers, 1.0 µL gDNA, water to a final volume of 50.0 µL with thermocycling conditions: 94°C/2 min; 94°C/30 sec, 60°C/30 sec, 72°C/60 sec, 10 cycles; 94°C/30 sec, 64°C/30 sec, 72°C/60 sec, 30 cycles; 72°C/7 min; hold at 15°C. The resulting PCR products were visualized on an ethidium bromide stained agarose gel (2.0 % (w/v)), extracted using a commercial kit (Zymoclean™ Gel DNA Recovery Kit, Zymo Research, Irvine, U.S.A.), ligated into the T-tailed cloning vector pGEM-Teasy (Promega, Madison, U.S.A.) and transformed into competent DH5α *E. coli* cells. Plasmids were purified (High Pure Plasmid Isolation Kit, Roche Diagnostics, Penzberg, Germany) and their inserts sequenced by an external contractor (Massey Genome Service, Massey University, N.Z.). From six godwit genomic DNA samples two distinct *Clk*polyQcds allele sequences were found and denoted *Clk*polyQ$_9$ and *Clk*polyQ$_{11}$ on the basis of variation in the length of a poly-glutamine (poly-Q) coding region (Figure 3.1). Using the *Clk*polyQ$_9$ and *Clk*polyQ$_{11}$

allele sequences godwit specific *Clk*polyQcds PCR primers were designed for further genotyping, forward primer *LimClk*For: 5'– TGTAAAACGACGGCCAGTTGGGACAGGTGGTGACGGCTTAC – 3'; reverse primer *Lim*ClkRev: 5' – GTTTCTGCTGAACGGTGGTGAG – 3'. The godwit *Clk*polyQcds sequences correspond to coordinates 63–85 and 217–200 of *Clk*polyQ$_{11}$ (GenBank acc. no.: JN676983). An 18 bp generic M13 'tag' sequence (5'– TGTAAAACGACGGCCAGT – 3') was included on the 5' end of *LimClk*For to allow florescent labelling of amplicons using the three primer PCR strategy of Schuelke (2000). A 5' 'PIG tail' sequence (5'– GTTT –3') was included on *LimClk*Rev in an effort to enhance consistency in amplicon size by promoting the addition of 3' A's by the *Taq* DNA polymerase (Brownstein *et al.* 1996). Godwit *Clk*polyQcds genotyping PCR reaction mixes consisted of: 1x MyTaq^{TM} HS Mix (cat. no. BIO25045, Bioline, London, U.K.), *LimClk*For (0.2 µM), *LimClk*Rev (0.6 µM;), and a florescent dye-labelled generic M13-tag primer (5' – TGTAAAACGACGGCCAGT – 3') (0.6 µM) Schuelke (2000), with thermocycling conditions: 94˚C/2 min; 94˚C/30 sec, 59˚C/30 sec, 72˚C/30 sec, 15 cycles; 94˚C/30 sec, 54˚C (ramping +0.2˚C/sec)/30 sec, 72˚C/30 sec, 25 cycles; 72˚C/5 min; 60˚C/30 min; 15˚C/hold. Amongst 20 godwit gDNA samples screened 6 were identified that collectively displayed 5 different *Clk*polyQcds length variant alleles, either in heterozygous or homozygous conditions. Using the generic *Clk*polyQcds primer pair of Johnsen *et al.* (2007) (i.e. Gen*Clk*For and Gen*Clk*Rev) the five different *Clk*polyQcds allelic sequences were amplified and sequenced (Figure 3.1). Five plasmids corresponding to the five *Clk*polyQcds alleles *Clk*polyQ$_{8\text{-}12}$, were then used as templates for the PCR with primers *LimClk*For and *LimClk*Rev in combination with 6-FAM™ labelled M13 tag primer. The resulting amplicons were diluted and pooled to generate 6-FAM™ labelled standards for each of the five *Clock*polyQcds alleles. To genotype the full set of godwits (n = 135) their corresponding gDNA samples were used as templates with the generic M13 tag primer labelled with either VIC® (green) or NED™ (yellow). Amplicon lengths were determined by an external contractor (ABI3730 Genetic Analyzer, GeneScan™-500 LIZ™ size standard, Massey Genome Service, Massey University, Palmerston North, N.Z.) for amplicon length estimations consisted of pools of: (i) the 6-FAM™ labelled standards (ii) amplification products from one godwit gDNA labelled with VIC® and (iii) amplification products from one bird gDNA labelled with NED™. Peak Scanner™ v2.0 (Life Technologies, Carlsbad, U.S.A.)

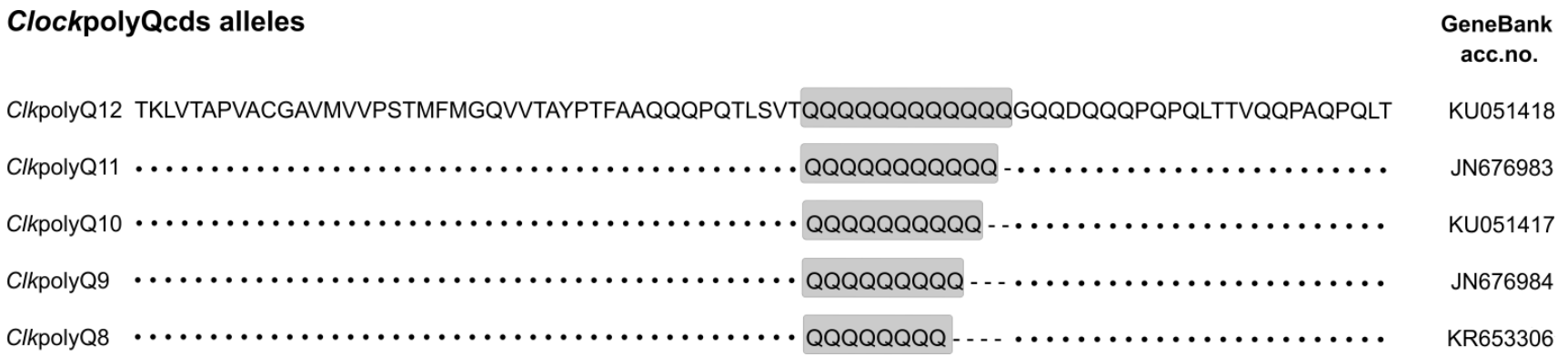


Figure 3.1: Alignment of predicted protein sequences corresponding to the five godwit *Clk*polyQcds alleles sequenced in this study. An exon of the godwit Clock gene orthologue containing a poly-glutamine (poly-Q) coding region was amplified and sequenced. Allele names denote the number of glutamine (Q) residues predicted to be encoded in a variable length region (shaded). Positions of predicted protein identity are indicated by dots (•) with gaps introduced to the poly-Q region to achieve alignment indicated by (-). GenBank accession numbers for the five allelic sequences are: *Clk*polyQ$_8$ (KR653306), *Clk*polyQ$_9$ (JN676984), *Clk*polyQ$_{10}$ (KU051417), *Clk*polyQ$_{11}$ (JN676983) and *Clk*polyQ$_{12}$ (KU051418).

was used to view and analyze the electropherogram data.

In preliminary experiments we found, as reported by Sutton *et al.* (2011), that the dye PET™ adds an apparent 3 bp to the estimated length of a DNA fragment when compared with FAM™, VIC® and NED™ labelling (data not shown). Therefore, only the dyes FAM™ (for the standards) and VIC® and NED™ (for genotyping) were used in this study. Repeat genotyping of three gDNA samples with either VIC® or NED™ labelled generic M13 tag primers returned the same genotype with examples of the genotyping electropherogram shown in Appendix 2, Figure 2.2. Therefore, it was concluded that genotypes generated using FAM™, VIC® and NED™ labelled generic M13 tag primers could be combined and compared.

Statistical analyses

General parameters (He, Ho, Na, allele frequencies) were calculated for the pooled dataset using GENEAIEX v.6.501 (Peakall & Smouse 2012). We tested for departure from HWE (Hardy-Weinberg equilibrium) using ARLEQUIN v.3.5 (Excoffier *et al.* 2005) with 1,000,000 steps in Markov chain and 1,000,000 dememorization steps. We checked for normality of poly-Q unit frequency distributions for each sex with a Shapiro-test. Differences between sexes in allele length for either locus were assessed with a t-test. *Clk*polyQcds genotype frequencies between sexes were compared with a Kruskal-Wallis test (non-parametric). Since we don't have *a priori* knowledge about the genotype-phenotype relationship (i.e. dominance, co-dominance, partial-dominance) we ran replicated analyses for the length of the shorter allele, length of the longer allele and mean length of the two alleles. We calculated Kendall's correlation (tau) (Kendall 1948) and the correspondent p-values (two-tailed) to test an association between allele length (either the shorter, longer and mean polyQ allele length) and migration departure time for the pooled population, and for each gender. All statistical analyses were performed in R v.3.3.0 (R Core Team 2014) and plots were generated in SIGMAPLOT v.10.0. Finally, to see whether *Fst* values from the polyQ genotypes were similar to those from putatively neutral regions of the genome, we compared the *Fst* value from the *Clk*polyQcds locus to those from microsatellites. For that, we used genotypes of 27 and 30 microsatellites (that is, including and excluding the 3 microsatellites that did not conform HWE) and calculated AMOVA-*F*st for the N.Z. (n = 135) and the Asian (n = 32) datasets classified into 'earlier' vs. 'later' *a priori* groups, and we calculated *F*st and their correspondent p-values for the functional *Clk*polyQcds locus. AMOVA-*F*st calculations were done in

ARLEQUIN v.3.5 and plots were created using SIGMAPLOT v.10.0. We corrected for multiple testing using Bonferroni approach (Rice 1989) ($p < 0.05$). We indicate both p-values uncorrected and corrected (Bonferroni) as the Bonferroni correction is quite conservative and can lead to false negatives (Perneger 1998; Narum 2006).

Results

Clock polyQ genotyping and general parameters

We successfully amplified the *Clk*polyQcds length region from 135 individuals. Sequences of standard size fragments (Q_7–Q_{12}) were identical except for the number of poly-Q repeats (Figure 3.1). Average observed heterozygosity was 0.79 (females = 0.77, males = 0.80). The most common alleles were Q_9 (36%), Q_{11} (24%) and Q_{10} (22%) (Table 3.1). Genotype frequencies did not deviate from Hardy-Weinberg equilibrium for sexes combined and separate (all $P > 0.8$). Molecular sexing indicated slightly more females (52%) than males (48%). *Clk*polyQcds allele and genotype frequencies were not different between two sexes (Table 3.1).

Table 3.1: Summary statistics of *Clk*polyQcds allele frequencies amongst godwits over-wintering in New Zealand. Note that eight individuals were not classified in any of the *a priori* groups as their chronophenotype fell at day 81. *Abbreviations*: n = number of individuals; *k* = number of alleles; Ho = observed heterozygosity.

Dataset	*n*	*k*	Ho	Q7	Q8	Q9	Q10	Q11	Q12
Total New Zealand	135	6	0.79	0.07	0.08	0.36	0.22	0.24	0.03
Females	70	6	0.77	0.06	0.08	0.36	0.21	0.26	0.04
Males	65	6	0.80	0.08	0.07	0.37	0.24	0.22	0.02
< Day 81 ('earlier departers')	88 (53♀/35♂)	6	0.75	0.10	0.06	0.39	0.20	0.22	0.03
> Day 81 ('later departers')	39 (13♀/26♂)	6	0.85	0.03	0.10	0.29	0.29	0.26	0.03

*Clk*polyQcds polymorphism and migratory departure dates

We summarized the *Clk*polyQcds genotypes as the mean of both *Clk*polyQcds alleles, the length of the longer *Clk*polyQcds allele, the length of the shorter *Clk*polyQcds allele or simply the *Clk*polyQcds genotype (Figure 3.2). No statistically significant relationship between any of the measures of *Clk*polyQcds genotype and migratory departure dates was detected, with the sexes analyzed separately or together (Figure 3.2, Table 3.2).

Table 3.2: Kendall's test (tau and two-sided *P*-values) results testing the correlation between the mean, longer and shorter *Clk*polyQcds alleles with chronophenotype for the N.Z. and Asian datasets with sexes separately and pooled.

Dataset	Subpopulation	*Clk*polyQcds genotype measure	tau	*P*
N.Z. departure	females (n = 70)	Mean	0.12	0.19
		Longer	0.08	0.38
		Shorter	0.09	0.32
	males (n = 65)	Mean	0.14	0.16
		Longer	0.16	0.11
		Shorter	0.12	0.23
	females + males (n = 135)	Mean	0.08	0.19
		Longer	0.07	0.28
		Shorter	0.09	0.18
Asia departure	Females (n = 17)	Mean	0.10	0.66
		Longer	0.21	0.32
		Shorter	0.17	0.41
	Males (n = 15)	Mean	0.28	0.22
		Longer	0.19	0.38
		Shorter	0.12	0.61
	females + males (n = 32)	Mean	0.16	0.24
		Longer	0.15	0.30
		Shorter	0.17	0.24

Population genetics of the ClkpolyQcds locus

The *F*st value for the N.Z. dataset grouped into '< day 81' vs. '> day 81' was not significant with 27 or 30 microsatellites (both *F*st = 0.003; *P* (Bonferroni) < 0.10; Interval of Confidence (IC) = 90%, Table 3.3; Appendix 2, Figure 2.3). For the Asian dataset,

population structure was shown by the set of microsatellites, but not by *Clk*polyQcds, between birds departing before versus after day 128 (*F*st = 0.012–0.014, *P* = 0.02, *P* (Bonferroni) = 0.05; IC = 90%). However, a marginally significant *F*st based on *Clk*polyQcds was found for birds departing before versus after day 135 (*F*st = 0.054; *P* = 0.075), which is not significant after Bonferroni correction (*P* = 0.225) (Table 3.3; Appendix 2, Figure 2.3). Using 27 microsatellites this comparison was significant (*F*st = 0.014; *P* = 0.034, *P* (Bonferroni) = 0.068), however it was not when using 30 microsatellites (*F*st = 0.010; *P* = 0.093, *P* (Bonferroni) = 0.128, Table 3.3; Appendix 2, Figure 2.3).

Table 3.3 Genetic differentiation of godwit sub-population timing groupings based on departure date divides. Using migration departure dates from N.Z. and Asia godwit were grouped into 'earlier' and 'later' departers groups, see text (Materials & Methods) for details. F_{ST} values were calculated between the groups indicated using the *Clk*polyQcds locus alone or microsatellite loci. In none of the three comparisons did the F_{ST} values calculated from the *Clk*polyQcds allele frequencies reach statistical significance. Significant F_{ST} values without or without Bonferroni correction are indicated in bold ($P < 0.05$). Marginally significant F_{ST} values with and without Bonferroni correction are underlined ($P < 0.08$). Numbers of genotypes in the groupings compared are indicated in parentheses *Abbreviations*: I.C. = interval of confidence.

Dataset	Dividing departure date	Loci	*Fst*	*P*-value	*P*-value (Bonferroni)	I.C. (%)
N.Z.	**Day 81**	27 microsatellite	0.003	0.102	0.102	90
	(88/39)	30 microsatellite	0.003	0.064	0.128	90
		*Clk*polyQcds	0.009	0.092	0.225	-
Asia	**Day 128**	27 microsatellite	**0.014**	**0.017**	0.051	90
	(15/17)	30 microsatellite	**0.012**	**0.031**	0.093	90
		*Clk*polyQcds	0.002	0.394	0.128	-
	Day 135	27 microsatellite	**0.014**	**0.034**	0.068	95
	(23/9)	30 microsatellite	0.010	0.093	0.128	90
		*Clk*polyQcds	0.054	0.075	0.225	-

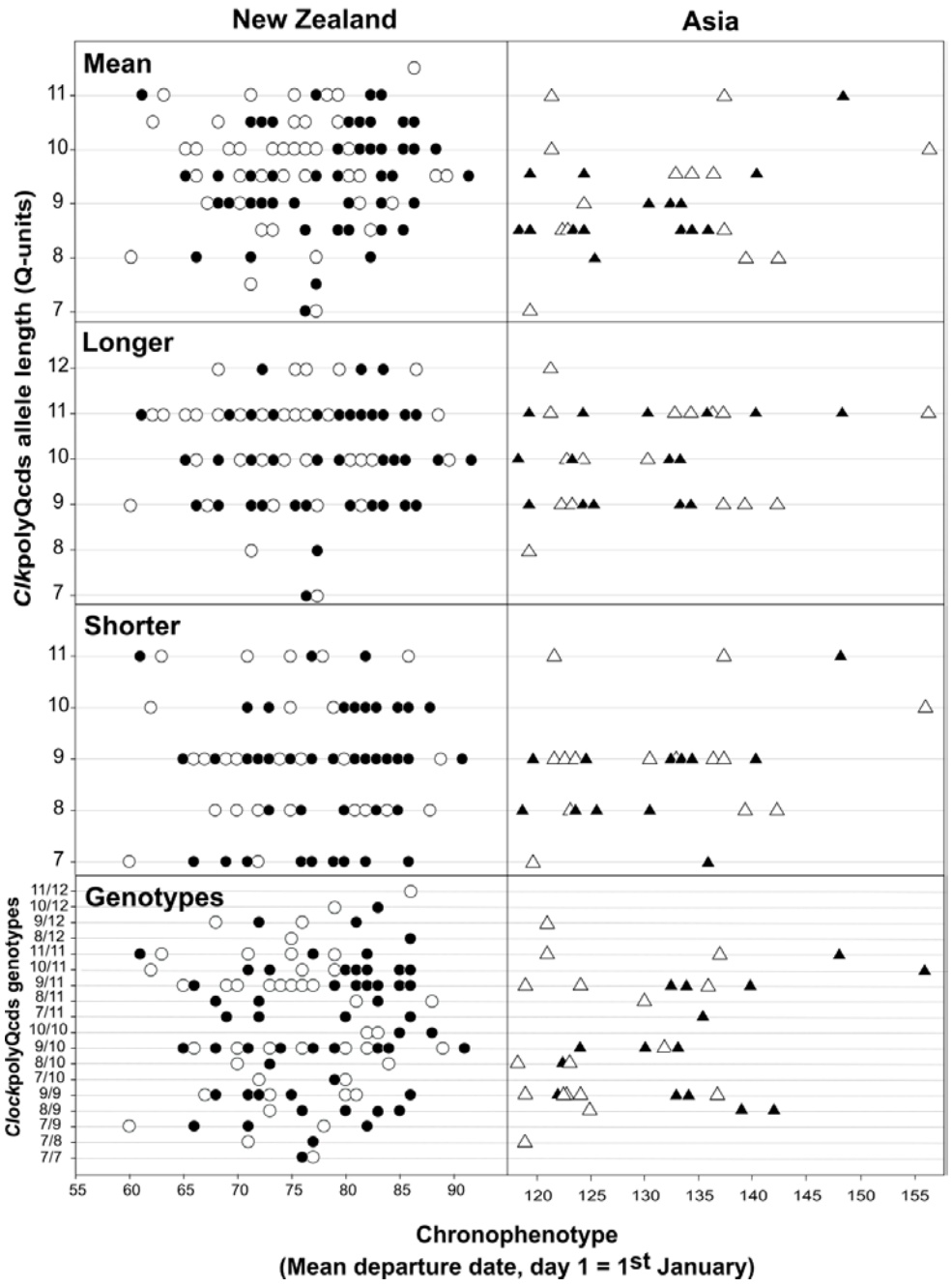


Figure 3.2: Relationships between godwit mean migratory departure dates and *Clk*polyQcds genotype. The *Clk*polyQcds genotypes of individual female and male godwits are summarized, as mean allele length, length of longer allele, length of shorter allele and actual genotype (allele combination). Mean migratory departure dates of individual godwits from N.Z. and Asia are given as Gregorian calendar date (day 1= 1 January). See materials and methods for details of how mean migratory departure dates were obtained and calculated. N.Z. departure dataset consists of 135 godwits (70 female (○), 65 male (●)); Asia departure dataset consists of 32 godwits (17 female (Δ), 15 male (▲)). No association between *Clk*polyQcds genotype and migration departure date was detected.

Discussion

New Zealand-wintering bar-tailed godwits provide an appropriate system in which to test for genetic influences on migration timing, as the measure of phenology is clear and unequivocal (the date that major trans-oceanic migratory flights are embarked upon), and is variable between individuals yet strongly consistent within individuals. We found that godwits were highly polymorphic at the *Clk*polyQcds (heterozygosity 0.79) with a wide range of genotypes so that relationships are unlikely to be strongly influenced by outliers. Despite this, we found no evidence for a relationship between *Clk*polyQcds and migration timing in godwits.

ClockpolyQcds polymorphism in the bar-tailed godwit

We identified six alleles within the 135 bar-tailed godwits, which is very similar to what previous studies reported from other species (Johnsen *et al.* 2007; Liedvogel *et al.* 2009; Liedvogel & Sheldon 2010; Dor *et al.* 2011a; Dor *et al.* 2011b; Caprioli *et al.* 2012; Chakarov *et al.* 2013; Kuhn *et al.* 2013; Bazzi *et al.* 2015; Saino *et al.* 2015; Stuber *et al.* 2016) (Appendix 2, Table 2.3). The number of poly-Q repeats in godwits ranged from Q_7 to Q_{12}. Previous studies found poly-Q alleles from Q_5 to Q_{16} (Johnsen *et al.* 2007; Dor *et al.* 2011a); therefore, godwit poly-Q repeats are within the range already reported. Heterozygosity in godwits (0.79) was the highest within those reported in previous studies (Appendix 2, Table 2.3) ranging from monomorphic in sedge warbler (*Acrocephalus schoenobaenus*), reed warbler (*Acrocephalus scirpaceus*) and European bee-eater (*Merops apiaster*), to 0.64 in blue tit (*Cyanistes caeruleus)* (Dor *et al.* 2011b; Saino *et al.* 2015; Bazzi *et al.* 2016b), indicating high diversity in poly-Q content in godwits. Bazzi *et al*. (2016a) found in trans-Saharan migrants (mainly passerine) that long-distance migration was associated with low gene diversity, which they interpreted as reflecting tighter migratory timing constraints. Bar-tailed godwits do not match this pattern as they are an extreme long-distance migrant yet with high *Clock* diversity (using the methodology of Bazzi *et al.* (2016a): $\hat{h} = 0.75$), suggesting that their observation of low *Clock* allelic diversity in long-distance passerine migrants is not generalizable.

Departure time, ClkpolyQcds and sexes

Our study found no significant differences in poly-Q content or genotype frequencies between sexes as expected of an autosomal locus. However, we obtained higher diversity of poly-Q genotypes in females than in males. Only one previous study compared the poly-Q polymorphism potential association with breeding time and they

found weak and positive association only in females (Liedvogel *et al.* 2009). Another study detected an inconsistent (two out of four migratory species) association between departure date for migration from a stop-over with the mean and with the longer *Clock* poly-Q allele (Saino *et al.* 2015). A more recent study (Bazzi *et al.* 2016c) reported no association between *Clock* poly-Q and spring migration timing in a long-distance migratory passerine (*Phylloscopus trochilus*), although they found another gene (i.e. *Npas2*) predicting spring migration timing significantly only in females. To date, there is no consensus about whether *Clk*polyQcds polymorphism has an adaptive function for only one of the sexes (Bazzi *et al.* 2016c).

Variation at ClkpolyQcds and migratory timing behaviour

We found no statistical support for an association between *Clk*polyQcds alleles or genotypes and the continuum of chronophenotypes for either the N.Z. or the Asian datasets, although later-migrating godwits tended to have longer alleles (Figure 3.2). We found a marginally significant – without Bonferroni correction $P = 0.075$ – *Clock* poly-Q genetic difference (*F*st) using the Asian dataset, when birds were divided into those departing before ('earlier departers') vs. after ('later departers') day 135. However, we also calculated AMOVA-*Fst* values for each microsatellite and compared them with that obtained for *Clk*polyQcds (Appendix 2, Figure 2.4) and found that *Clk*polyQcds *F*st was surpassed by some of the microsatellite *F*st values. Hence, we cannot rule out the marginally significant *Clk*polyQcds result from the Asian dataset at the day 135 divide being a product of the genetic background, reflecting demographic events, i.e. non-random mating rather than differential selection on *Clock* across the breeding range (Chapter 2). In conclusion, we do not have a clear signal for *Clk*polyQcds being associated to migratory departure time behaviour in godwits.

Similarly, despite some strong claims of a real association between migratory timing and *Clock* poly-Q polymorphism, these findings are arguably marginal and inconsistent. Bazzi *et al*. (2015), for instance, claimed that certain rare genotypes were associated with specific migratory schedules in barn swallows (*Hirundo rustica*), but due to these genotypes being rare, any relationships are inevitably driven by just a few individuals while the vast majority of the population showed no variation in *Clock*. Similarly, Saino *et al.* (2015) claim (in their title) that clock polymorphism "predicts phenology of long-distance migration in birds", yet no relationships were detected for two of the four species they studied, and in the remaining two the relationship was found

only with one allele type (longer, but not mean length: nightingale, *Luscinia megarhynchos*) in one sex only and potentially influenced by a few rare genotypes (tree pipit, *Anthus trivialis*). It could be argued that there is yet no convincing association at the individual level between *Clock* variation and migration time (see Appendix 2, Table 2.3 for a summary of *Clock*-migratory traits findings on birds).

Nevertheless, given the function attributed to the *Clock* gene as a core element in biological rhythms, *Clk*polyQcds may yet be related to other time-regulated traits such as moulting or breeding. Other studies have found some support for variation in *Clock* being associated with breeding time, though again results are variable between species (Liedvogel *et al.* 2009; Caprioli *et al.* 2009; Saino *et al.* 2013; Bourret & Garant 2015; but see Liedvogel & Sheldon 2010; Bazzi *et al.* 2016c).

Other factors and limitations that might have affected the results of this study

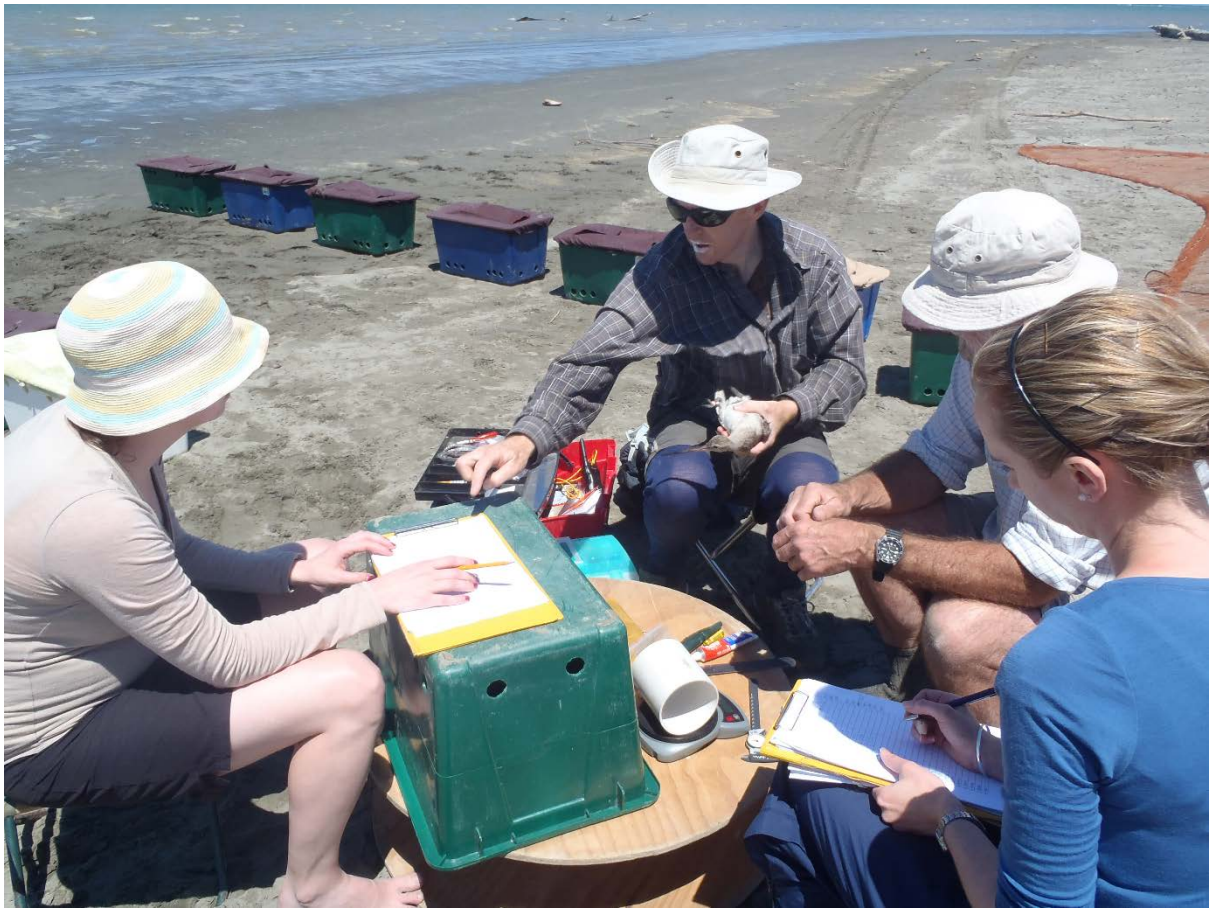
Migratory departure date is influenced not only by the internal "clock" but also by environmental conditions (Marra *et al.* 2005; Bauer *et al.* 2008; Bourret & Garant 2015). Over a decade of observations at the over-wintering sites during the departure period for migration have confirmed that N.Z. has, in general, favorable conditions with unsuitable winds typically lasting for only a few days in sequence (e.g. Conklin & Battley 2011). Therefore, we believe that environmental variation influences an individual's decision on when to start its migration journey on a scale of days, rather than weeks. There are some suggestions that migrants may fine-tune their migration (Mettke-Hofmann & Gwinner 2003; Chernetsov *et al.* 2004; Thorup *et al.* 2007; Mueller *et al.* 2013b; Evens *et al.* 2017), with first-time northward migrants having lower repeatability in migration departure date than adults (Battley 2006; Battley *et al.* 2011). This suggests that the experience that an individual may gain after years of completed migratory cycles could blur the strength in which timing-associated genes could be detected (Berthold 2001). A significant improvement in our study would be using immature individuals' first departure dates for migration, as these may be more genetically driven than subsequent migrations in which experience may play a role. We expect, however, that the genetic influence on an individual migratory behaviour remains determinant in the adult's average departure date (i.e. an immature early chronophenotype would become an adult early chronophenotype).

Conclusions

Our results indicate that *Clock* poly-Q polymorphism is not associated with the timing of northwards migration in bar-tailed in godwits. It adds to a growing list of studies that have failed to establish a link between *Clock* and migration timing. However, we limited our analysis to a single trait; other traits under circannual control (i.e. fat storage, phenology of moult, breeding time) might be good candidates for future studies. We think that the most plausible explanation for these discordant results between all the studies on *Clock* polyQ is that migratory timing is indeed a complex trait governed by numerous genes (polygenic) in which *Clock* variability seems not to contribute substantially or consistently to the observed chronophenotype variability (Pulido & Berthold 2003; Bazzi *et al.* 2016b; Hess *et al.* 2016). Phenotype-genotype studies are complex since an individual phenotype assignation can be influenced by many factors (i.e. environment), and it could potentially change according to a particular environment (i.e. phenotype plasticity). Another alternative explanation is that poly-Q may not play a role in departure time decisions for migration at all. Finally yet importantly, genetic components are not the only molecular-related sources of behavioural variations: epigenetic-phenotype associations are still in their early stages of being investigated, thus, with the potential to play an important role in future discovering (Crews 2011; Powledge 2011; Liedvogel & Lundberg 2014; Baerwald *et al.* 2016). Indeed, a recent study found evidence of methylation level at the *Clk*polyQcds playing an important role regulating individual variation in migration timing (Saino *et al.* 2017). To what extent epigenetics is involved in determining variation in migratory timing of individuals and populations across taxa is not yet sufficiently explored, but is a promising avenue for future work.

Chapter 4

Polygenic architecture underlies departure time in a long-distance migratory bird



Phil Battley, David Melville and two volunteers working together to ring and take measurements of captured godwits at the Manawatu River estuary (November, 2013).

Ángela M. Parody-Merino, Phil F. Battley, Andrew E. Fidler, Francisco Prosdocimi, Nicholas C.B. Lima, Jesse R. Conklin, Murray A. Potter, Murray P. Cox

Abstract

Describing mechanistic links between genetic variation and behaviour-level variation in free-living animals presents a major scientific challenge. Migratory bar-tailed godwits that breed in Alaska and over-winter in Australasia perform the longest recorded avian non-stop migratory flight, directly from Alaska to New Zealand (N.Z.), 12,000 km away. The control of avian migration timing is believed to involve changes in daylength as perceived by the birds, but this insight alone does not explain the observed inter-individual variation in different departure schedules. It is anticipated that genetic variation is causally and mechanistically associated with such phenotypic variation. To investigate this possibility, we looked for associations between individuals' migration departure time (chronophenotypes) and genotypic variation at loci that form elements of the endogenous circadian clock, of the Hypothalamic-Pituitary-Gonadal axis, of photoreception, and fat metabolism. A reference bar-tailed godwit genome was used to identify the candidate genes and 19 additional genomes which served to prioritize 4,919 SNPs that were genotyped on 265 individuals. 3,412 SNPs (harboring 120 candidate genes) passed our filtering parameters. Our assessment of the genetic architecture underlying migratory departure time together with a population structure analysis suggested that this phenotype has a highly polygenic basis. Although we did not identify any strong association of any specific candidate genes with an 'earlier' or 'later' chronophenotype, we found that 124 SNPs can explain significantly around 10–36% of it. Whether this finding is principally founded on a smaller number of SNPs is something that deserves further investigation.

Introduction

The annual rhythms of migratory birds are governed by an endogenous program, with external clues governing its arrangement in time (i.e. synchronizing it to external seasonal-related events) (Berthold & Helbig 1992; Gwinner 1996; Visser *et al.* 2010; Mueller *et al.* 2016). A growing number of studies aim to identify the genetic architecture, including genes and pathways, that shape this 'internal circannual clock', both to better understand how it works (Lundberg *et al.* 2017) and to help predict the adaptability of species to climate change (Barton & Keightley 2002; Pulido 2007b; Hut & Beersma 2011; Helm *et al.* 2013). While some elements have been identified (e.g. hormones, circadian genes), how genetic variation mediates timing in natural populations remains far from clear (Liedvogel *et al.* 2011).

Bar-tailed godwits (*Limosa lapponica baueri*, godwits hereafter) of the sub-species *baueri* are shorebirds that perform one of the most extraordinary annual migrations seen in nature. From February to April, godwits leave New Zealand (N.Z.) enroute to their breeding grounds in Alaska, stopping for 5–7 weeks in eastern Asia to refuel (Gill *et al.* 2009; Battley *et al.* 2012). When the boreal winter starts, they fly directly from Alaska to the southern hemisphere summer in N.Z. in what is the longest non-stop journey ever recorded in birds (Battley *et al.* 2012). Surveys of up to 10 years at three sites show striking consistency in the dates when individual birds leave N.Z. (Figure 4.1) (Battley 2006; Conklin *et al.* 2013). Within N.Z., godwits comprise a single genetic population as estimated from 27 microsatellite markers (Chapter 2), even further increasing the power of genotype-phenotype association studies (Rönnegård *et al.* 2016).

Zugunruhe, the restless behaviour that migratory birds exhibit when they are 'internally' ready to migrate (Berthold *et al.* 2000; Helm & Gwinner 2006), is known to be influenced by circadian/circannual mechanisms and by changes in photoperiod (Helm & Gwinner 2006). A key set of genes known as the Core Circadian Oscillator (e.g. *clock*, *bmal1*, *per2*, *per3*, *cry1*, *cry2*) are frequent targets of study (Cassone 2014). The Hypothalamic-Pituitary-Gonadal axis (e.g. DIO3, TSHβ, TRH) also modulates the internal clock (Yasuo *et al.* 2003; Csernus *et al.* 2004; Bell-Pedersen *et al.* 2005; Csernus 2006; Racz *et al.* 2008; Yoshimura 2010; Greives *et al.* 2012; Gwinner *et al.* 2012; Cornelius *et al.* 2013; de Miera *et al.* 2013; Herwig *et al.* 2013; Dardente *et al.* 2014), likely by affecting how birds perceive changes in photoperiod (Cassone 2014). Roles have also been proposed for genes related to photoreception (i.e., opsins and the melatonin pathway) (Dawson *et al.* 2001; Ramenofsky *et al.* 2003; Halford *et al.* 2009; Wyse & Hazlerigg 2009; Ikegami & Yoshimura 2012), and fat metabolism and storage (Stuber *et al.* 2013; Goymann *et al.* 2017).

Here, we explore candidate genes in these and other potential pathways, as well as a broader panel of genome-wide genetic diversity, to begin to identify the genetic basis of departure time for migration in a natural population of wild shorebirds. By associating genetic variation with longitudinal data on individual departure times in bar-tailed godwits, we aim to determine whether departure time is controlled by many genes working in concert (polygenic adaptation) or a select number of highly influential genetic triggers.

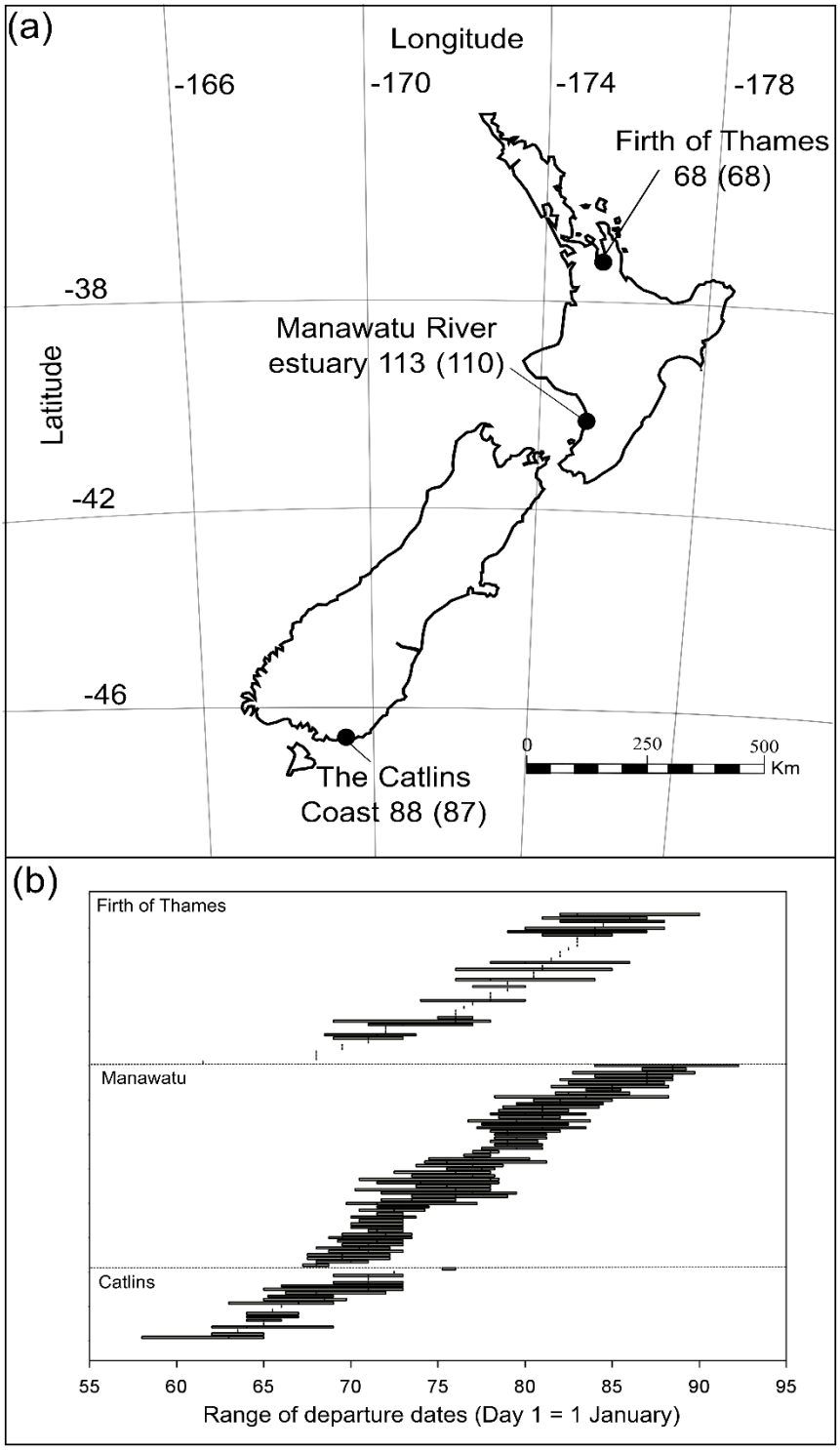


Figure 4.1: (a) Geographical location of the three N.Z. sites where godwits blood samples and departure dates were obtained. Number of sampled individuals are indicated, with the number in parenthesis being the final number of individuals used for the analyses. (b) Individuals' departure data spans at the three sites across multiple years. Bars correspond to a single individual. Note that, to facilitate the visualization of this graph, only individuals with ≥ 2 years (for Firth of Thames and Catlins) and ≥ 4 years (for Manawatu) of recorded departure dates were included.

Materials and methods

Sample collection, DNA extraction

Cannon nets or mist nets were used to trap godwits from 2004–2016 in N.Z. at two sites in the North Island: Firth of Thames, 37.17˚S 175.32˚E, n = 68; Manawatu River estuary, 40.47˚S, 175.22˚E, n = 113; and at one site on the South Island: Catlins Coast, 46.48˚S, 169.70˚E, n = 88 (total = 269 individuals). Blood samples were collected using microhaematocrit capillary glass tubes from the metatarsal or brachial vein and preserved in either 96% (v/v) ethanol or Queen's Lysis buffer (QLB) (10 mM Tris, 10 mM NaCl,10 mM EDTA, 1.0% (w/v) *n*-lauroylsarcosine, pH 8.0) (Seutin *et al.* 1991) before long-term storage at ambient temperature. Godwit capture and blood sampling were carried out under both N.Z. Department of Conservation (DOC) permits and Animal Ethics Committee approval from Massey University (#07/163, # 12/90) and the University of Otago (#66/03). Genomic DNA was extracted from the blood samples using a commercial kit following the manufacturer's protocol (Quick-gDNA^{TM} MiniPrep, cat. no. D3050, Zymo Research, Orange, U.S.A.).

Phenotyping: Determination of individual chronophenotypes

Determination of migratory departure dates is explained in detail in Chapter 2). Briefly, departure dates of godwits leaving N.Z. were recorded based on daily observations of individually-marked godwits in the three sites previously detailed over the years 2004–17. Migration departure dates were assigned by designating 1 January as day 1. Mean and standard deviation (SD) were calculated for each individual, confirming that individuals' departures were consistent within a week (mean SD <= 8) – except for one bird: MY12701, SD = 9.2 – along the 10 years of recorded data. We standardized phenotypes to account for year-to-year variation by the mean and standard deviation of departure date each year and per site (Appendix 3, Table 3.1). The actual standardized phenotypes will, hereafter, be referred to as 'longitudinal standardized chronophenotype'. Since most of the analyses require a unique phenotype value per individual (not allowing for longitudinal data/ repeated measurements of phenotypes along several years) we used the mean of all departure dates available as the measure of an individual's chronophenotype, hereafter referred as 'mean standardized chronophenotype' (Appendix 3, Table 3.1). Note that recorded departure dates from Manawatu (452 observations) are more accurate (and abundant) than those obtained from the Firth of Thames (131

observations) and Catlins (168 observations). For this reason, we repeated all the statistical analyses using two different datasets: all sites and only Manawatu alone. Conclusions from these two parallel analyses did not differ, therefore, for simplicity, here we report results using all sites. Last, the Catlins subpopulation, which is the only subpopulation from the South Island and the most recent to be tracked (since 2013), has a shifted phenology – by 11days earlier on average – than those of Manawatu and Firth of Thames (unpublished data). This was accounting for by adding *site* as covariate in all association tests.

Additionally, 34 individuals from Manawatu carried geolocators, enabling the determination of departure dates (1 or 2 years of departure date) from the stop-over in Asia. Tracking was conducted in 2008, 2009, 2013 and 2014, and the method used to determine departure dates is explained in detail in Chapter 2. Briefly, the geolocators recorded conductivity (i.e. marine wetness) every 10 minutes to 4 hours depending on the brand and model so the date of dramatic change in conductivity was used to determine departure date from Asia (see (Battley & Conklin 2017). These individuals and their mean departure dates from Asia are included in Appendix 3, Table 3.1.

Godwit de-novo genome assembly

A female godwit was used to create a reference godwit genome. Its genomic DNA was extracted from blood using a commercial kit following the manufacturer's protocol (Quick-gDNA^{TM} MiniPrep, cat. no. D3050, Zymo Research, Orange, U.S.A.). Sequencing was done in Illumina HiSeq2000 (DHMRI, USA). Three paired-end libraries with average fragment sizes of 170 bp, 500 bp and 800 bp were generated, in addition to one mate-pair library with an average fragment size of 4 Kbp. The library with 170 bp inserts was run on a single lane; the libraries with 500 bp and 800 bp were pooled at an equimolar ratio, then run on a single lane; the mate pair library was run on a single lane. The quality of the reads was checked using FastQC (Andrews 2010). Sequences were selected by quality and trimmed when they showed adaptors using Trimmomatic (Andrews 2010). Only the reads with mean quality (Q) higher than 30 and with a length higher than 70 bp were retained. A Q score reflects that the probability of an incorrect base call is 1 in 1000 for an inferred base call accuracy rate of 99.9%. All the libraries were used in the assembly using AllPaths-LG (Gnerre *et al.* 2011).

Candidate genes

We used a Candidate Gene Approach (CGA), therefore genes were chosen based on literature searches known to have a role as elements of the endogenous circadian clock, and its outputs, components of the Hypothalamic-Pituitary-Gonadal (HPG) axis, and photoreceptors. We also included other genes that have been described as playing a role in fat metabolism, seasonal modification of the brain physiology, protein degradation (Virshup & Forger 2007), transcription factors and chromatin modellers (see complete list of candidate genes and references in Appendix 3, Table 3.2).

We used gene and protein orthologues sequences of candidate genes extracted from chicken and zebra finch genomes from two databases: NCBI (https://www.ncbi.nlm.nih.gov/), and Ensembl (http://www.ensembl.org/index.html). Then we used BLASTN and BLASTP to annotate predicted genes in the godwit genome. We included +/- 1Kb of the flanking regions, therefore hereafter, candidate genes will be referred as candidate gene regions, which comprise the sequence of the candidate gene (i.e. exons and introns) and its +/- 1Kb flanking regions.

Whole genome re-sequencing

With the aim to facilitate the selection of the final SNPs to be genotyped in the whole dataset, as well as to add data to the final reference godwit genome assembly, 19 unrelated individuals were selected from the extremes of the behavioural continuum (Appendix 3, Table 3.3). Genomic DNA was extracted from blood samples using a commercial kit following the manufacturer's protocol (Quick-gDNA^{TM} MiniPrep, cat. no. D3050, Zymo Research, Orange, U.S.A.). Four paired-end libraries of insert size 500 pb were prepared and sequencing was performed on a HiSeq2000 (Illumina) with New Zealand Genomics Limited (NZGL, N.Z.), which generated 160 FastQ read files. Illumina adapters were removed from the sequence reads by NZGL. SolexaQA++ v3.1.4 (Cox *et al.* 2010) was used to perform three additional quality control operations: checking that all reads were correctly paired, trimming to the largest contiguous sequence for which all bases have a probability error < 0.05; and removing all reads < 50 nucleotides in length. This generated three outputs: paired-end reads and single-end reads (both passed quality control) and discarded reads (which failed quality control). Reads were mapped using Bowtie2 v2.2.6 (Langmead & Salzberg 2012) to the godwit reference genome. Paired-end and single-end reads from the trimming process were mapped separately. Read files were processed using '--very-sensitive' end-to-end mapping and

each sample was assigned a unique read group ID (i.e. '10E' for sample 10 of the early bird group). We used SAMtools v1.2 (Li *et al.* 2009) to: convert into binary BAM files, merge all BAM files for an individual in a single BAM file, sort reads by genomic coordinates, and to remove duplicate reads (this was done in SAMtools v0.1.19 due to a known error in the 'rmdup' function of SAMtools v1.2). Final BAM files were then indexed. Genome coverage was calculated for all 20 samples using the 'pileup' function of SAMtools v1.2. Two genomic tests for relatedness were performed using VCFtools: the *Ajk* statistic of Yang and colleagues (Yang *et al.* 2010), and the O statistic of Manichaikul and colleagues (Manichaikul *et al.* 2010).

Variant (SNPs) calling

Variants were inferred with the Bayesian haplotype-aware variant caller, FreeBayes v0.9.21 (Garrison & Marth 2012). Variant calling was performed in a population-aware manner, where markers are determined by comparing all sequenced individuals simultaneously. Indels variants (i.e. polymorphisms that produced a length change relative to the reference) and all variants with a quality score less than Q20 were removed using VCFtools v0.1.12a (Danecek *et al.* 2011).

SNPs selection

We calculated *Fst* (Weir & Cockerham 1984) using VCFtools v0.1.14 between 'early' and 'late' migrants in the re-sequenced group for all 8,941,974 SNPs to help identify SNPs that showed some potential difference between the extremes of the migration timing continuum. SNPs with higher *Fst* were prioritized. We created two lists of SNPs and ranked them from the highest to the lowest *Fst* values including: 1) only SNPs within candidate gene regions and; 2) all the SNPs. For the first list we set a threshold ($Fst > 0$). For the second list, we selected the SNP with the highest *Fst* value of all the SNPs (regardless of the type of protein in which it was placed) in addition to 37 SNPs located in predicted genes that fell in one of the categories from which initial candidate genes were selected (i.e. elements of the endogenous circadian clock, its outputs, components of the Hypothalamic-Pituitary-Gonadal (HPG) axis, elements of the fat metabolism). More details about these 'new' candidate genes that were added in the second instance are in Appendix 3, Table 3.1. We finally generated a file (BED format) containing 5,229 SNPs (i.e. 5,192 SNPs within initial candidate genes regions + 37 SNPs from 'new' candidate gene regions) which were genotyped in the 269 individuals. The

final number of candidate genes contained in the 5,229 SNPs was 140 (113 from the initial list + 27 from the 'new' list).

Genotyping and variant calling

Probes for genotyping could be designed for 4,942 SNPs of the selected 5,229 SNPs (i.e. 287 SNPs were placed in regions of repetitive elements). The 4,942 SNPs were genotyped in the 270 samples with Capture-Seq technology (RAPiD Genomics, USA). Two samples were duplicates for control purposes, therefore the number of individuals was actually 269. Raw data were demultiplexed using Illuminas BCFofastq, as well as cleaned and trimmed (i.e. 3' end were trimmed by removing Low quality basis with < 20 quality score and reads were filtered to remove reads with more than 10% of the read with < 20 quality score). Clean reads were aligned to the reference godwit genome (acc. number: LXVZ00000000.1) with MOSAIK (Lee *et al.* 2014). Freebayes v1.1.0 was used for the SNPs calling.

Filtering and quality control

We used VCFtools v0.1.14 (Danecek *et al.* 2011) to filter SNPs using the following parameters: minimum allele count (mac) = 3, minimum depth (minDP) = 3, minimum allele frequency (maf) = 0.05, minimum quality (minQ) = 40, minimum genotype quality (minGQ) = 30. We set minDP to 3 (i.e. keeping genotypes with 3 or more reads) because FreeBayes, which has been used to call the variants, are proven to confidently call genotypes with few reads since variants are assessed across all samples simultaneously (Bao *et al.* 2014). We also calculated the cut-off proposed by Li (2014) (i.e. the average depth plus a multiplier of 3 times the square root of average depth) for the maximum mean depth parameter (max-meanDP = 289). We applied the Hardy Weinberg Equilibrium filter (hwe = 0.05) to remove potential erroneous variant calls as suggested in Li (Li 2014).

One sample (MY13391) was replicated and sent in different plates in order to control for batch effects (Laurie *et al.* 2010). Concordance of the genotypes of these replicates was 99.5%. We checked that there were no duplicated samples (a part from the two replicates) using VCFtools v0.1.14. We calculated a relatedness matrix (VanRaden 2008) using a presumably unlinked set of SNPs (i.e. SNPs located at least 100 Kb from each other, which was obtained using VCFtools v0.1.14 --thin tool). This was done taking into consideration the fact that Linkage Disequilibrium (LD) in birds is on the order of

10–100 kb (Edwards & Dillon 2004; Li & Merila 2010; Stapley *et al.* 2010a). This kinship matrix was generated using the R package GAPIT (Lipka *et al.* 2012).

Statistical analyses

We used Bayesian Sparse Linear Mixed Model (BSLMM) (Zhou *et al.* 2013), which is a hybrid of the Linear Mixed Model (LMM) and Bayesian Variable Selection Regression Model (BVSR). While LMM assumes that all variants have an effect on the phenotype (polygenic basis) with effect sizes normally distributed, BVSR assumes that a relatively small proportion of all variants have an effect on the phenotype (mono/oligogenic basis). We accounted for site and kinship (VanRaden matrix) as covariates. This kinship matrix was obtained using the software TASSEL v.5.2.39 (Bradbury *et al.* 2007) using a subset of 124 SNPs presumably not in LD, spaced at least 100 Kb from each other. These covariates were included when fitting BSLMM model in the software GEMMA (Genome-wide Efficient Mixed-Model Association) (Zhou & Stephens 2012) with a burn-in of 5 million steps and subsequent 20 million MCMC steps (sampling every 1000 steps). We used this method to: 1) obtain insight about the genetic architecture of departure time for migration (i.e. closer to a LMM case, meaning a highly polygenic basis of the phenotype, or a more a BVSR case, which means that there are a few major effects loci (mono/oligogenic basis of the phenotype); and 2) to estimate the "chip heritability", which is defined as the proportion of variance in phenotypes explained by all SNPs, and is given by the hyperparameter PVE (Proportion Variance Explained).

We used LMM to assess the effect of SNPs on migration timing using the software TASSEL v.5.2.39. We assessed the possible inflation of test statistics and therefore, the lower control over the occurrence of false positives, through the QQ plot (Voorman *et al.* 2011). We accounted for *site* as a covariate and kinship (i.e. cryptic relatedness) as a random factor (Yu *et al.* 2006). This is the LMM (Bradbury *et al.* 2007) equation of the model:

$$\text{(LMM) } mdd_i = \mu + ß_1*SNP_i + ß_2*site + ß_3*K + e_i$$

where the phenotype (mdd_i: mean departure date) of individual *i* is a function of the population phenotypic mean (μ), the SNP genotype, *site* and kinship as random effect, and the residual effect (e_i).

We tested the power of these models to detect significant SNPs through a simulated analysis. Briefly, this consisted of adding three 'dummy' SNPs from a study of

migration timing in steelhead (*Oncorhynchus mykiss*) (Hess *et al.* 2016). We took the three SNPs from that study that were significantly associated with steelhead migration timing and simulated the equivalent distribution across our godwit chronophenotype data. Our analyses could detect those three influential SNPs in our association test (Appendix 3, Figure 3.1), confirming that this analysis/method had the power of detection for an effect of that size.

RepeatABEL package for R (Rönnegård *et al.* 2016) allowed us to use longitudinal phenotypic data (that is, repeated measures of phenotypes over several years). This method accounts for kinship (K) through the calculation of a relatedness matrix (in principal the same as the VanRaden relatedness matrix with the sole difference that each SNP is corrected for its variance in genotype coding rather than the variance for all SNPs). The advantages of this method are that it allows using of longitudinal data (repeated measurements of the phenotype) and it increases the power of the GWAS (Rönnegård *et al.* 2016). For this analysis, we used the longitudinal standardized chronophenotypes. We also fitted the model including *site* as covariate. Heritability was estimated following the RepeatABEL package tutorial.

As migration timing in godwits seems to be a polygenic trait (i.e. many SNPs with very little effect involved in the trait; see Results) we tested for association between combination of SNPs and mean standardized chronophenotypes using the pipeline Lamplink (Terada *et al.* 2016). We performed this analysis with both the recessive exclusive model (in which the homozygote of the minor allele (aa) is assigned 1 and the homozygote of the major allele (AA) and the heterozygote (Aa) are assigned 0) and the dominant exclusive model (in which the homozygote of the minor allele (aa) and the heterozygote (Aa) are assigned 1 the and homozygote of the major allele is assigned 0). The statistical test used was a Mann-Whitney U test and the significance level was set at 0.05.

Gienapp *et al.* (2017) showed that taking the interaction of genotypes with environmental variables into account can improve the success of detecting SNPs associated to the trait of interest. Besides, the Catlins subpopulation has a shifted phenology comparing to the other two sites. This suggests that there are potentially environmental factors associated with site that might interact with the genotypes. In

pursuit of this possibility, we used Plink v1.07 (Purcell *et al.* 2007) to test the following model:

$$mdd_i = \mu + ß_1*SNP_i + ß_2*site + ß_3*(site{:}SNP_i) + e_i$$

where the phenotype (mdd_i: mean departure date) of individual *i* is a function of the population phenotypic mean (μ), SNP genotype, *site*, the interaction between SNP genotype and *site* and the residual effect (e_i). We tested for phenotype-genotype association for the three genetic models: additive, dominant and recessive.

Population structure analyses: comparison between neutral (microsatellites) and candidate SNPs loci

With the aim to obtain SNPs putatively not in LD, two subsets of SNPs were generated using the function --thin in VCFtools. These two subsets of SNPs and the complete set of 3,412 SNPs were used to assess genetic background with the program fastStructure (Raj *et al.* 2014), which uses a Bayesian framework. Numbers of clusters (*K*) tested ranged from 1 to 5. The software Distruct v2.2 (Rosenberg 2004) was used to generate the Q-plots. We compared these results with those obtained with microsatellites in Chapter 2. The correlation between mean departure date and probability of ancestry to cluster 1 (Q1) when assuming $K = 2$ was assessed for each of the sites using a Mantel test (Mantel 1967), which was performed with 10,000 permutations and using the R package *ade4* (Dray & Dufour 2007).

To correct for multiple testing and control false positives, we calculated Benjamin-Hochberg (BH) p-values (Benjamini & Hochberg 1995) every time that was needed.

Analyses using birds with geolocators and departure dates from Asia

The subset of individuals (n = 34, all from Manawatu) with departure dates from Asia derived from geolocator data was used to perform some of the previously described analyses in order to explore the consistency of the findings obtained in the larger dataset on a smaller but more accurate sample. Briefly, we performed univariate (i.e. LMM in TASSEL) and multivariate (i.e. combination of SNPs in Lamplink) models, following the previously described methods. Weak population genetic structure was suggested according to Chapter 2 ($Fst = 0.016$, $p < 0.05$), and this was taken into account by doing a Principal Component Analyses (PCA) (Zhu *et al.* 2002; Price *et al.* 2006) using the microsatellite data to obtain the individuals' eigenvectors, which then were included in

the LMM as covariant. This was done in R (R Core Team 2014) using the package *adegenet* (Jombart 2008).

Population genetic structure analyses (i.e. fastStructure) was also done using the same 124 SNPs previously described and Q values obtained when assuming $K = 2$. Probability of belonging to one of the clusters was plotted against mean departure dates. The same procedure was done with Q values obtained when assuming $K = 2$ from 27 microsatellites in Chapter 2. A Mantel test was used to assess correlation between mean departure date and Q values for both 124 SNPs and 27 microsatellites.

Results

Godwit reference genome

The generated godwit reference genome had 32,319 scaffolds (N50 = 283 Kbp). The total length of the assembly was 1,034,770,521 bp (with gaps). This assembled genome was deposited in NCBI (acc. number: LXVZ00000000.1). We successfully predicted 16,538 genes in the godwit genome; however, the genes coding for AANAT, EYA3 and DRD4 were, unfortunately, not found (Appendix 3, Table 3.2).

Whole genome re-sequencing, variant calling and SNPs selection

78% of the reads (76–80% across individuals) mapped to the godwit genome. Read depth averaged 4.3-fold (2.7 to 4.9-fold) (Appendix 3, Table 3.4). 8,941,974 polymorphic variants were identified across the 19 individuals after performing quality control. 15,221 SNPs fell within candidate gene regions.

Final SNPs and individuals in the genotype-phenotype analysis.

We had a coverage of 99.5% of the targeting SNPs (that is, we successfully genotyped 4,919 SNPs of the initial 4,942) in 270 samples. Note that one sample was from the same individual (duplicates, for control purposes), therefore our sample size is 269 individuals. After the filtering step (i.e. quality, depth, minimum allele frequency) the number of SNPs was 3,412. This number corresponds to 120 candidate genes (of the initial 140). Regarding the quality checking of individuals, the kinship matrix showed evidence of four individuals being closely related (potentially first siblings). Consequently, we decided to discard these samples (samples IDs: MY13383, MY13300, MY13581 and CY12826) from further analyses, reducing our sample size to 265.

Genetic architecture of departure time for migration (BSLMM)

We investigated the type of architecture of migration departure time using all sites, in which we included *site* as a covariate. In these analyses, we focussed primarily on the following hyperparameter: PVE (Proportion of Variance Explained), PGE (Proportion of Genetic variance explained by sparse Effect, also called 'chip heritability'), n.gamma (number of variants with major effects) and PIP (Posterior Inclusion Probability). PIP is the frequency that a variant is estimated to have a sparse effect in the MCMC (Markov Chain Monte Carlo), in other words, the number of times the variant appears in the MCMC with a non-zero sparse effect. Results gave a PVE = 52%, from which 69% (PGE) is explained by a few variables (n.gamma = ~7). Site was the variable with the strongest effect and highest PIP value. Similar estimates were obtained from the estimation of heritability using the longitudinal standardized chronophenotypes, which estimated that 49% of the phenotype was heritable.

Univariate and multivariate candidate-gene association study

No SNP showed significance after adjustment to account for multiple testing in any of the univariate models tested (i.e. LMM, LMM with repeated measurements), or including the interaction of *site* with SNPs in the LMM. Similarly, no combination of SNPs showed significance under any of the three genetic models tested (i.e. additive, recessive and dominant).

Population structure test on candidate SNPs and microsatellites

The filtering of SNPs based on a LD of 10 Kb and 100 Kb resulted in two subsets of SNPs: 329 SNPs in 120 candidate gene regions and 124 SNPs in 119 candidate gene regions, respectively. FastStructure results performed with all the candidate SNPs, and with two subsets of 124 and 329 SNPs putatively not in LD supported $K = 1$ as the best number of clusters that explains structure in the dataset. Despite this result, the bar-plots assuming $K = 2$ show a subtle difference between earlier and later birds (Appendix 3, Figure 3.2), which is more evident at the Manawatu site and when using the subset of 124 unlinked SNPs (Figure 4.2). This pattern is not present when using the 27 microsatellites used in Chapter 2. Correlation of proportion of ancestry (i.e. probability of belonging to one or another cluster) derived from the 124 candidate SNPs was correlated with mean standardized chronophenotypes for each site ($r = 0.20$, $r = 0.32$, $r = 0.46$ for Catlins, Firth of Thames and Manawatu, respectively, Figure 4.3a). The Mantel tests were significant ($p < 0.05$) for Manawatu and Firth of Thames ($p = 0.0001$, $p = 0.003$ respectively),

although not for Catlins (p = 0.13). In contrast, equivalent analyses using neutral variation (27 microsatellites) rather than candidate SNPs showed no correlation with mean standardized chronophenotypes (Figure 4.3b). The Mantel tests of the 27 microsatellites with the mean chronophenotypes were not significant for any of the sites (p > 0.05).

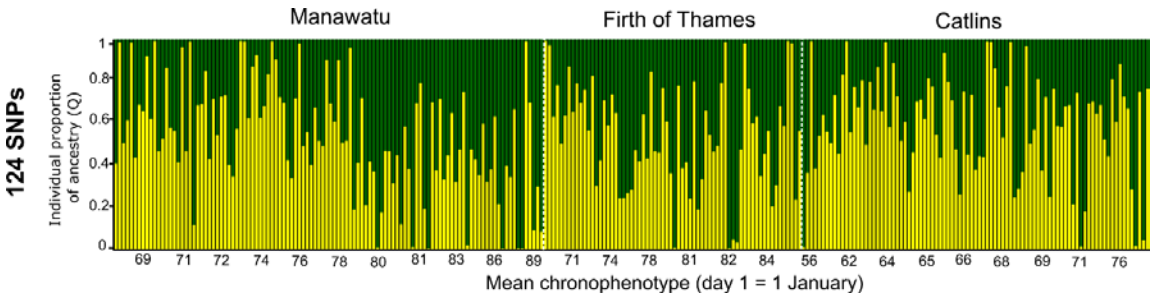


Figure 4.2: FastStructure Q-plot of individual ancestry (Q) at *K* = 2 using the subset of putatively unlinked 124 SNPs (i.e. separated by at least by 100 kb). Individuals were sorted by their mean standardized chronophenotype (x-axis) and organized by site (i.e. Manawatu, Firth of Thames and Catlins), which are separated with dash lines. Mean chronophenotypes are showed for every ten birds.

Results from birds with geolocators

The univariate analysis by which we tested the association between the 3,412 SNPs and migratory departure date from Asia (n = 34) resulted in no significant SNP, after BH correction for multiple testing (Appendix 3, Figure 3.4).

Population structure analyses using the 124 SNPs from candidate genes produces the same pattern observed in the bigger dataset of individuals from Manawatu with departure dates from New Zealand. The correlation was significant (Mantel Test, p = 0.0003) and stronger (r = 0.60) than that of Manawatu N.Z. mean departure dates (Figure 4.3).

Discussion

We sought to explain individual variation in migration timing in bar-tailed godwits, which provide an excellent system of natural variation in which to undertake genetic studies for several reasons. (1) We monitored birds embarking on the first flight of their migration towards breeding grounds, so there is no influence of previous events

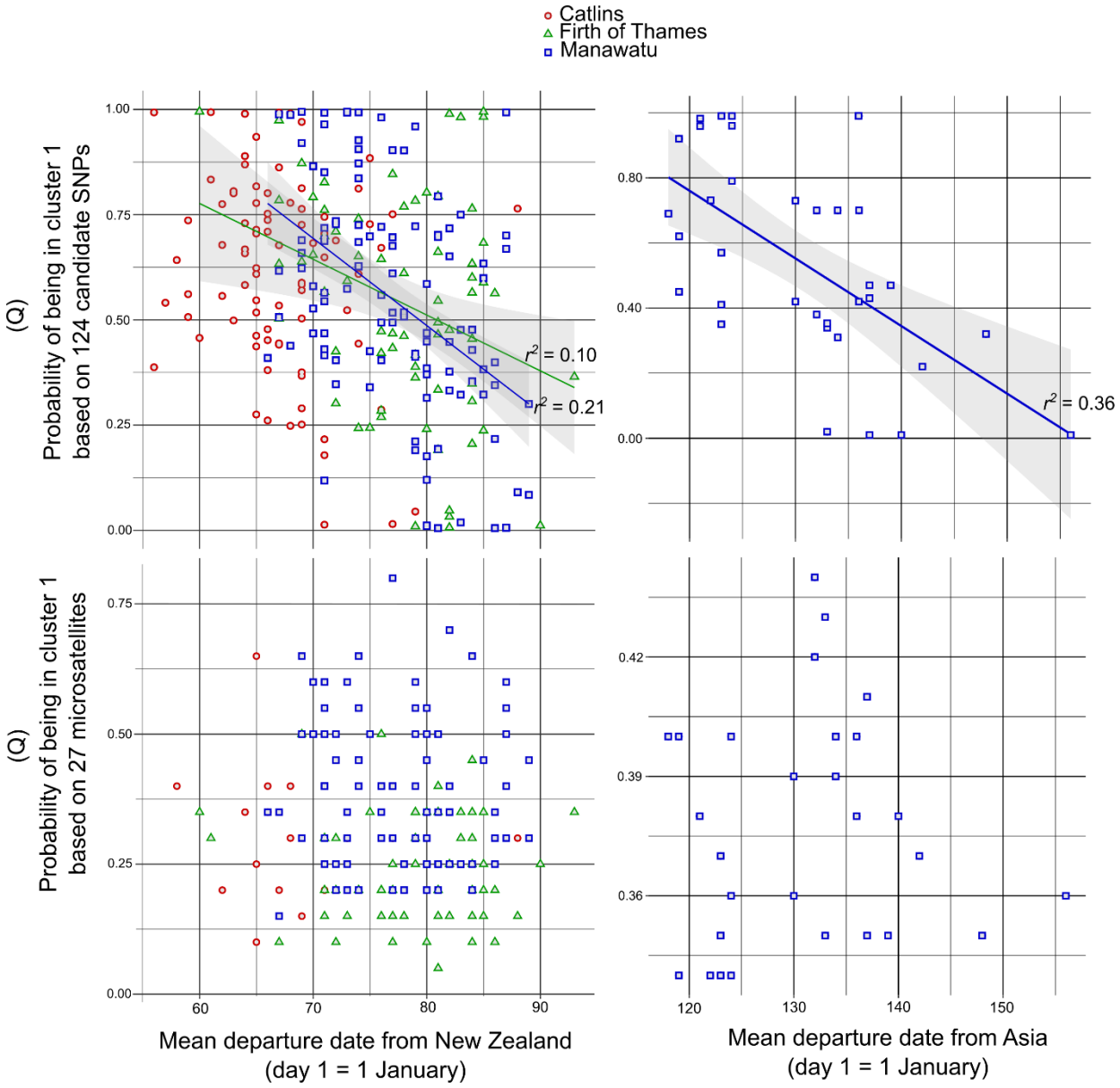


Figure 4.3: Correlation between the probability of belonging to cluster 1 (Q) and mean departure dates from the three sites in New Zealand (left side of the panel) and from Asia (right side of the panel) for the 124 SNPs (upper graphs); and for 27 microsatellites used in Chapter 2 (lower graphs). Note that departure dates from Asia were available for 34 birds, from which departure dates from New Zealand was also available. Different shapes indicate the origin from which blood samples were obtained: Catlins (circle), Firth of Thames (triangles) and Manawatu (squares). Interval of confidence (95%) is shown in shadow and r^2 is indicated only for significant correlations. The work with microsatellites was done with a total of 169 individuals (16 from Catlins, 61 from Firth of Thames and 91 from Manawatu). There are 141 individuals (15 from Catlins, 41 from Firth of Thames and 85 from Manawatu) in common between the present study and Chapter 2. Results from using these 141 individuals do not differ extensively from what is shown here (see Appendix 3, Figure 3.3).

on migration timing (differential timing of arrival on the Alaskan breeding grounds, Conklin & Battley 2012). (2) Multiple years of monitoring confirms the consistent nature of the individual differences in migration timing. (3) The underlying reason for the broad-scale variation between individuals is understood (differential timing of arrival on the Alaskan breeding grounds, Conklin *et al.* 2010). (4) Weather conditions seldom halt migration for more than a few days (Conklin & Battley 2011), so migration dates of birds can be assumed to reflect individual preference to a large degree. (5) Within a site, all individuals experience the same environmental conditions (i.e. daylength change), so variation between individuals is likely to reflect internal rather than external factors. Finally, (6) the measure of migration timing (departure date) could be measured simply and, at our main study site (Manawatu Estuary) with certainty, so we have a reasonably large sample size for a study of free-living migratory birds. New Zealand-wintering godwits are therefore an ideal candidate in which to test for genetic influences on individual migration timing. Accordingly, we coupled one of the largest genomic datasets with phenotypic records of departure times to evaluate the genetic determinant of migratory timing within a natural population.

Despite screening genetic diversity in 120 genes, including all genes known to be associated with internal clocks in other species, we found none that had strong individual effect on departure time. Nevertheless, 124 SNPs in 119 genes – 1 SNP placed in one candidate gene, approximately – explain 10–20% of the phenotypic variance, suggesting a moderate level of heritability in migratory departure time. The fact that these results are consistent with those obtained when using a smaller dataset from which departure dates from Asia were available – the 124 SNPs explain 36% of the phenotypic variance – suggests that this finding is not merely a coincidence. It therefore seems that a complex interaction of many genes underpins migration timing in godwits. This is quite different to the clock mechanism in some non-avian species, such as the steelhead, where just three genes explain 46% of the variance in the annual migration timing (Hess *et al.* 2016). This difference may be because as-yet unknown genes influence this trait in godwits (Lundberg *et al.* 2017), or that we have simply missed key genes, including those that do not presently appear to have orthologs in the godwit genome (e.g. *aanat*, *eya3*, *adcyap1*, *dio3*).

As far as we know, this is the first time that the genetic architecture of migration timing has been studied in a long-distance migratory shorebird. Polygenic architecture is

common for quantitative traits in natural populations (Santure *et al.* 2013; Delmore *et al.* 2016; Silva *et al.* 2017). However, SNPs with larger effects occur in other taxa (Hess *et al.* 2016), suggesting that the genetic basis of 'internal timing' may differ widely among species, perhaps shaped by species-specific evolutionary history or evolutionary convergence (Pulido 2007a).

Comparison of the genome-wide diversity in 19 early-departing and late-departing godwits, although limited by statistical power, does not reveal any particular region with high allelic frequency differences between these groups (Appendix 3, Figure 3.5 and Appendix 3, Figure 3.6). Furthermore, SNPs within candidate genes did not exhibit higher allelic frequency differences, as indicated by *Fst*, than randomly chosen regions of equivalent length (Appendix 3, Table 3.6). As with the candidate gene analyses, this suggests that many genes may each be making relatively small contributions to the overall timing effect. Different combinations of haplotypes might also lead to similar chronophenotypes and the complex interaction of loci (epistasis) and epigenetic modifications are yet other possibilities (Carlborg & Haley 2004; Le Rouzic *et al.* 2008). However, it remains challenging to study these factors in natural populations, principally due to the very large sample sizes needed and the lack of controls that non-experimental systems permit (Carlborg & Haley 2004; Kardos *et al.* 2016).

Beyond addressing the ever-present issue of sample size and statistical power, there is still little information on how environmental factors (such as daylength, rate of change in daylength, wind conditions, and sun position) might interact with an individual's specific 'internal clock' thus modifying the response (behaviour) of individual birds. Controlling for these environmental factors improves the modelling of genotype-phenotype interactions (Carlborg & Haley 2004; Gienapp *et al.* 2017). Finally, a broad range of models – additive, dominant, recessive and more complex models – may well be required to identify the basis of this trait. It is, however, important that these models are applied within natural population settings and all the complexity that it involves (Helm *et al.* 2017). There is still a long way to go to fully understanding the genetic mechanisms of complex behaviours. We are still exploring the potential of new technologies to unravel the extent to which genes ultimately influence behaviours in natural populations, but we are just beginning to discover how complex and challenging these systems are.

Chapter 5

Synthesis

A godwit waiting for its turn to be carefully measured, sampled and released, at Manawatu River estuary (November 2014).

Introduction

For many decades, the field of chronobiology developed principally through experiments, first on model-organisms and then on semi-wild individuals. Chronobiologists have gained considerable knowledge in understanding the properties and characteristics of biological Clocks in different organisms. Recently, the big step towards understanding biological Clocks in the real world (that is, in natural conditions) has taken off (Schwartz *et al.* 2017). Field biologists are studying the annual routines of wild individuals in ways not possible previously, with miniaturisation of remote tracking technologies enabling the tracking of even small passerine birds (Hallworth & Marra 2015; Dominoni *et al.* 2017). We can now ask questions of how biological Clocks work in natural populations, and whether we can explain timing with what we already know from controlled experiments. Chronobiology and ecology are converging due to common interests in understanding the properties and adaptability of chronotypes in the real world (Helm *et al.* 2017). The need for this interdisciplinary approach, together with the still-in-expansion possibilities that genomics brings, opens the door to explore unknown worlds, albeit with associated and uncharted risks. In my thesis, I took this (exciting) risk in trying to gain insight into the genetics underlying migratory departure time using an excellent system, the New Zealand bar-tailed godwit.

It is fair to say that at the outset of this study, we all (supervisors and myself) had high hopes for a nice, clean and significant finding. The system was fascinating and clearly described, and had the potential to throw up an answer that would revolutionise understanding of the genetics of avian migration. Instead, I had the experience of how science more commonly develops, through incremental advances and insight into complex biological systems such as the links between genetics and behaviour leading to the conclusion that our initial expectations may have been rather naive. Nevertheless, my research was able to address successfully the initial questions, which were:

1) Is there genetic structure in New Zealand godwits that could confound genotype-phenotype association studies? (Chapter 2)
2) How does migration departure date relate to background genetic structure? (Chapter 2)
3) Does the *Clock*polyQ polymorphism distribution show any evidence of being involved in determining chronophenotypes in godwits? (Chapter 3)

4) Do any candidate genes associated with the endogenous clock have a strong effect in determining individual chronophenotypes? (Chapter 4)
5) What does the genetic framework underlying migratory departure time in godwits look like? (Chapter 4)

In this chapter, I will integrate the main findings of my research with published work as well as discuss the implications of these findings to understand a bit more this world around us (specifically the world of how migratory timing behaviour relates to genetics). Last, I address new questions and highlight those questions that remained unanswered together with my own perspective of how these should be addressed in future studies.

The first godwit genome

Godwits are, without doubt, an example of a remarkable long-distance migratory bird that awakes intrigue and excitement to most people who know about them. This thesis contains the first – and so far unique – work done in New Zealand godwits addressing its genetics. As part of this project, we generated the first godwit genome (i.e. *Limosa lapponica baueri*) available to the public, which joins a list of 95 other bird genomes available in the NCBI so far, being the only species representing the genus *Limosa* and the second of the Scolopacidae family (with the ruff, *Calidris pugnax*). Specifically, 84% (1.0 Gbp) of the estimated complete genome size of the godwit (i.e. 1.2Gbp) was successfully sequenced and assembled. This contribution will serve for future studies that will benefit from having a broader range of genomes available. Potentially, future studies would aim to do comparative multi-genomics analyses to answer evolutionary questions of broad scale (i.e. new phylogenetic studies establishing the relationship between species and their evolutionary histories), as well as for, possibly, understanding more fundamental questions regarding the common characteristics of eukaryote genomes and its relationship with evolutionary processes. Future studies specific to the bar-tailed godwit may also benefit from having a reference genome.

Population structure background in godwits and its relationship with migration departure time

Chapter 2 aimed to first assess the potential population genetic structure in godwits associated with geography on the breeding grounds and with migratory timing.

This was an important first step, as it would highlight the size of any underlying genetic structure that might have to be accounted for in associations between functional genes and migration timing. Given that godwits show some evidence of a morphological cline on the breeding grounds (i.e. northern breeders are smaller and darker than southern breeders) as well as a suggestion of an associated behavioural cline (i.e. earlier departers from New Zealand and from Asia breed at more southern areas than later departers, Conklin *et al.* 2011), there is the potential for there to be at least some genetic structure associated with migration timing. Population genetic structure in New Zealand godwits was assessed using quite a large number of microsatellites (27) and three commonly used analysis methods. These methods are based on slightly different algorithms but in essence, they search for the number of clusters that best fit taking into consideration the assumption of Hardy-Weinberg equilibrium as the expected allelic frequency model that defines a population in panmixia (random mating).

I found no evidence for population genetic structure associated with migration timing within New Zealand godwits, but very slight differences associated with geography in Alaska (northern versus southern breeders) and with migration timing from Asia. The congruence between the latter two is not surprising, as the relationship between migration timing from Asia and breeding destination is very strong (i.e. little overlap in timing between northern and southern breeders: Figure 2.2a). It was perhaps more surprising that no association was found with timing of migration from New Zealand given the documented relationship between geographic location in Alaska and departure dates from New Zealand (Conklin *et al.* 2010). Examination of the migration dates of known-breeding-ground godwits in New Zealand (Figure 2.2b) revealed that there is greater overlap of migration timing of northern and southern birds in New Zealand and, consequently, the weak genetic structure becomes undetectable. The implications for my study is that population genetic structure would not confound any phenotype-genotype associations detected in New Zealand (Chapter 4). The *Fst* between earlier and later birds from Asia is very low (*Fst* = ~ 0.01) and the other methods (i.e. STRUCTURE and *K*-means) were not capable of detecting more than one cluster. This weak genetic structure background could confound results for the aims of Chapter 4, and therefore it was taken into account. Having identified that breeding region (and associated migration timing in Asia) has implications for population dynamics, the question that remains unanswered is whether this weak population structure indicates an incipient microevolutionary process

of separation, and in which direction this process is taking. This question will remain for future studies.

*Clock*polyQ polymorphism in godwits

The length of the *Clock*Poly-Q region has been described as a modifier of the functionality (i.e. as transcription-activating factor) of the protein CLOCK, thereby influencing timing-related phenotypes (Gekakis *et al.* 1998). The discovering of a poly-Q region (i.e. region rich in glutamine) in the avian *Clock* gene and its polymorphism gave rise to studies trying to understand the meaning of the existence of this diversity at the *Clock*poly-Q region in the different aspects of the phenology (i.e. breeding time). During the last decades, many studies addressed this question using different systems, principally passerines, and from a diverse perspective, principally testing its potential relationship with breeding timing and/or breeding latitudinal cline. Three previous studies analysed the relationship between migration timing and *Clock*polyQ polymorphism in birds – both in passeriforms – (Bazzi *et al.* 2015; Saino *et al.* 2015; Contina *et al.* 2018), and the claims of these studies show inconsistencies. Exploring the possibility of association of this controversial polymorphism using the precision of godwits' departure date for migration could potentially give light to this confused panorama.

In Chapter 3 I described the poly-Q polymorphism in Godwits, which consists of six polymorphisms (i.e. Q7–Q12). There is a growing bibliography on *Clock*polyQ variation in different taxa, many on migratory bird species (Appendix 2, Table 2.3). Species vary greatly in how diverse the *Clock*polyQ polymorphism is with high diversity in blue tits (Liedvogel & Sheldon 2010) and godwits (Chapter 3), but low variability in barn swallows (Dor *et al.* 2011b), juncos (Peterson *et al.* 2015) and Wilson's warblers (*Cardellina pusilla*) (Bazzi *et al.* 2016b). The reason why *Clock*polyQ is diverse in some species but not in others is still not clear. Bazzi *et al.* (2016) analysed species-level *Clock*polyQ diversity in 23 trans-Saharan migratory species and compared levels of diversity against a number of migratory traits. They found that *Clock*polyQ length was negatively related to migration distance, with shorter-distance migrants having higher diversity than longer-distance migrants. Their proposed explanation was that the low diversity of *Clock* in long-distance migrants could be a result of strong stabilizing selection on phenotypes, so that *Clock* diversity declines due to tight time constraints in long-distance migrants. This pattern clearly does not match the *Clock*polyQ diversity in

godwits, as these are very long-distance migrants with highly consistent migration timing, yet with high *Clock*polyQ diversity (Chapter 3). The role of the *Clock*polyQ diversity in migratory timing traits is still not clear.

The main finding in this chapter was that migration departure date (from New Zealand or from Asia) was not associated with the distribution of the different alleles of *Clock*polyQ or the genotypes (Chapter 3, Figure 3.2). This finding contrasts with other claims about *Clock*, such as the title of Saino *et al.*'s paper "Polymorphism at the *Clock* gene predicts phenology of long-distance migration in birds" (Saino *et al.* 2015). As discussed in Chapter 3, the evidence that Saino *et al.* (2015) base their claim on comes from just one out of four species they studied in total, and when taking into consideration one type of genetic model (i.e. dominant or additive). In short, the main claim in their paper is debatable given their findings. Chapter 3 joins the increasing number of studies of *Clock*polyQ and migratory timing behaviour showing no evidence an association. Nevertheless, there is evidence that underlying mechanisms that drive bird phenological characteristics may be different between species (Liedvogel & Sheldon 2010), even between distinct populations (Saino *et al.* 2015; Bazzi *et al.* 2016b), therefore, we cannot still generalize that the *Clock*polyQ polymorphism does not have some role in generating migratory timing phenotypes in other migratory species. It also appears as if opinions are changing as to the role that *Clock* may play in the annual routines of birds. The paper that essentially started the interest in *Clock* (Johnsen *et al.* 2007) proposed that variation in *Clock* may relate to microevolutionary responses to photoperiod related to latitude. Subsequent studies implied that *Clock* may play a direct role in migration timing (O'Malley & Banks 2008; Saino *et al.* 2015; Bazzi *et al.* 2016c), however, many other studies failed to find such a role (Bazzi *et al.* 2016a; Bazzi *et al.* 2016b; Bazzi *et al.* 2016c; Contina *et al.* 2018). Interestingly, Zhang and colleagues found that *Clock*polyQ length of individual birds is associated to laying date (Zhang *et al.* 2017), which is in accordance with previous studies on passerines (Liedvogel *et al.* 2009; Caprioli *et al.* 2012; Bourret & Garant 2015). This study takes the evidence of an experimental study done in rats in which they found that CLOCK binds to E-box elements in the gonadotropin hormone (GnRH) gene (Resuehr *et al.* 2007). This hormone plays a role in the upper stream of the Hypothalamic-Pineal-Gonadal axis, which is the route that triggers physiological changes that prepare the organism for the reproductive period. Thus, it is possible that through a mechanism – which it is still not clear –, the length of *Clock*polyQ affects this binding

and consequently it produces the observed inter-individual variation in breeding/laying time (Zhang *et al.* 2017). In short, the most recent studies consider that the role of *Clock* is more an adaptation to photoperiodic regimes at the breeding grounds with, possibly, implications for breeding time rather than a direct determination of migratory timing.

The genetics of migration timing in bar-tailed godwits.

Chapter 4 is arguably the most comprehensive study of the genetic basis of migration timing ever done in birds so far. In this chapter, I did an extensive selection of 140 candidate genes, which include all the genes of the CCO, in addition to other genes that had been previously described to have a function related to timing events in other organisms. In addition, I compared 19 genomes from individuals of the extremes of the chronophenotype continuum, which served as an exploration of other regions apart from those candidate genes. Results show no significant association of individual genes to migratory departure time and a strong indication of a multigene effect, as shown by 124 SNPs (from 119 candidate genes) that can explain ~ 10–20% of the phenotypic variance in godwits. Moreover, this same finding (even stronger, ~ 36%) was obtained using the Asian departure data from the 34 geolocator-tracked birds. As we did not have complete genome coverage (i.e. 84%) it is possible that some influential polymorphic genes exist that were not sequenced. Nevertheless, Chapter 4 suggests that migratory timing behaviour is governed by a large number of genes with small effects.

The main finding of Chapter 4 is that 124 SNPs (harbouring 119 candidate genes) are capable of explaining ~10–36% of the phenotypic variation as shown by a population genetic approach based on allelic frequencies. This finding suggests that the complexity behind the timing of migration in godwits is higher than expected initially. Either each of these 124 SNPs is contributing equally to explain the totality of the 10–36% of phenotypic variance, or there are a few that might explain larger percentages of the explained phenotypic variation. This question could be explored through a Random Forest analysis (e.g. Holliday *et al.* 2012; Brieuc *et al.* 2015), which is a method that does not require assumptions regarding the type of relationship between the phenotype and the genotype (i.e. additive, dominant or recessive). This approach was used, for example, by Hess *et al.* (2016) in combination with a univariate methodology (i.e. LMM and GLM) in order to rank candidate SNPs by their importance in explaining migratory timing in steelhead. Given that my univariate SNP-migratory timing analyses showed that no SNP had strong

effects on chronophenotypes, I would not expect that a few genes would be ranked significantly at higher positions than the others. Nevertheless, the Random Forest test would be a suitable analysis to further test the importance of these 124 SNPs in determining migratory departure time, and probably to give more support to the finding of a polygenic scenario underlying the determination of migratory departure time in godwits.

Issues & challenges

Without doubt, I had to face several challenges in the course of this research, some of which are outlined here:

As mentioned in the introduction, the genes *adcyap1* and *clock*, were the genes the most studied for genetics of migratory traits studies. Initially, both *adcyap1* and *clock* were going to be the elements of study for Chapter 3, however preliminary laboratory work done in ~25 birds resulted in only two alleles being detected in *adcyap1*. I decided not to invest more time and money in the work on *adcyap1* given that it was not looking very promising (i.e. low genetic diversity).

Chapter 4 was, initially, going to assess genetic and epigenetic influences of a few candidate genes (i.e. those of the CCO), however, in the early stages of my thesis I decided not to include the epigenetic component on my research. There were two reasons for this decision: 1) I was not in the position to obtain suitable samples to adequately conduct such epigenetic studies because epigenetic modifications are tissue-specific (Ledon-Rettig *et al.* 2013). Therefore, for studying epigenetic effects on migratory timing behaviour the sampling of brain tissue would be ideal, but this is not possible for the godwit, as it would require collection of large numbers of birds of known phenology; neither practicable nor culturally and ethically desirable; 2) we felt it was wiser to focus on the genetic perspective more in depth (i.e. with a genome-informed larger number of candidate genes). There were new studies reporting potential candidate genes related to migration timing, therefore, it was possible to include more candidate genes, a part from those of the CCO, which would broaden the genetic scope of my study. Given the current view regarding how the genetics of complex traits work and the need for more genetic studies of large scale in natural populations, this decision was arguably the correct one. Time and resources were then focused in this direction.

Not all the initial candidate genes selected from bibliographic work passed through to the final genotype-phenotype tests. One of the limiting steps was that not all these genes were found in the godwit reference genome. The godwit genome version used for this thesis, as is usual in *de novo* genomes, contains a proportion that, due to technical limitations, was not possible to sequence (~16%). The Chicken (*Gallus gallus*) genome is the most complete of the avian genomes, and contains 25,062 genes. The closest genome available to the godwit genome is that of the ruff, which has 18,583 predicted genes. The godwit genome contains 16,538 predicted genes. Thus, we would expect that 35% of the 138 initial candidate genes would be missing (i.e. not found in the godwit genome). However, only eight genes were not found (i.e. *aanat*, *acdyap1*, *dio3*) which means that 95% of the initial candidate genes were found in the godwit genome. Obviously, this expectation of 35% missing genes is an inaccurate approximation. One of these missing genes is the gene *aanat* (arylalkylamine-N-acetyltransferase) which was one of the most promising candidate genes. The transcription of this gene has been found to be regulated by the elements of the CCO, and it is involved in melatonin synthesis in the pineal gland (i.e. it is responsible for the rhythmic production of melatonin) (Kim *et al.* 2007). Under the – albeit unlikely – possibility that this important gene was not present in the godwit genome, I amplified this gene from a few godwits samples together with a few other bird species. As expected, the godwit samples showed amplification bands consistent with the gene *aanat*, implying that it was unfortunately not sequenced in the current version of the godwit genome.

Values of Linkage Disequilibrium (LD) vary along genomes, and the estimation of this parameter is more accurate when genomes have chromosome and phase information of the genotypes. I could not obtain good estimations of LD for the godwit genome, and chromosome information was not available. Specifically, estimates of LD varied largely across the godwit genome (i.e. 10,000 randomly chosen pair of SNPs showed a LD of 1 Mb or less). In order to avoid losing important SNPs by simply using a general value of LD in birds, I kept all SNPs within candidate gene regions, knowing that likely some of them are in LD. This is not a problem for univariate testing, but has implications when correcting for the multiple-testing, since this correction assumes that all tests are independent. Nevertheless, in order to have an idea of how much the rigour of the multiple-testing correction could have masked otherwise significant SNPs, I performed the univariate testing using two general estimations of LD in birds (10 Kb and

100 Kb) and confirmed that no SNP was significant in either comparison. The issue of a potential lack of power due to the multi-testing issue was also tested in two ways. First, I tested the association between only six candidate genes (those of the CCO, which were the initial pre-genome planned targets: *Clock*, *bmal1*, *per2*, *per3*, *cry1*, *cry2*) and migratory timing behaviour, and nothing came out as significant. Secondly, I performed a simulated study including in my dataset three SNPs strongly associated with timing reported in Hess *et al.* (2016), which is a study on the genetics of migration timing in steelhead, with a similar number of individuals and three times more SNPs than in the godwit Chapter 4 study. My simulated analysis could detect those three significant SNPs, showing that a lack of power was not a problem for detecting influential SNPs.

One of the challenges in GWAS in natural populations is the fact that samples sizes are often not too large (i.e. < 1000 individuals) (Kardos *et al.* 2016). Consequently, the resolution that GWAS in natural populations can reach allows for detecting SNPs with medium to large effects. The sequencing of 19 additional genomes from the extreme of the chronophenotype continuum was expected to improve the resolution by enriching the final list of SNPs to be genotyped with the ones more potentially associated to the chronophenotype, therefore, increasing/ improving the power of detection. This approach was useful in that it gave me an idea of how random SNPs associate to the phenotype compared to those SNPs from candidate gene regions (Appendix 3, Figure 3.6), but the variability within each extreme group was higher than expected, and therefore the enrichment of the most promising SNPs might have not been as useful. Nevertheless, this high variability between the two extreme groups (i.e. 10 early vs. 9 late genomes) is in accordance with a scenario of higher complexity than a few genes underlying migratory departure time, which other methods support as shown in Chapter 4.

It is possible that other limitations implied by the methodologies used in Chapter 4 impede the detection of SNPs associated to chronophenotypes. For instance, univariate analyses (i.e. LMM) use either the additive or dominant models, but they don't check for over-dominant or recessive models, and this might have limited the finding of associated SNPs. In that chapter I used a methodology that fits SNP effects in a LMM allowing repeated measurements of the phenotype to be used instead of the means. This method claims to increase the power of detection (Rönnegård *et al.* 2016). This model could have been improved by the inclusion of the interaction of site with genotypes, however, it was not possible due to this methodology not having implemented this option. I used

Lamplink to test whether combination of SNPs could explain chronophenotypes. There is a limitation in Lamplink, which is that it is based on a single genetic model for assessing the association of chronophenotypes with a combination of SNPs. It is more plausible to think that every SNP may work under a different genetic model, therefore combination of SNPs should be possible to be tested under the combination of different genetic models, and this is not (yet) implemented in any pipeline (as far as I know). This could be one of the reasons why this method could not detect the 124 SNPs that, from population genetic methodologies, are capable of explaining some percentage of the chronophenotypes.

The selection of the 124 SNPs that gave rise to the main finding of Chapter 4 could be questioned, because these SNPs were selected not randomly (the bioinformatic tool --thin from VCFtools V0.1.14 used takes the first SNP per scaffold and the next one distant at least a pre-determined number of pair bases). The use of this subset of SNPs was planned to serve as a proxy of SNPs most likely not in LD. It is possible that a different selection of 124 SNPs (i.e. random selection of SNPs within genes instead of only the first one) would have been a more reasonable approach. Nevertheless, under the assumption that those 124 SNPs are representatives (in linkage) of the other 'near-by' SNPs, then I would not expect a different result as that shown in Chapter 4.

In Chapter 4, I corrected the dataset with departures by population structure background for departures from Asia but not from N.Z. (based on findings in Chapter 2). As a consequence of this even quite minor correction, the resolution of the univariate genotype-phenotype association tests is affected such that any apparent genetic effects on timing that were of similar size to the level of the population genetic background becomes insignificant. But what if this level of population structure is actually masking a true association between a SNP and godwit chronophenotypes? There is indeed one SNP that is significant ($p = 0.04$, after BH correction) using a General Linear Model (results not shown) – which is the scaffold1956_position38041 (Gonadotropin Releasing Hormone Receptor –, when not correcting for population structure. Such a SNP could potentially harbour a functional variant. However, such a significant association is not present in the larger N.Z. population where there was complete overlap of genotypes along the chronophenotypes continua. This demonstrates that it is more likely that the significant SNP is actually a false-positive product of the subtle population structure described/reported in Chapter 2, as shown by the spatial distribution of genotypes by the two principal components from the 27 microsatellites (Figure 5.1).

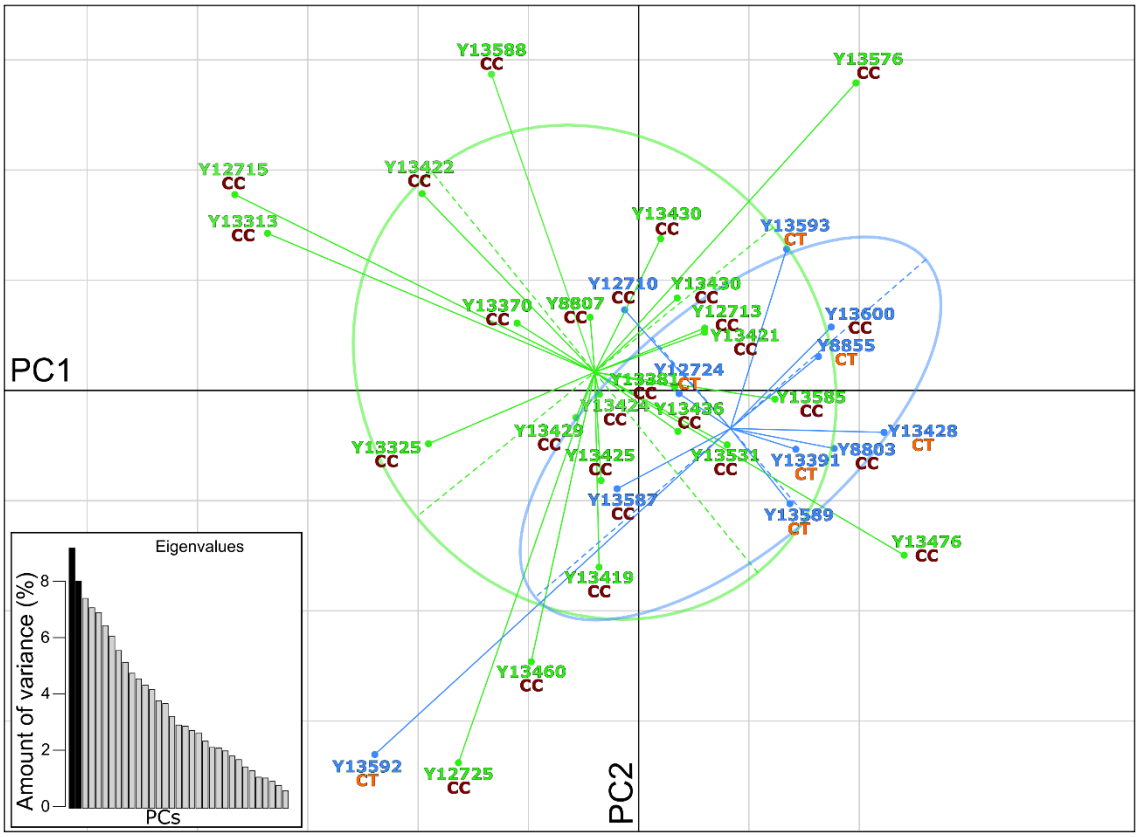


Figure 5.1: Plot of the first two Principal Components from the 27 microsatellites on the 34 individuals with geolocators that showed a significant weak genetic structure (*Fst* = 0.013) when using day 135 as boundary between early and late birds (see Chapter 2). Individuals are indicated by its Y-number (the metal leg band code) and the colours represent birds with mean departure date before (green) or after (blue) day 135. Below each Y-number is the bird's genotype for the SNP (scaffold1956_position38041) that was significant in the univariate analyses. Note how the distribution of these genotypes (i.e. CC and CT) matches the underlying weak population structure, indicating that this gene is not necessarily associated to migratory departure time but is likely a product of non-absolute random mating in the godwit population.

There was no *a priori* expectation of a different phenology between the three sites from which godwits were sampled for this study. The Catlins population, the only one from the South Island, was first tracked and sampled in 2013, that is, at the same time that this project was initiated. To our surprise, the Catlins population showed a significantly shifted phenology with respect to those of Manawatu and Firth of Thames (on average ~ 10–11 days earlier, P.F. Battley *et al*. in prep.). This changed the direction of Chapter 4, in which I accounted for site in the analyses. I would not question that the molecular/genetic mechanisms by which variation of chronophenotypes arises in godwits is the same for different populations. The observed shifted phenology is likely generated by an environmental factor (probably associated to daylength) in a way that we still do not understand. Nevertheless, this difference in phenology between sites was accounted in Chapter 4 by including site as a covariate.

Implications for understanding timing of migration

A primary challenge in biology these days is to understand the effect of genetic variants on complex traits. The study of complex traits, such as behaviours, has undergone a genomic and technological revolution in the last decades (Mackay *et al.* 2009). First studies took approaches to dissect the genetics of migratory timing focused on a few candidate genes (e.g. Fidler *et al.* 2007; Lehtonen *et al.* 2012; Liedvogel *et al.* 2012; Steinmeyer *et al.* 2012; Chakarov *et al.* 2013). It was becoming more and more evident that to fully understand the genetics underlying complex traits, it was necessary to use more exhaustive strategies, and thus current studies are applying genomic approaches with greater coverage (e.g. Santure *et al.* 2013; Kardos *et al.* 2016). However, most of these studies report not having found associations between single genes and the trait in question, but a suggestion of a polygenic effect underlying the genetics of the trait. For example, Santure *et al.* (2013) studied the genetic architecture of clutch size and egg mass in great tits (*Parus major*), and found that the most plausible explanation is a scenario where many loci of small effect contribute to trait variation. In another study, 70 genomes of migratory and non-migratory species were compared at important candidate genes for migration (i.e. *clock*, *per2*, *cry1*, *bmal1*) and found that none of these genes could distinguish migrants from non-migrants (Lugo Ramos *et al.* 2017). In a another comparative genomic study, of the migratory willow warbler (*Phylloscopus trochilus*), no clear differences between subspecies of different migratory behaviour could be found, although the genomic regions with stronger differences were found in a few haplotype blocks (Lundberg *et al.* 2017). A considerable proportion (44%) of the variation in timing of migration of steelhead was explained by 25–50 SNPs (Hess *et al.* 2016). Last, the study on the genetic architecture of avian breeding time done in great tits highlights the absence of (or difficulty in finding) genes of large effect and highlights the necessity of including environmental variables when trying to identify genes associated to complex traits (Gienapp *et al.* 2017). In summary, there is more and more evidence that behind complex traits there is greater complexity than previously expected, at least in natural population of vertebrates. In addition, there is still a general failure in explaining most of the expected genetic heritability, not even from the sum of the small-effects of a number of genes. The findings of Chapter 4 about the genetics underlying migratory timing in godwits is another example that shows how GWAS approaches in natural populations is only able to explain

a limited amount of heritability. This leads to an unavoidable question: if not genetic, then what?

If not genetic, then what (or what else)?

Despite evidence of strong genetic influences in several complex traits, as reported for migratory timing in birds (Pulido *et al.* 2001), there is still a debate on what has been called the 'missing heritability problem' (Manolio *et al.* 2009). This term was mentioned for the first time in 2008 in a study that questioned why, after the identification of thousands of genetic variants from thousands of people, most of these variants were still unable to explain the majority of the heritability of complex traits (i.e. diseases, phenotypes, behaviours). For example, for human height, there were found to be 40 loci associated to human height, yet these loci could explain only 5% of the variance (Visscher 2008). In a study of blackcaps (*Sylvia atricapilla*), migration restlessness was shown to be associated with the *adcyap1* polymorphism, but this association could explain only 2.7–3.5% of the variance of the trait. This is giving us an idea that it is possible that the majority of the variance that the genome code explains is shared by a large number of genes with small effects on the phenotype. If the genetic code cannot explain much by itself, are there other genetic (or non-genetic) aspects that can be a source of phenotypic variance? Indeed, there is more work to be done to understand what are the sources of phenotypic variation in natural populations.

Future directions

Most individuals are capable of modulating their behaviour when exposed to different environmental conditions (i.e. favourable winds, social interaction), showing what is known by plasticity (Helm *et al.* 2005; Dingemanse & Wolf 2013). There is still limited evidence of how birds fine-tune and are able to change their migratory scheduling. Several studies show that populations are changing their phenologies accordingly to changes in temperature (i.e. global warming) (Jonzen *et al.* 2006; Knudsen *et al.* 2011). In godwits, some degree of plasticity or adaptability has been observed on the last years (i.e. mean departure date at Manawatu has advanced; Conklin & Battley, in prep). Plasticity of behaviours have been studied through what is called "behavioural reaction norms" (Dingemanse & Wolf 2010). Some part of this plasticity is due to non-genetic factors (i.e. learned) and some other is attributed to have a genetic basis (Scheiner 1993;

Pigllucci 1996; Dingemanse & Wolf 2013). In contrast, adaptability means evolutionary changes in response to selection which acts on the population (Charmantier & Gienapp 2014). There is still a lack of knowledge about whether these changes are due to population adaptation and/or individual plastic responses (Porlier *et al.* 2012; Charmantier & Gienapp 2014; Usui *et al.* 2017). It is logical to think that temperature could be a determinant environmental factor involved in promoting plasticity of migration timing (Tøttrup *et al.* 2010; Buskirk *et al.* 2012). It might be that the increase of the mean temperature on the breeding grounds in the last years is promoting earlier departures in godwits, but still it is not clear whether this response is a result of individual plasticity or microevolutionary processes or both (Réale *et al.* 2003; Buskirk *et al.* 2012). More studies are necessary to unravel what underlies phenotypic plasticity (i.e. of migratory timing) in order to understand how populations track short-term environmental fluctuations (Charmantier *et al.* 2008; Teplitsky *et al.* 2008) or even how these processes can lead to speciation in some cases (Thibert-Plante & Hendry 2011).

The majority of behaviours in nature are refined by individuals' experience. There is still poor knowledge about how is experience added on the top of a 'genetic program'. Several studies show that first-time migrants vary more in their migratory behaviours (i.e. departure time, orientation) than experienced adults. Thus, it is logical to think that studying migration traits in inexperienced individuals would bring more possibilities to dissect the genetics underlying migratory traits. Indeed, performing GWAS using phenotypes from inexperienced individuals promises to be a good direction for future genetic studies on migratory traits. This possibility would bring not only insight about how experience could be affecting GWAS in which experience individuals have been used, but also it is an improved approach to unravel potential large-effect genes associated to migratory traits.

As explained in the introduction, it is interesting to note that in godwits, young individuals stay in their over-wintering grounds for 2 or 3 years before performing their first migration. This means that there is a mechanism that works, somehow, 'switching on' the migratory behaviour. Epigenetic marks have been pointed out as a mechanism that can mediate the transition between behavioural states (Ledon-Rettig *et al.* 2013). Under the assumption that epigenetic marks might be modulating the change between non-migrant to migrant stages, it would be interesting to compare epigenetic profiles (i.e. at candidate genes) of individuals at the stage of young (non-migrant) and at the adult

state (migrant). Indeed, in addition to nucleotide variation in the DNA epigenetic polymorphism could be a pivotal aspect of genetics explaining migratory traits (i.e. migratory timing), and it has been poorly explored (i.e. only in barn swallows and in the gene Clock, Saino *et al.* 2017). In the context of explaining timing traits, a good idea would be to assess the epigenetic profiles of a few genes (i.e. genes of the CCO: *Clock, bmal1*, *per2*, *per3*, *cry1*, *cry2*; or genes related to the HPG axis: *aanat*, *acdyap1*, *dio3*, *eya3*) or – much better although more costly– the genome-wide profiles and test for association to earlier-later migrants as well as before-after migratory periods. A second possibility for future epigenetic studies on migration traits in birds would be to explore the possibility that individual migration phenotypes might be determined from birthplace. It is possible that in a system like godwits, where individuals are born along a wide range of photoperiodic conditions, this could have generated an epigenetic mark that determines adults' migratory timing 'preference' via environmental-conditioned genomic mark (Barrett & Page 1989; Canal-Corretger *et al.* 2001; Ciarleglio *et al.* 2011). However, carrying out epigenetic studies of circannual-related behaviors in protected species like godwits would be limited to blood samples (Saino *et al.* 2017), but this is not the ideal as epigenetic marks are tissue-specific (Ledon-Rettig *et al.* 2013). I believe it is still not clear whether epigenetic marks measured in blood can serve as a proxy of the epigenetic changes in other tissues, and this can compromise the robustness of the study. There must be other systems of similar characteristics in which these questions can be addressed.

Final remarks

In conclusion, this thesis provides evidence of the need for including more complexity when studying complex behaviours in natural populations. In the particular case of migratory departure date in godwits, this complexity is reflected at its genetic level, with possibly a large number of genes being involved. Genetics factors are surely not the only factor determining an individual's behaviour, at least in vertebrates, with a number of other genetic and non-genetic factors contributing to it. Furthermore, is precisely this complexity which makes it difficult to reach to generalizable conclusions regarding the sources of phenotypic variation in nature, since it is probably shaped by species-specific contexts. Nevertheless, science, advances in technology and human creativity – and persistency – would be surely the ingredients that would allow

overcoming the difficulty imposed by this complexity – as it has always been. There will without doubt be currently unthought of discoveries in this field to come.

Appendices

I know, it is not a godwit, it is a lesser knot (also known as red knot, *Calidris canutus*) that I am releasing in this picture, but in essence, the release of godwits looks the same. Photo taken by Phil Battley at the Manawatu River estuary, November 2013.

Appendix 1

Table 1.1 Summary statistics for the 30 microsatellite loci used in this study. GenBank accession numbers for sequences of one allele of each locus are listed. The three microsatellite loci that may harbour null alleles (as detected using MICRO-CHECKER v.2.2.3) or which display statistically significant deviation from HWE ($P < 0.05$ after Bonferroni correction, as detected using ARLEQUIN v.3.5.1.3.) are indicated in underlined italic script. *Abbreviations*: n, number of individuals genotyped at the locus; *Na*, number of alleles detected; *Ho*, observed heterozygosity; *He*, expected heterozygosity, HWE, Hardy-Weinberg equilibrium; No, not statistically supported; Yes, statistically supported. [1]analyses using MICRO-CHECKER v.2.2.3; [2]analyses using ARLEQUIN v.3.5.1.3.

Locus name/ GenBank acc. no.	n	Na	Ho	He	Possible null alleles[1]	Possible deviation from HWE[2]
LIM5 FJ652579	222	7	0.324	0.354	No	No
LIM6 FJ652593	218	4	0.142	0.133	No	No
LIM8 FJ652580	221	7	0.466	0.513	No	No
LIM12a FJ652583	218	4	0.431	0.446	No	No
LIM12b FJ652584	221	4	0.231	0.227	No	No
LIM25 FJ652588	220	2	0.359	0.323	No	No
LIM30 FJ652590	219	11	0.840	0.849	No	No
LIM32 FJ652591	220	8	0.668	0.751	No	No
C046 KT967115	218	8	0.789	0.798	No	No
C994 KT967116	219	3	0.183	0.186	No	No
C856 KT967117	219	10	0.836	0.803	No	No
C586 KT967118	222	4	0.212	0.253	No	No

C734 KT967119	218	4	0.248	0.274	No	No
C008 KT967120	220	6	0.582	0.555	No	No
C348 KT967121	220	9	0.541	0.604	No	No
C501 KT967122	218	5	0.486	0.493	No	No
C367 KT967123	219	5	0.333	0.352	No	No
C772 KT967124	221	3	0.041	0.040	No	No
C822 KT967125	222	5	0.320	0.343	No	No
C697 KT967126	217	11	0.733	0.760	No	No
C355 KT967128	218	8	0.610	0.628	No	No
C980 KT967129	219	4	0.242	0.219	No	No
C444 KT967131	220	9	0.727	0.758	No	No
C893 KT967132	188	5	0.426	0.422	No	No
C463 KT967133	220	5	0.227	0.244	No	No
C129 KT967134	220	8	0.609	0.664	No	No
C311B KT967135	220	5	0.718	0.718	No	No
LIM26 FJ652589	203	14	0.650	0.822	Yes	Yes
C267 KT967127	212	4	0.288	0.335	Yes	Yes
C998 KT967130	220	11	0.650	0.801	Yes	Yes

Table 1.2 PCR primer pairs (n = 30) used for amplification of godwit microsatellite loci. Names of the loci are indicated with those from Verkuil *et al.* 2009 prefixed 'LIM' and those from this work prefixed 'C'. GenBank accession numbers for representative allelic sequences at each locus are listed. Underlined primer sequencers are (i) forward primers, M13(-21) universal primer 5' tag sequence and (ii) reverse primers, PIG-tail to promote addition of 3' As.

Locus/GenBank Accession nº	Primer sequence (5'-3')	Repeat motif/ nº repeats
LIM5 **FJ652579**	F: TGTAAAACGACGGCCAGTACTGCTGCTTCCAAATGACA R: GTTTTCTCCCCTCCATCTGAAAAG	$(GA)_{10}AA(GA)_2$
LIM6 **FJ652593**	F: TGTAAAACGACGGCCAGTTATTCAGCAGAACACACGCAC R: GTTTACTGGTATTCTTTAACCCCGACTG	$(AC)_5GC(AC)_6$
LIM8 **FJ652580**	F: TGTAAAACGACGGCCAGTCTGAAGTGATCAGGCAAGGTG R: GTTTTGTGGAGGAAGGAGGCTTG	$(CA)_{12}$
LIM12a **FJ652583**	F: TGTAAAACGACGGCCAGT GGTTCTCTTTGGGCTGTCTG R: GTTT AGGAAACTCATCTGGGGTCTG	$(CA)_{10}$
LIM12b **FJ652584**	F:TGTAAAACGACGGCCAGTGCCAAAATATTTGACAGACCCCAG R: GTTT GAGTTTCCCAGCACTTTGCCC	$(CA)_{11}$
LIM25 **FJ652588**	F: TGTAAAACGACGGCCAGTTGACACCAGACAGTGTTGCAT R: GTTTCCGACTTTATTTGGTTTTCCAG	$(CA)_4AAAA(CA)_8$
LIM26 **FJ652589**	F: TGTAAAACGACGGCCAGTTCAGGTGGACGTACAGCAC R: GTTTGCAAGTCAAGGGGTGGAAC	$(CA)_{11}$
LIM30 **FJ652590**	F: TGTAAAACGACGGCCAGTACCTTAGTACATGGGGAACAG R: GTTTTGAAGGCATATCTGGGGATGTC	$(CA)_{10}$
LIM32 **FJ652591**	F: TGTAAAACGACGGCCAGTTCAGACGTCGATCACCTGAG R: GTTTTGAAACTATAAATCCTGCGGG	$(CA)_9CC(CA)_6$
C046 **KT967115**	F: TGTAAAACGACGGCCAGTGTGCTCAGAGATAGGCAAACC R: GTTTCTGAAGTGAAAAGAGCGGGG	$(TATC)_9$
C994 **KT967116**	F: TGTAAAACGACGGCCAGTTTGATGAGAATCCTACCCCTG R: GTTTGCGTCACAGTATGAACCCAC	$(TTTG)_8$
C856 **KT967117**	F: TGTAAAACGACGGCCAGTTGTGAACCAGGGTGCTACTC R: GTTTTTTAAGGTGGCCAAAGCAGG	$(TAT)_{12}$
C586 **KT967118**	F: TGTAAAACGACGGCCAGTGCACTCTCAGTCCACAATGC R: GTTTACACCTGCCAGTAAGAAACAG	$(ATTT)_7$
C734 **KT967119**	F: TGTAAAACGACGGCCAGTACTTACAGTGCTTCTCTTGGC R: GTTTTAGCCTGGGTTTTTCCTGTG	$(GTTT)_8$
C008 **KT967120**	F: TGTAAAACGACGGCCAGTCAGGCTTCCCCAAAATGTCC R: GTTTGGAACTGGCAGCTTAATGGC	$(AAT)_7$
C348 **KT967121**	F: TGTAAAACGACGGCCAGTCTGGACGGACATGGAGTCTG R: GTTTCACAACAACTGAGTAGGGGC	$(TG)_{13}$
C501 **KT967122**	F: TGTAAAACGACGGCCAGTTTGCTGACTCCTGAGACCAC R: GTTTTGCCAGTTCTCACCCCATAC	$(TTCA)_7$
C367 **KT967123**	F: TGTAAAACGACGGCCAGTAACTGTTGCGTAGTGTTCGG R: GTTTTTAGCAGGGCCCAAAGAAAC	$(GAAG)_7$
C772 **KT967124**	F: TGTAAAACGACGGCCAGTTGCAGAGACCCCTAAAGCTG R: GTTTTTTGGCCCTTTACATTCGGC	$(CAT)_7$
C822 **KT967125**	F: TGTAAAACGACGGCCAGTTGCCCCCTAAGATACACTGC R: GTTTTCACGCCCCTGTGACTAATG	$(TTG)_8$
C697 **KT967126**	F: TGTAAAACGACGGCCAGTTCCATGAGGCAAATGCCAAG R: GTTTGGGATGTGTAGCTGGGGTAG	$(GATA)_8$
C267 **KT967127**	F: TGTAAAACGACGGCCAGTTACAGCTGCAACAAACGAGC R: GTTTGCAACATCTAGTGGCACACC	$(TG)_{12}$

C355 **KT967128**	F: TGTAAAACGACGGCCAGTCATTCTCCTTCGGGCTTTGC R: GTTTGTTTGTCAGGTGCTTCCTCG	$(AAC)_8$
C980 **KT967129**	F: TGTAAAACGACGGCCAGTGTCCATGCAAAGCCCATCAG R: GTTTTCTTCCAGTTCAGTCTTGGC	$(ATT)_8$
C998 **KT967130**	F: TGTAAAACGACGGCCAGTTAGTTCAACAGCAGGACCCC R: GTTTGATCTACAGGGTGGAGGAGC	$(ATT)_9$
C444 **KT967131**	F: TGTAAAACGACGGCCAGTAGCTGGGAAGATGTGCCTG R: GTTTGCACAGGGGAAGGTGAAATG	$(GAAG)_9$
C893 **KT967132**	F: TGTAAAACGACGGCCAGTTCCATCTTTAGTCCCCTGCC R: GTTTTGCTCTCTGAGGTGAATGCC	$(AT)_{13}$
C463 **KT967133**	F: TGTAAAACGACGGCCAGTTCTCCCAGTGTCTGTTTCCG R: GTTTGAGGAGCAGTTTTACGTGCG	$(GGA)_8$
C129 **KT967134**	F: TGTAAAACGACGGCCAGTCAAGTGGAGCTGTTCCCTCC R: GTTTAACATTTCTCCTCCCCGCTC	$(AGAA)_{10}$
C311B **KT967135**	F: TGTAAAACGACGGCCAGTCCCTGCTCACCTAACAGACC R: GTTTCACGCTGCATGGAAGGATG	$(TTG)_7$

Table 1.3 Summary of population genetic structure results in relation to southern (i.e. Yukon-Kuskokwim Delta) and northern breeders (i.e. Seward Peninsula and North Slope) (Alaska_geoloc.); to 'earlier' and 'later' chronophenotype-based groups for the Asian (Asia_chronoph.) and N.Z. (N.Z._chronoph.) datasets using *a priori* defined timing divides at days 128 and 135 for the Asian and at day 81 for the N.Z. datasets (see text for more details). Comparisons were made using three methods: AMOVA-*F*st, Bayesian and *K*-means clustering and using 30 microsatellites. (*) denotes significant *Fst* at $P < 0.05$ without Bonferroni correction; No p-value was significant after Bonferroni correction ($P > 0.05$). *F*st values with Interval of Confidence (IC) = 95% are underlined. Numbers in italics are the numbers of individuals per group in the comparisons. *Abbreviation*: n, number of microsatellites; YKD, Yukon-Kuskokwim Delta; SP, Seward Peninsula; NS, North Slope.

Dataset	Divides	AMOVA-F_{ST}	Bayesian (*K*)	*K*-means (*K*)
Asia_geoloc. n = 36	YKD vs. SP+NS *19/17*	0.010(*)	1	1
Asia_chronop. n = 36	Day 128 *16/20*	0.011(*)	1	1
	Day 135 *25vs.11*	0.013(*)	1	1
N.Z._chronop. n = 165	Day 81 *111/45*[1]	0.002	1 (2)	2 (1)
N.Z._chronop. n = 36	Day 81 *24/11*[2]	0.005	1	1

[1] Nine individuals and [2] one individual were excluded for having their chronophenotype (mean departure date) = day 81.

Figure 1.1 Bayesian clustering mean log likelihood values (*L(K)*) plotted as a function of number of genetic clusters (*K*) using the software STRUCTURE, setting the admixture and prior-location models for the Asian dataset (n = 36), where light-levels of geolocators and biometrics where used to classify individuals into southern and norther breeders (see text for more details). Two datasets of microsatellites were used: 27 microsatellites (a) and 30 microsatellites (b).

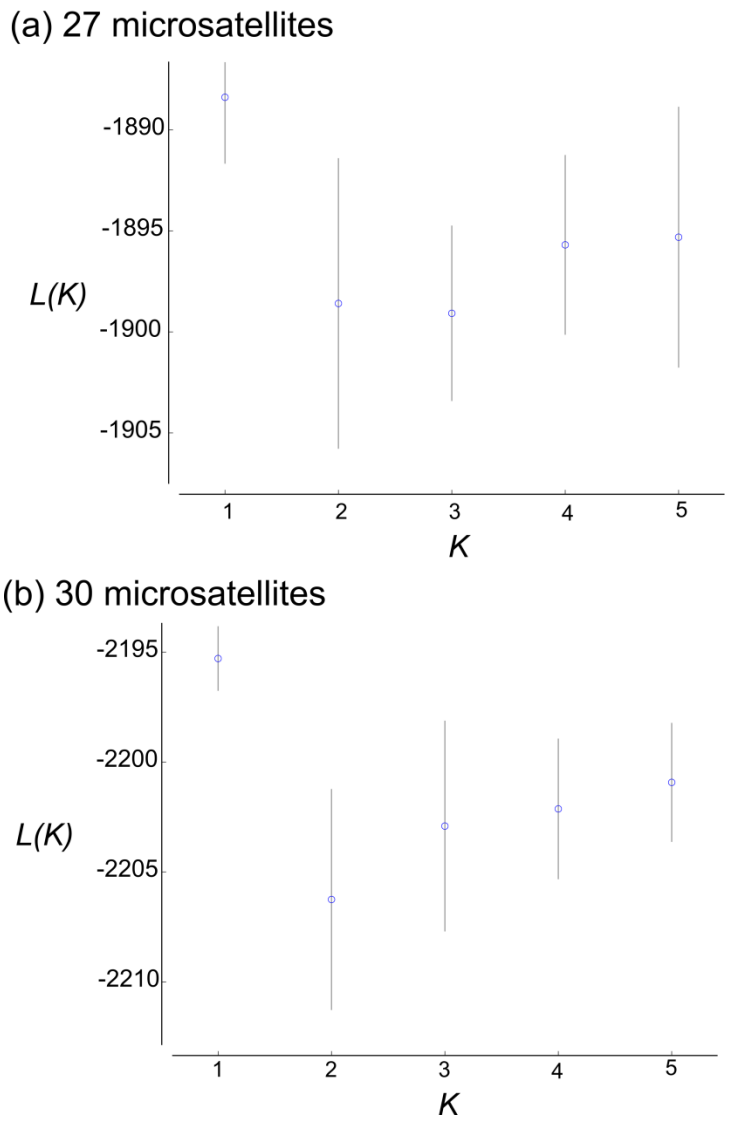


Figure 1.2 *K*-means clustering Bayesian information criterion (BIC) values plotted as a function of number of genetic clusters (*K*). Analyses of microsatellite genotypes from 36 individual godwits with geolocators and from which breeding latitude was derived from light-level information of geolocators and biometrics (see text for more details). Analyses were carried out using (a) 27 microsatellite loci and (b) 30 microsatellite loci.

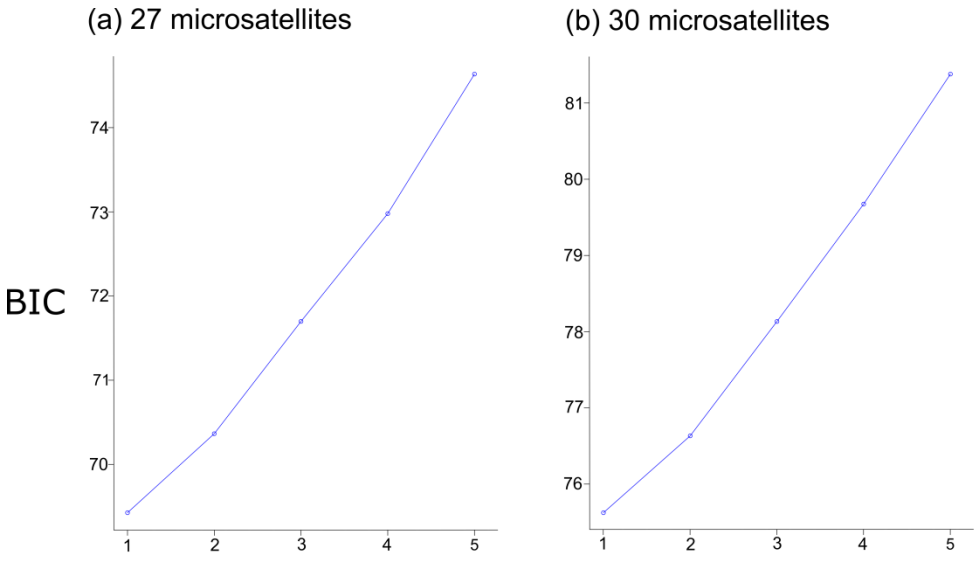


Figure 1.3 Bayesian clustering mean log likelihood values (*L(K)*) plotted as a function of number of genetic clusters (*K*) using the software STRUCTURE, setting the admixture and prior-location models for the Asian dataset. (a, b) Chronophenotype divide at day 128, groups: '< day 128', n = 16; '> day 128', n = 20; (c, d) chronophenotype divide at day 135: '< day 135', n = 25; '> day 135', n = 11. Two datasets of microsatellites were used: 27 microsatellites (a, c) and 30 microsatellites (b, d).

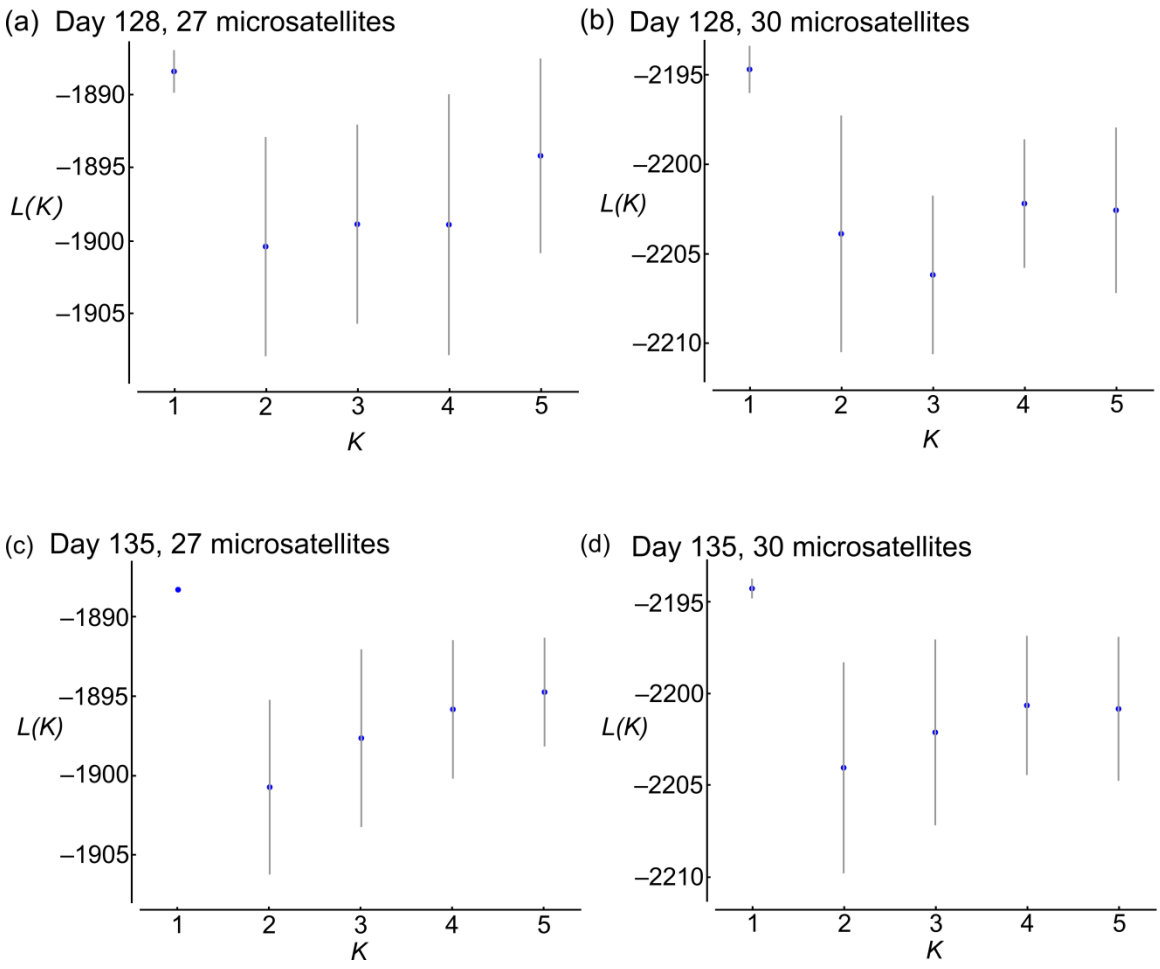


Figure 1.4 *K*-means clustering Bayesian information criterion (BIC) values plotted as a function of number of genetic clusters (*K*). Analyses of microsatellite genotypes from 36 individual godwits with departure dates from eastern Asia (stop-over) were carried out using (a, c) 27 microsatellite loci and (b, d) 30 microsatellite loci. *A priori* groups were formed based on two chronophenotypes divides: (a, b) '< day 128, n = 16; '> day 128', n = 20; (c, d) '< day 135', n = 25; '> day 135', n = 11.

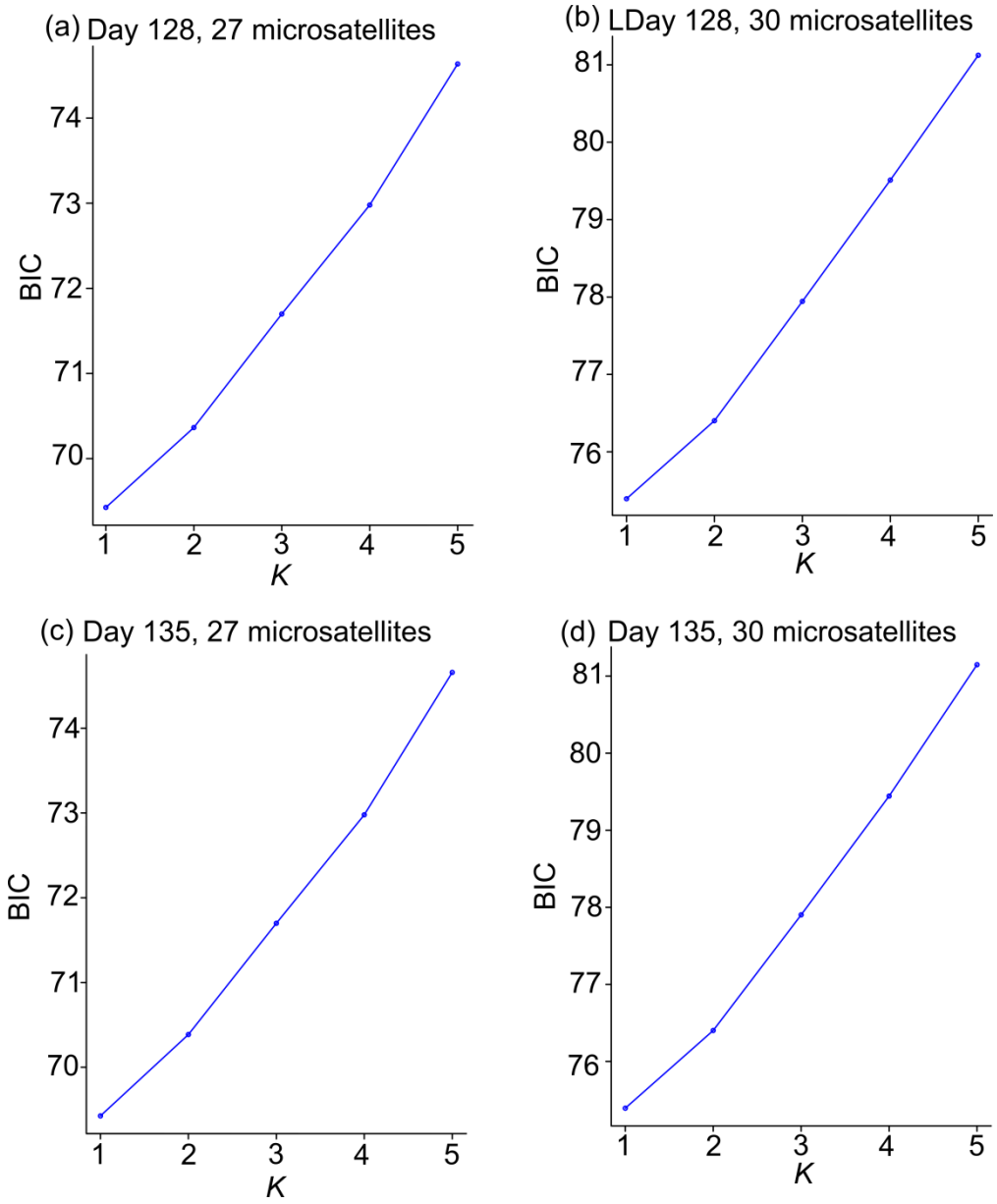


Figure 1.5 Number of clusters (*K*) estimated by STRUCTURE for the N.Z. dataset assuming *a priori* groups based on a chronophenotype divide at day 81. STRUCTURE was run using the admixture and prior-location models. Groups compared are: '< day 81', n = 111; '> day 81', n = 45. Two datasets of microsatellites were used: (a) 27 microsatellites, (b) 30 microsatellites. *Abbreviations*: *L(K)*, Bayesian clustering mean log likelihood.

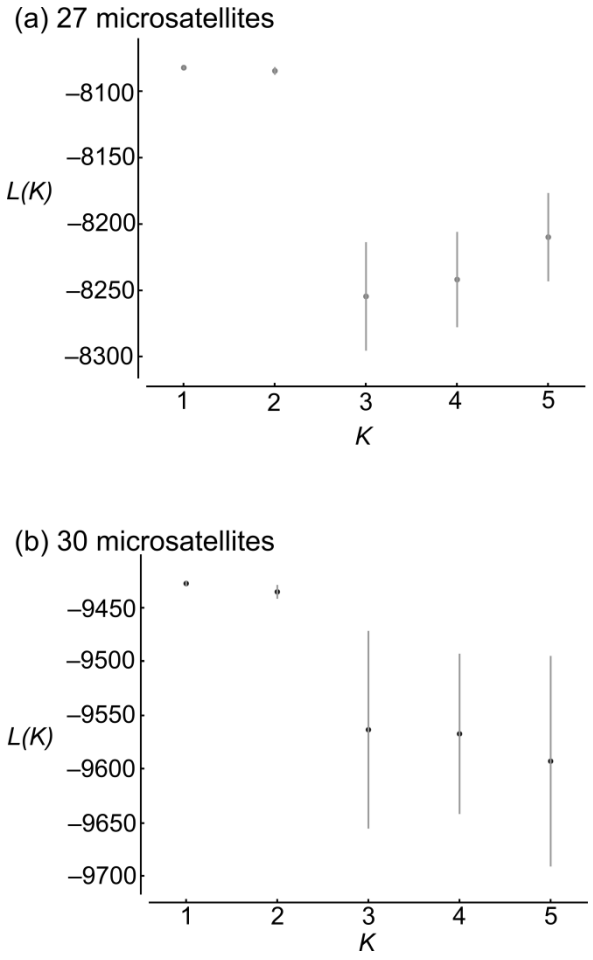


Figure 1.6 Number of estimated clusters (*K*) by *K*-means algorithm for the 156 godwits with N.Z. departure dates grouped by presumptive breeding latitude based on a timing divide at day 81 (111 birds '< day 81' and 45 birds '> day 81') using *K*-means clustering Bayesian information criterion (BIC) values plotted as a function of number of genetic clusters. Analyses were carried out using 27 (left) and 30 (right) microsatellite loci.

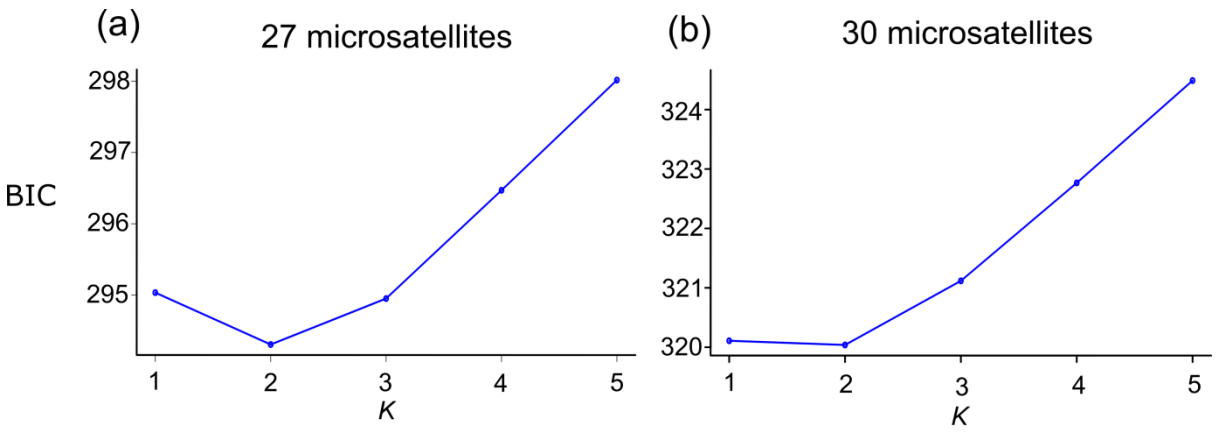


Figure 1.7 Bayesian clustering mean log likelihood values (*L(K)*) plotted as a function of number of genetic clusters (*K*) using the software STRUCTURE, setting the admixture and prior-location models for the Asian dataset (n = 36), where chronophenotypes and a timing divide at day 81 were used to classify individuals into 'earlier' vs. 'later' groups (see text for more details). Note that one individual was discarded for these analyses since its departure date felt at day 81. Two datasets of microsatellites were used: 27 microsatellites (a) and 30 microsatellites (b).

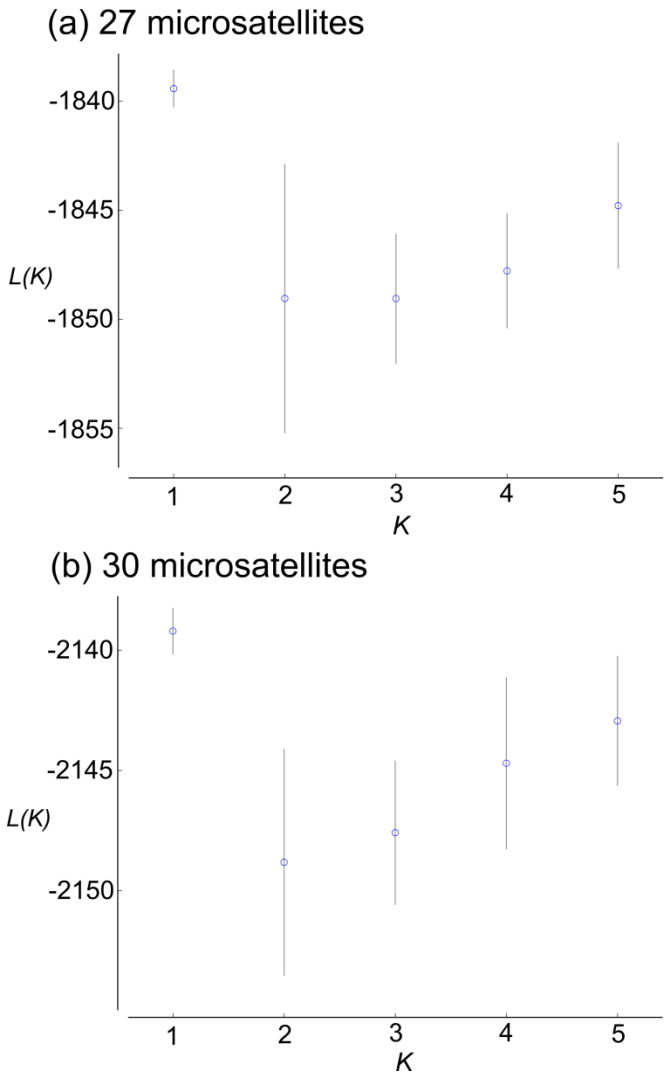


Figure 1.8 Number of estimated clusters (*K*) by *K*-means algorithm for the 36 godwits with geolocators classified in *a priori* timing groups based on their chronophenotype from N.Z. and on a timing divide at day 81 (24 birds '< day 81' and 11 birds '> day 81') using *K*-means clustering Note that one individual was discarded for these analyses since its departure date felt at day 81. Bayesian information criterion (BIC) values plotted as a function of number of genetic clusters. Analyses were carried out using 27 (left) and 30 (right) microsatellite loci.

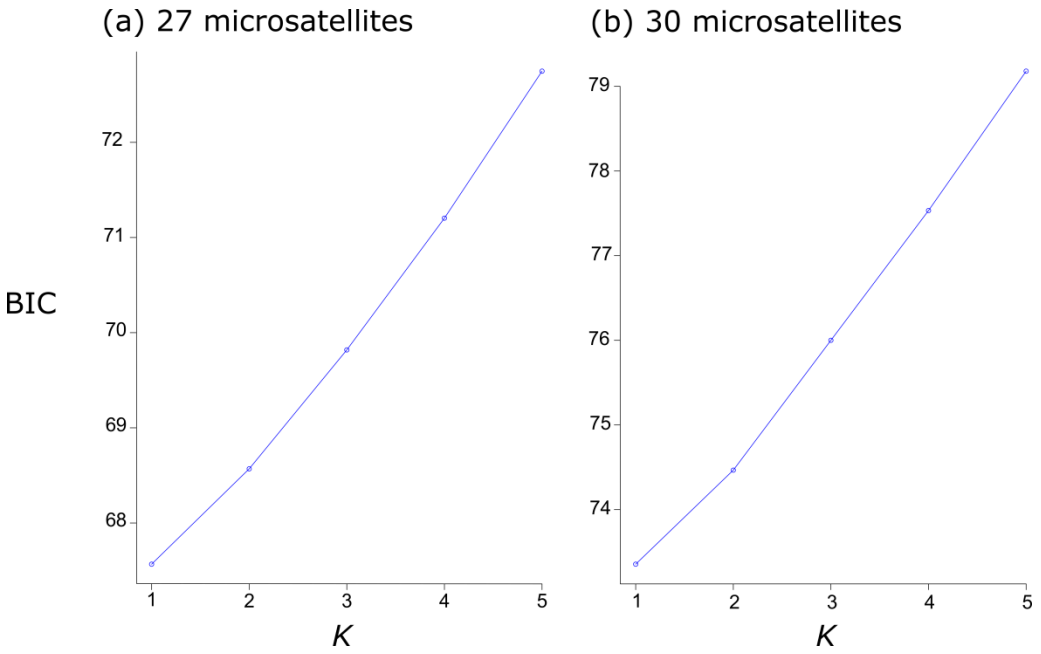


Figure 1.9 Bar-plots generated by STRUCTURE using the set of 30 microsatellites in the different datasets (i.e. the Alaskan, Asian and N.Z. datasets). (a) Using the Alaskan breeding regions derived from light-levels of geolocators and biometrics (see text for more details) as prior location. Groups are: 19 birds 'southern breeders' (Yukon-Kuskokwim Delta) and 17 birds 'northern breeders' (Seward Peninsula and North Slope). (b) using the Asian dataset (n = 36) classified in *a priori* 'earlier' and 'later' departers groups based on their chronophenotype derived from 1–2 recorded departure dates from eastern Asia (stop-over) and using two timing divides at day 128 and (c) at day 135. (d) Using godwit population with departure date for migration from N.Z. (n = 165) and *a priori* groups based on their chronophenotype and a date divide at day 81 with 111 birds '< day 81' and 45 birds '> day 81'. Note that 9 individuals of the initial 165 godwits in the sample were excluded for this analysis as their chronophenotype felt at day 81. (d) Using the godwit dataset with geolocators and chronophenotypes from N.Z. Analyses were made with *a priori* groups based on a date divide at day 81 with 24 birds '< day 81' and 11 birds '> day 81'. Note that one individual of the initial 36 individuals in the sample was excluded for this analysis as their chronophenotype felt at day 81. Bars indicate the admixture proportion of individuals [0;1] assuming $K = 2$. When using chronophenotypes (i.e. (b) (c) (d) and (e)), individuals are organized from the earliest to the latest departers from left to right. Day 1 = 1 January.

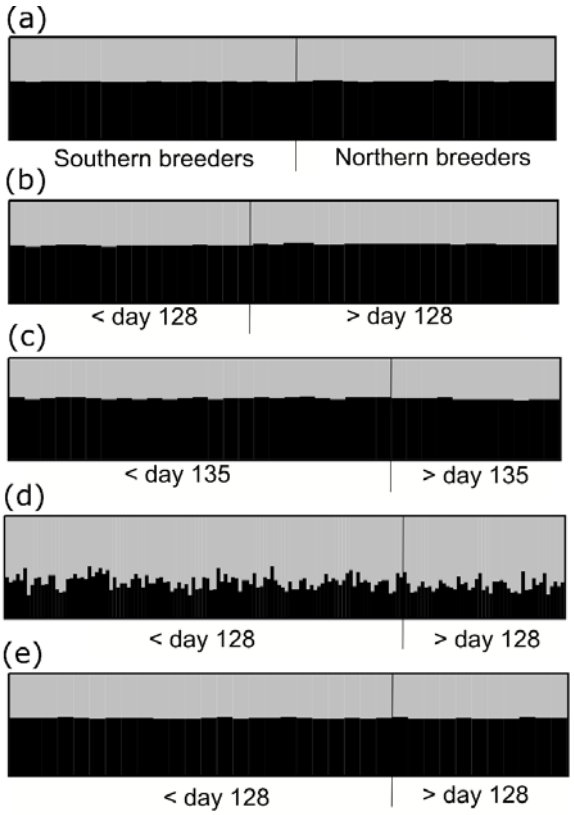


Table 1.4 Breeding locations and mean departure dates (chronophenotypes) from eastern Asia (stop-over) and N.Z. (over-wintering sites) of 165 godwits. Breeding locations (i.e. Yukon-Kuskokwim Delta, Seward Peninsula, North Slope, denoted 'YKD', 'SP' and 'NS' respectively) were determined from geolocators light levels and biometrics (Conklin *et al.* 2010). Departure dates from eastern Asia of 36 godwits derived from geolocators over the years 2008–14 (all from the Manawatu River estuary). Departure dates from N.Z. were collected by direct observation over the years 2004–16. Mean departure dates from eastern Asia and N.Z. for each individual were determined from all the years available (1 or 2 departure dates for the Asian dataset; 1–9 departure dates per bird for the N.Z. dataset). Godwits are listed in order from the earliest to the latest departers from N.Z. *Abbreviations*: Bird ID denotes the wintering site as the first letter (C = Catlins Coast; F = Firth of Thames; M = Manawatu Estuary) followed by a unique metal band number starting with Y, e.g. Y12630; YDD, number of years with departure date available. Note that day 1 = 1st January. Individuals whose known breeding location does not coincide with predicted breeding location by their departure date from Asia or N.Z. are shaded.

Bird ID	YDD	Breed_loc	Mean departure date from eastern Asia (2008–14)	Mean departure date from N.Z. (2004–16)
CY12630	1			59
CY12616	3			60
FY13144	2			61
FY13240	2			62
CY12618	3			63
CY12617	3			65
CY12622	3			65
CY12619	4			66
CY12386	2			66
CY12629	4			66
CY12615	2			67
MY13419	3	YKD	123	67
CY12624	4			68
CY12626	4			68
MY8861	3			68
FY13172	2			68
CY12627	4			69
MY11438	1			69
MY11440	1			69
FY13239	1			69
MY13313	6	YKD	118	70
MY13476	2	YKD	119	70

MY12485	9			70
CY12620	1			71
MY13311	4			71
MY12716	3			71
MY12725	3	YKD	119	71
MY13331	3			71
MY13471	3			71
MY13599	4			71
FY13142	1			71
FY13238	3			71
FY13297	4			71
MY13422	6	YKD	121	71
MY13585	3	YKD	130	71
MY13586	3			71
FY13244	2			72
FY13277	1			72
MY12723	2			72
MY12721	3			72
MY13583	4			72
MY13393	6			72
MY12718	3			72
MY13531	3	SP	130	72
CY12631	2			73
MY13587	4	YKD	136	73
FY13282	2			73
MY13578	6			73
MY13300	3	YKD	125	73
MY8807	4	YKD	123	73
FY13179	1			73
FY13180	1			73
FY13208	1			73
MY13425	9	YKD	123	73
MY13429	6	YKD	123	73
MY13581	4	YKD	124	73
MY12705	3			73
FY13175	1			74
FY13237	1			74
MY13582	3			74
FY13335	1			75
MY13299	8			75
MY13588	4	SP	134	75
MY12717	3			75
MY13323	3			75

MY13314	9			75
MY13328	5	YKD	122	75
CY12628	4			76
FY13104	2			76
FY13138	1			76
FY13173	1			76
FY13228	1			76
MY13373	2			76
MY13515	3			76
MY13591	2			76
MY13325	9	YKD	121	76
FY13176	2			77
FY13183	1			77
FY13265	1			77
FY13167	3			77
MY13389	4			77
MY13384	4			77
MY12726	3			77
FY13147	1			78
MY12722	2			78
MY8851	4			78
MY12720	1			78
MY12719	3			78
FY13162	2			79
FY13280	1			79
FY13294	1			79
FY13343	1			79
FY13165	5			79
MY13424	8	YKD	124	79
MY13592	3	SP	137	79
MY13460	4	NS	134	79
MY12708	1			79
MY13506	1			79
MY13590	1			79
MY13394	6			79
MY13329	7			79
FY13177	1			80
FY13344	1			80
MY13598	2			80
MY13426	9	YKD	119	80
MY13600	3	NS	139	80
MY8803	3	SP	136	80
MY12706	4			80

MY12713	3	YKD	124	80
MY13596	1			80
MY8855	4	SP	136	80
FY13163	3			81
FY13110	1			81
FY13195	1			81
FY13227	1			81
FY13274	2			81
FY13341	1			81
MY12710	2	SP	137	81
MY13377	1			81
MY13383	2			81
FY13136	1			82
FY13181	2			82
FY13229	1			82
MY13576	6	NS	132	82
MY11427	1			82
MY13301	1			82
MY13302	4			82
MY13597	2			82
MY8854	4			82
MY13589	3	SP	148	82
FY13283	2			83
FY13296	2			83
FY13125	2			83
FY13170	1			83
FY13171	2			83
FY13295	1			83
FY13336	1			83
FY13112	3			83
MY13430	9	SP	133	83
MY13375	1			83
MY13593	1	YKD	137	83
MY13370	9	SP	133	83
FY13102	4			84
FY13278	1			84
FY13346	1			84
MY12704	4			84
FY13334	2			85
FY13231	3			85
FY13287	3			85
FY13137	3			85
MY13391	6	NS	142	85

MY12724	1	SP	156	85
FY13249	4			86
FY13345	3			86
MY13381	5	SP	133	86
MY12715	3	SP	132	86
MY13388	2			86
MY13428	6	SP	140	86
MY8814	1			86
MY13421	5	YKD	124	86
FY13226	1			88
MY13318	3			88
MY13333	4			88
FY13245	2			89
FY13279	1			91

Figure 1.10 The relationships between migration timing and breeding latitude that were used to derive *a priori* behavioural-based groupings, based on Conklin *et al.* (2010). Day 1 = 1 January. For N.Z. departures, day 81 separated the chronophenotype groups '< day 81' (Yukon-Kuskokwim Delta breeders, 'earlier departers') and '> day 81' (Seward Peninsula and North Slope breeders, 'late departers'); birds with chronophenotype on day 81 were omitted from analyses. For Asia departures, day 128 separated < day 128 (Yukon-Kuskokwim Delta breeders, 'earlier departers') and > day 128 (Seward Peninsula + North Slope, 'later departers') and day 135 separated < day 135 (Yukon-Kuskokwim Delta + Seward Peninsula, 'earlier departers') and > day 135 (North Slope, 'later departers'). Figures were modified from Conklin *et al.* (2010).

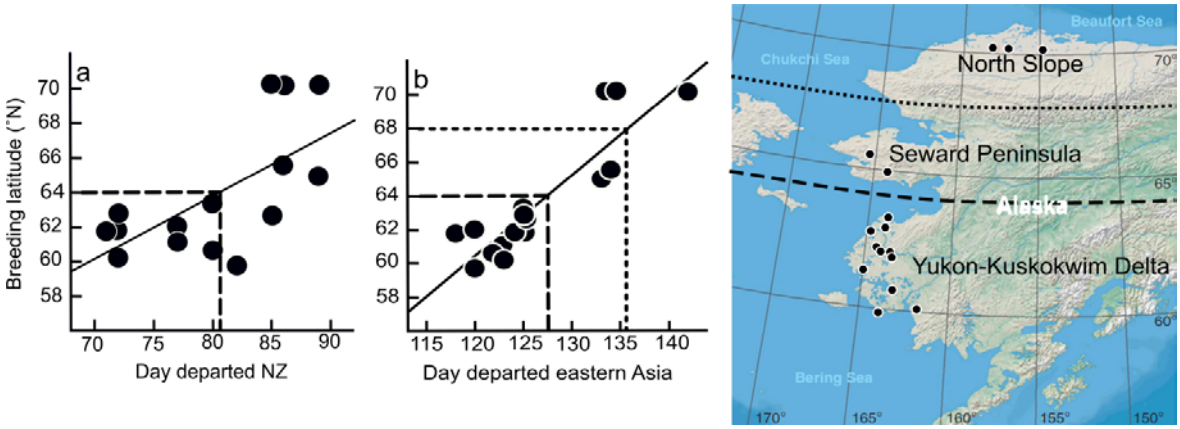


Table 1.5 Bayesian clustering results summary using even samples sizes. These analyses were performed in the software STRUCTURE. Samples sizes were equalized by resampling the larger group to match the sample size of the smaller in each comparison (see numbers in italics). Comparisons were made using 27 or 30 microsatellites. For all the analyses, either using 27 or 30 microsatellites, we obtained that $K = 1$ was the most likely number of clusters but with one exception: The N.Z. using 27 microsatellites. Despite 7 out of the 10 replicates gave $L(K)$ values at $K = 2$ slightly higher than at $K = 1$, standard deviations were always larger. Therefore, we interpreted this result as most likely no structure with a hint for some structure noted as 1 (2) in the table. *Abbreviations*: $n_{original}$, sample sizes per group (southern/northern) in the main dataset; $n_{equalized}$, sample sizes per group in the standardized dataset; n, number of microsatellites.

Dataset	Divides	n	Bayesian (*K*)
Asia_chronop.	Day 128	27	1
n = 36	$n_{original}$ = 25/11		
	$n_{equalized}$ = 11/11	30	1
N.Z._cronop.	Day 81	27	1
n = 36	$n_{original}$ = 25/11		
	$n_{equalized}$ = 11/11	30	1
N.Z._chronop.	Day 81	27	1 (2)
n = 156	$n_{original}$ = 111/45		
	$n_{equalized}$ = 45/45	30	1

Appendix 2

Table 2.1 Photoperiods (daylengths) in the three sampling sites (Firth of Thames, Manawatu River estuary, Catlins Coast) over the godwit's migratory period (mid-February to early-April). These photoperiods were obtained using the website of The Astronomical Application Department of the U.S. Naval Observatory (http://aa.usno.navy.mil/data/docs/RS_OneDay.php). Last column indicates differences in photoperiod between the two most distant sites (Firth of Thames and Catlins Coast, 9˚ degrees latitude difference). Highest photoperiod difference between these two sites is indicated in bold and underlined. *Abbreviations*: Hr. = hours; Min. = minutes.

Gregorian calendar	Day 1 = 1st January	Firth of Thames (Longitude E175° 25', Latitude S37° 4')		Manawatu River estuary (Longitude E175° 13', Latitude S40° 28')		Catlins Coast (Longitude E169° 38', Latitude S46° 28')		Firth of Thames photoperiod (Min.)	Catlins Coast photoperiod (Min.)	Difference (Min.)
		Daylength		Daylength		Daylength				
		Hr.	Min.	Hr.	Min.	Hr.	Min.			
15-Feb	**46**	13	28	13	39	14	2	808	842	**34**
16-Feb	**47**	13	26	13	36	13	58	806	838	32
17-Feb	**48**	13	24	13	34	13	55	804	835	31
18-Feb	**49**	13	22	13	31	13	52	802	832	30
19-Feb	**50**	13	19	13	29	13	48	799	828	29
20-Feb	**51**	13	17	13	26	13	45	797	825	28
21-Feb	**52**	13	15	13	23	13	42	795	822	27
22-Feb	**53**	13	13	13	21	13	39	793	819	26
23-Feb	**54**	13	9	13	18	13	35	789	815	26
24-Feb	**55**	13	7	13	16	13	32	787	812	25
25-Feb	**56**	13	5	13	13	13	29	785	809	24
26-Feb	**57**	13	2	13	10	13	26	782	806	24
27-Feb	**58**	13	0	13	7	13	22	780	802	22
28-Feb	**59**	12	58	13	5	13	19	778	799	21
29-Feb	**60**	12	57	13	3	13	17	777	797	20

1-Mar	**61**	12	55	13	0	13	15	775	795	20
2-Mar	**62**	12	53	12	56	13	12	773	792	19
3-Mar	**63**	12	50	12	53	13	8	770	788	18
4-Mar	**64**	12	46	12	51	13	5	766	785	19
5-Mar	**65**	12	45	12	48	13	2	765	782	17
6-Mar	**66**	12	43	12	46	12	59	763	779	16
7-Mar	**67**	12	41	12	43	12	55	761	775	14
8-Mar	**68**	12	38	12	41	12	52	758	772	14
9-Mar	**69**	12	36	12	38	12	49	756	769	13
10-Mar	**70**	12	33	12	34	12	45	753	765	12
11-Mar	**71**	12	32	12	32	12	42	752	762	10
12-Mar	**72**	12	29	12	29	12	38	749	758	9
13-Mar	**73**	12	27	12	26	12	35	747	755	8
14-Mar	**74**	12	24	12	24	12	32	744	752	8
15-Mar	**75**	12	22	12	21	12	28	742	748	6
16-Mar	**76**	12	20	12	19	12	25	740	745	5
17-Mar	**77**	12	17	12	16	12	21	737	741	4
18-Mar	**78**	12	15	12	13	12	18	735	738	3
19-Mar	**79**	12	12	12	11	12	15	732	735	3
20-Mar	**80**	12	10	12	8	12	12	730	732	2
21-Mar	**81**	12	7	12	5	12	9	727	729	2
22-Mar	**82**	12	5	12	3	12	5	725	725	0
23-Mar	**83**	12	2	12	0	12	2	722	722	0
24-Mar	**84**	12	1	11	57	11	59	721	719	2
25-Mar	**85**	11	58	11	55	11	55	718	715	3
26-Mar	**86**	11	56	11	51	11	52	716	712	4
27-Mar	**87**	11	53	11	48	11	49	713	709	4

28-Mar	**88**	11	51	11	46	11	45	711	705	6
29-Mar	**89**	11	48	11	43	11	42	708	702	6
30-Mar	**90**	11	46	11	41	11	39	706	699	7
31-Mar	**91**	11	43	11	38	11	35	703	695	8
1-Apr	**92**	11	42	11	35	11	32	702	692	10
2-Apr	**93**	11	40	11	39	11	30	700	690	10
3-Apr	**94**	11	37	11	37	11	27	697	687	10
4-Apr	**95**	11	35	11	34	11	23	695	683	12
5-Apr	**96**	11	32	11	31	11	20	692	680	12

Table 2.2 List of individuals, *Clk*polyQcds genotypes and mean departure dates (MDD) collected by direct observation over the years 2004–16 from the over-wintering sites (n = 135) and by recovered geolocators' from the stop-over in Asia (n = 32) in the years 2008, 2009, 2013 and 2014. Sex was determined molecularly. Godwits were ordered from the earliest to the latest departers based on their mean departure date from N.Z. *Abbreviations*: Bird ID, in which wintering site is indicated by the first letter (i.e. C= Catlins Coast; F= Firth of Thames; M= Manawatu Estuary); *Clk*polyQ-A1 = Allele 1; *Clk*polyQ-A2 = Allele 2. Note that day 1 = 1st January.

Bird ID	Sex	*Clk*polyQ-A1	*Clk*polyQ-A2	Mean departure date from N.Z. (2004–2016)	Mean departure date from Asia (2008–14)
CY12616	female	Q7	Q9	60	
FY13144	male	Q11	Q11	61	
FY13240	female	Q10	Q11	62	
CY12618	female	Q11	Q11	63	
CY12617	female	Q9	Q11	65	
CY12622	male	Q9	Q10	65	
CY12619	female	Q9	Q11	66	
CY12386	female	Q9	Q10	66	
CY12629	male	Q7	Q9	66	
MY13419	female	Q9	Q9	67	123
CY12626	female	Q9	Q12	68	
CY12624	male	Q9	Q9	68	
MY8861	male	Q9	Q10	68	
FY13172	female	Q8	Q11	68	
CY12627	female	Q9	Q11	69	
FY13239	male	Q7	Q11	69	
MY13313	female	Q8	Q10	70	118
MY13476	female	Q9	Q11	70	119
MY12485	female	Q9	Q10	70	
CY12620	female	Q9	Q9	71	
MY12725	female	Q7	Q8	71	119
MY13599	female	Q9	Q10	71	
MY13422	female	Q11	Q11	71	121
MY13311	male	Q7	Q9	71	
MY13331	male	Q9	Q10	71	
FY13142	male	Q9	Q9	71	
FY13238	male	Q7	Q9	71	
FY13297	male	Q10	Q11	71	
MY13585	male	Q9	Q10	71	130

MY13586	male	Q10	Q11	71	
FY13244	female	Q9	Q9	72	
FY13277	female	Q7	Q11	72	
MY13393	female	Q7	Q10	72	
MY13531	female	Q8	Q11	72	130
MY12723	male	Q9	Q9	72	
MY13583	male	Q9	Q12	72	
CY12631	male	Q10	Q11	73	
MY13587	female	Q9	Q11	73	136
MY13300	female	Q8	Q9	73	125
MY8807	female	Q9	Q9	73	123
FY13179	female	Q9	Q9	73	
FY13208	female	Q10	Q11	73	
MY13425	female	Q8	Q10	73	123
MY13581	female	Q9	Q9	73	124
MY12705	female	Q9	Q9	73	
MY13578	male	Q9	Q10	73	
FY13180	male	Q9	Q10	73	
MY13429	male	Q8	Q10	73	123
FY13175	female	Q9	Q10	74	
FY13237	female	Q9	Q11	74	
FY13335	female	Q11	Q11	75	
MY13299	female	Q9	Q11	75	
MY13314	female	Q8	Q12	75	
MY13588	male	Q9	Q9	75	134
MY13323	male	Q9	Q9	75	
MY13328	male	Q9	Q9	75	122
CY12628	male	Q8	Q9	76	
FY13104	female	Q10	Q11	76	
FY13173	female	Q9	Q11	76	
FY13228	male	Q7	Q7	76	
MY13373	female	Q9	Q10	76	
MY13325	female	Q9	Q12	76	121
FY13183	female	Q9	Q10	77	
FY13265	female	Q7	Q7	77	
FY13176	male	Q7	Q8	77	
FY13167	male	Q11	Q11	77	
MY13384	female	Q9	Q11	77	
MY12726	female	Q9	Q10	77	
MY13389	male	Q9	Q10	77	
MY8851	female	Q7	Q9	78	

FY13165	male	Q7	Q10	79	
FY13280	female	Q9	Q11	79	
FY13294	female	Q10	Q11	79	
FY13343	male	Q9	Q10	79	
MY13592	female	Q11	Q11	79	137
MY12708	female	Q10	Q12	79	
MY13394	female	Q10	Q12	79	
MY13329	female	Q9	Q11	79	
MY13424	male	Q9	Q10	79	124
MY13460	male	Q9	Q11	79	134
MY13506	male	Q9	Q10	79	
FY13177	female	Q9	Q10	80	
MY13598	female	Q10	Q11	80	
MY13426	female	Q9	Q9	80	119
MY13596	female	Q7	Q10	80	
MY13600	male	Q8	Q9	80	139
MY12706	male	Q10	Q11	80	
MY8855	male	Q7	Q11	80	136
FY13274	female	Q10	Q11	81	
FY13341	female	Q8	Q11	81	
FY13163	female	Q9	Q11	81	
FY13110	male	Q9	Q11	81	
FY13195	male	Q9	Q12	81	
FY13227	male	Q10	Q11	81	
MY12710	female	Q9	Q9	81	137
MY13383	male	Q9	Q11	81	
FY13136	male	Q10	Q11	82	
FY13181	male	Q9	Q11	82	
FY13229	male	Q7	Q9	82	
MY13576	female	Q9	Q10	82	132
MY13301	female	Q10	Q10	82	
MY13302	male	Q10	Q11	82	
MY8854	male	Q9	Q11	82	
MY13589	male	Q11	Q11	82	148
FY13125	male	Q9	Q11	83	
FY13171	female	Q8	Q9	83	
FY13336	female	Q10	Q10	83	
FY13296	male	Q8	Q9	83	
FY13170	male	Q10	Q12	83	
FY13295	male	Q8	Q9	83	
FY13112	male	Q8	Q10	83	

MY13430	male	Q9	Q11	83	133
MY13375	male	Q8	Q11	83	
MY13370	male	Q9	Q10	83	133
FY13102	female	Q9	Q10	84	
FY13278	male	Q9	Q10	84	
FY13346	male	Q9	Q10	84	
MY12704	female	Q8	Q10	84	
FY13334	female	Q10	Q11	85	
FY13287	female	Q9	Q11	85	
FY13231	male	Q9	Q11	85	
FY13137	male	Q10	Q10	85	
MY13391	male	Q8	Q9	85	142
MY12724	male	Q10	Q11	85	156
FY13249	female	Q11	Q12	86	
FY13345	female	Q9	Q11	86	
MY13421	female	Q9	Q11	86	124
MY13381	male	Q9	Q9	86	133
MY13388	male	Q7	Q11	86	
MY13428	male	Q9	Q11	86	140
MY8814	male	Q10	Q11	86	
MY13333	female	Q8	Q11	88	
MY13318	male	Q10	Q10	88	
FY13245	female	Q9	Q10	89	
FY13279	male	Q9	Q10	91	

Figure 2.1 Comparison of the departure time for migration between sexes in the (a) N.Z. (Females = 70; males = 65), and (b) Asian dataset (Females = 17; males = 15). Filled dots and triangles (individuals) were used for the N.Z. and the Asian datasets respectively. * denotes statistically significant ($P < 0.05$), ** ($P < 0.01$). Note that the N.Z. dataset is a subset of a larger dataset with birds' departure dates recorded but without *Clock* genotype information. There is no difference in N.Z. departure time between the sexes in the full dataset (results not shown).

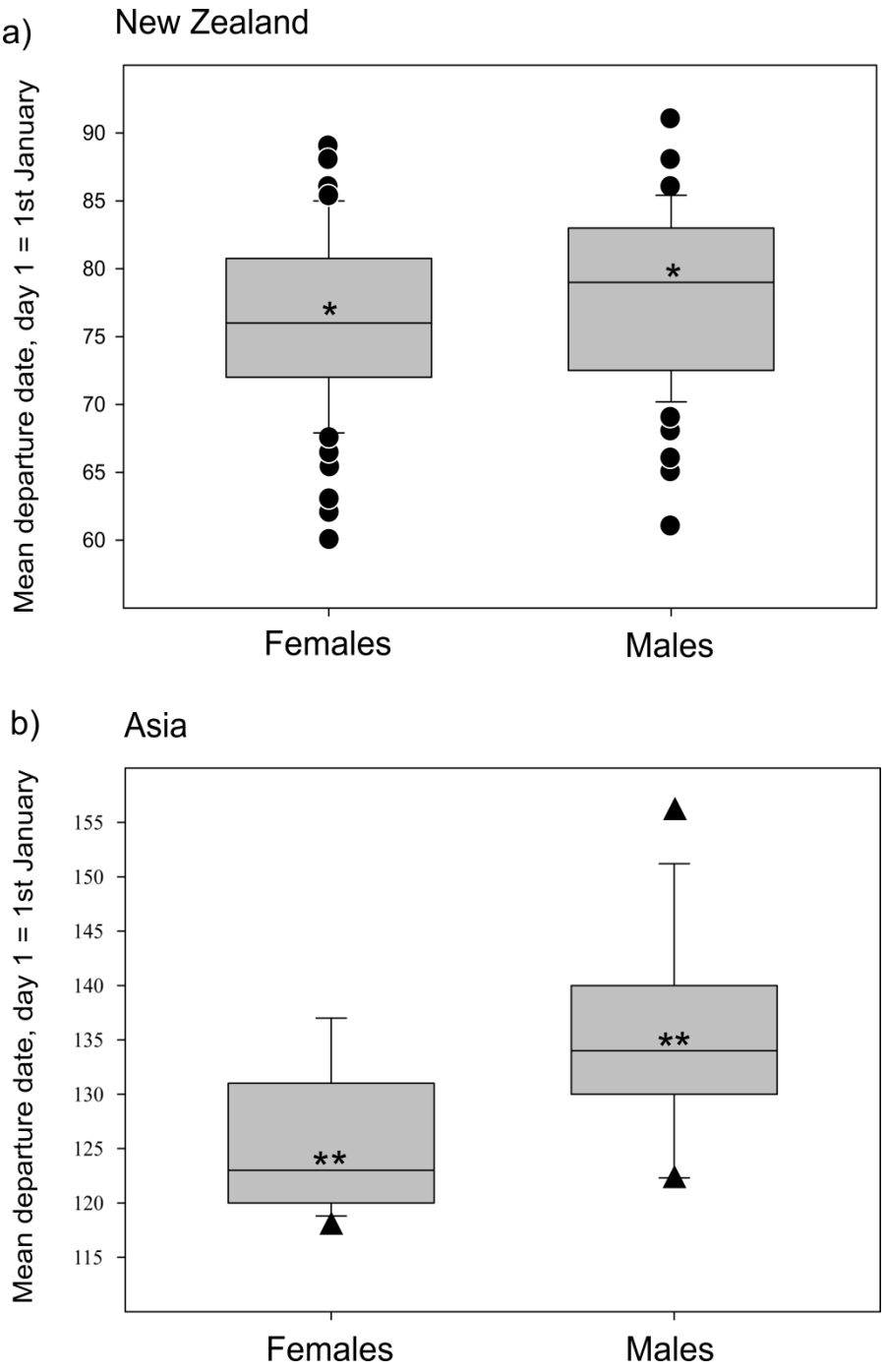


Figure 2.2 Confirmation that labelling identical amplicons with the alternative florescent dyes VIC® and NED™ did not affect corresponding estimated amplicon sizes. (a, b) Bird ID = MY13302; (c, d) = MY13394; (e, f) MY13391. Electropherograms show that the two peaks corresponding to the two *Clk*polyQcds alleles found in heterozygous individuals had the same size whether labelled with VIC® (green, A, C, E) or NED™ (black, B, D, F). The five blue peaks shown correspond to the FAM™ labelled amplicons generated from five cloned *Clk*polyQcds alleles while *GeneScan*™-500 LIZ™ size standards are shown in orange.

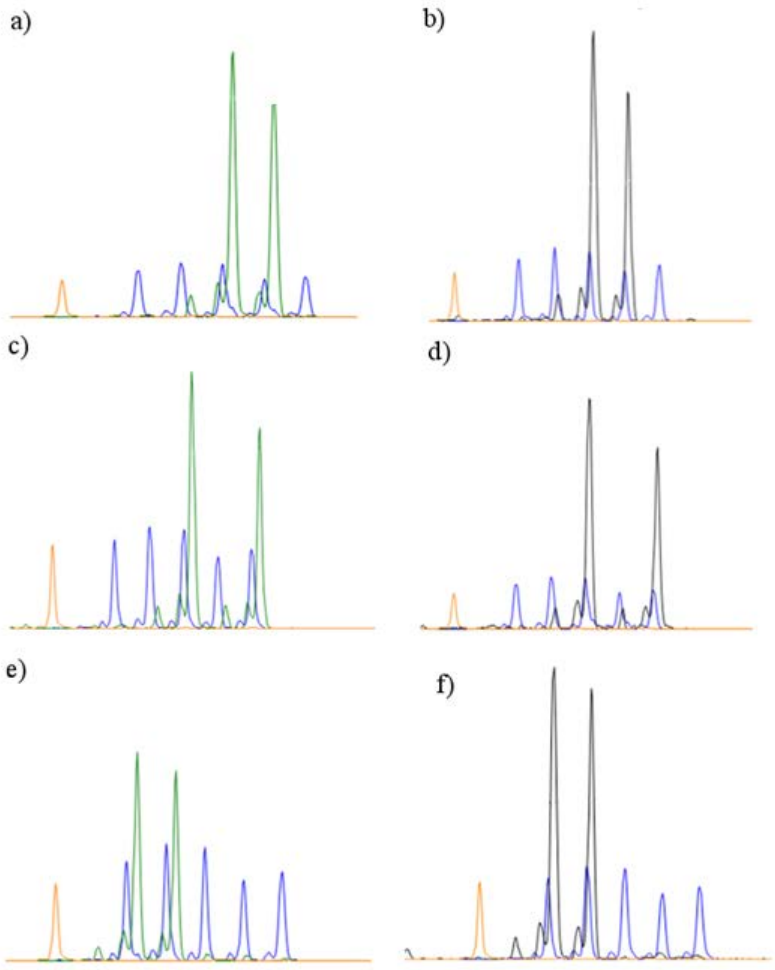


Figure 2.3 Comparisons between F_{ST} values calculated using godwit microsatellite and *Clk*polyQcds allele frequencies in *a priori* 'earlier' and 'later' departing groups. Using migration departure dates, godwits over-wintering in N.Z. or stopping-over in Asia were grouped on discrete 'earlier' and 'later' departers' groupings as described in detail in the text. F_{ST} values were calculated between the departure time groups using allele frequencies at 27 microsatellite loci or the *Clk*polyQcds locus. Note that using 30 microsatellites does not change this graphic significantly. * Statistically significant ($P < 0.05$) F_{ST} value based on microsatellite loci; ‡ indicates marginally significant for *Clk*polyQcds ($P < 0.08$). Numbers of individuals per *a priori* groups are indicated in parentheses.

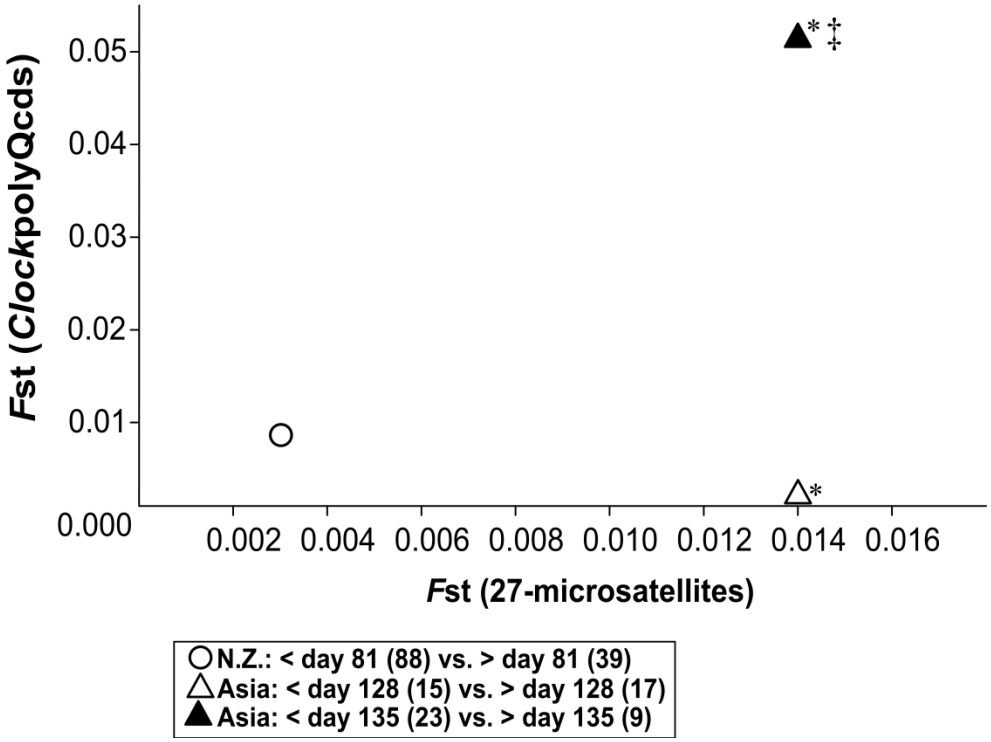


Table 2.3 Summary of previous studies on the *Clk*polyQcds polymorphism and diverse phenology-related traits in birds. Studies are listed by year of publication. Observed heterozygosity (Ho) is indicated when the publication reported it. * indicates number of alleles when alleles were not reported (genotyped).

Reference	Species	Range of polymorphism	Ho	Reported findings
(Johnsen *et al.* 2007)	*Luscinia svecica*	Q10–Q16	0.21	Longer alleles more common in higher latitude at the breeding site (only in *Cyanister caeruleus*)
	Cyanister caeruleus	Q9–Q17	0.49	
(Liedvogel *et al.* 2009)	*Cyanistes caeruleus*	Q10–Q16 (Q15 not detected)	0.56	Breeding time weakly associated with *Clk*polyQcds (only in females). Larger polyQ-alleles found in females with more breeding success.
(Liedvogel & Sheldon 2010)	*Parus major*	Q10–Q16 (Q14 accounted for 96% of the overall allelic diversity)	0.08	No evidence for association of *Clk*polyQcds with breeding time
(Dor *et al.* 2011b)	*Tachycineta bicolour*	Q6–Q9	0.47	No evidence for association of *Clk*polyQcds with latitude, clutch initiation or incubation duration
	T. thalassina	Q6–Q9	0.35	
	T. albilinea	Q7–Q9	0.34	
	T. leucorrhoa	Q7–Q9	0.05	
	T. meyeni	Q6–Q8	0.44	

(Dor *et al.* 2011a)	*Hirundo rustica*	Q6–Q8	0.03	Very low polymorphism (>96% homozygotes). Conclusion: High diversity at *Clk*polyQ is not general across all avian species.
(Caprioli *et al.* 2012)	*Hirundo rustica*	Q5–Q8 (Q7 accounted for 96.7% of the overall allelic diversity)	0.07	Evidences of association between breeding date and two rare *Clk*polyQ genotypes (Q7/Q8 vs. Q7Q7).
(Kuhn *et al.* 2013)	*Ficedula hypoleuca*	Q10–Q14	0.50	No significant genetic variation over time.
(Chakarov *et al.* 2013)	*Buteo buteo*	Q8	-	*Clk*polyQcds is monomorphic in a large clade of birds of prey. Conclusion: *Clk*polyQcds polymorphism seems not to be necessary for adaptive variation in phenology.
	Permis apivorus	Q9	-	
	Accipiter gentilis	Q9–Q11 (Q10 not detected)	-	
	Milvus milvus	Q6–Q8 (Q7 not detected)	-	
	Milvus migrans	Q8	-	
	Falco tinnunculus	Q7	-	
	Falco peregrinus	Q7	-	
	Falco cherrug	Q7	-	
(Saino *et al.* 2013)	*Hirundo rustica*	Q6–Q8	-	Q7/Q8 barn swallows at the *Clk*polyQcds delayed moult compared to the other individuals with different genotypes. However, sample size of this study is quite small.

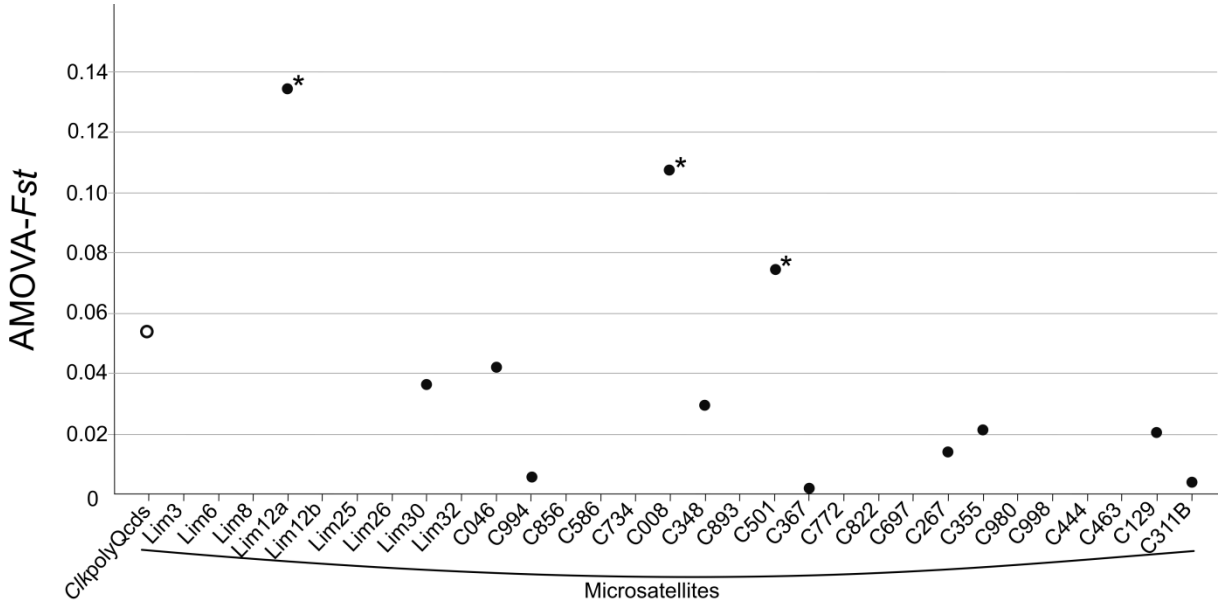
(Peterson *et al.* 2015)	*Junco hyemalis hyemalis*	4*	0.27	No evidence for a predictable relationship between migratory behaviour (sedentary vs. migratory) and lengths of *Clk*polyQcds
	J. h. carolinensis	3	0.36	
	J. j. aikeni	3	0.49	
	J. h. oreganus	3	0.19	
	J. junco thurberi	4	0.31	
	J. h. caniceps	3	0.14	
	J. h. meamsi	4	0.32	
	J. h. insularis	1	0.00	
	J. phaeonotus phaeonotus	4	0.37	
	J. p. alticola	3	0.56	
	J. p. bairdi	4	0.16	
(Bourret & Garant 2015)	*Tachycineta bicolour*	Q6–Q9	0.51	Association between laying date and *Clk*polyQcds only in females. Association between incubation duration and *Clk*polyQcds only in females.

(Saino *et al.* 2015)	*Luscinia megarhynchos* *Ficedula hypoleuca* *Anthus trivialis* *Saxicola rubetra*	Q9–Q13 Q10–Q15 (Q14 not detected) Q6–Q10 Q9–Q16 (Q10 not detected)	0.55 0.48 0.25 0.13	Longer alleles are more common in later migratory birds in *Anthus trivialis* using the *Clk*polyQcds mean allele length. Longer alleles are more common in later migratory females in *Anthus trivialis* using the *Clk*polyQcds longer allele length. Negative and significant relationship between mean *Clk*polyQcds allele length and wing length.
(Bazzi *et al.* 2015)	*Hirundo rustica*	Q6–Q8	Low (not specified)	Two males with genotype Q6Q7 departed earlier compared to those with genotype Q7Q7. A single female Q7Q8 departed later compared to those genotyped Q7Q7. Three Q6Q7 individuals had similar phenology between them compared to individuals with genotype Q7Q7.
(Bazzi *et al.* 2016b)	*Cardellina pusilla*	-	0.02	98% of the individuals were homozygous at *Clock* (almost no variability at the *Clock* gene). Rare heterozygotes did not deviate from homozygous migration phenology.
(Bazzi *et al.* 2016c)	*Phylloscopus trochilus*	-	-	*Clock* did not predict time of spring migration or moult speed. They obtained significant predictability of *Npas2* allele size and spring migration date only in females and of *Creb1* allele size and moult speed only in males.

(Bazzi *et al.* 2016a)	*Acrocephalus arundinaceus*	-	-	Note: This is a between-species study (species level). Species breeding at northern latitudes had significantly longer *Clk*polyQcds alleles. There was no association between migratory date and *Clock* allele size. *Clock* gene diversity (see Nei & Roychoudhury 1974, formula 12.33 in Nei & Kumar 2000) was found to be related to migratory behaviour: Specifically long-distance migratory species were found to be associated with lower Clock diversity.
	Acrocephalus schoeobaenus	-	-	
	Acrocephalus scirpaceus	-	-	
	Anthus trivialis	-	-	
	Caprimulgus europaeus	-	-	
	Ficedula hypoleuca	-	-	
	Hippolais icterina	-	-	
	Jynx torquilla	-	-	
	Lanius senator	-	-	
	Luscinia megarhynchos	-	-	
	Merops apiaster	-	-	
	Muscicapa striata	-	-	
	Oeanthe oeanthe	-	-	
	Oriolus oriolus	-	-	
		-	-	

(Bazzi *et al.* 2016a)	*Oriolus oriolus*	-	-	
	Phoenicurus phoenicurus	-	-	
	Phylloscopus sibilatrix	-	-	
	Phylloscopus trochilus	-	-	
	Saxicola rubetra	-	-	
	Streptopelia turtur	-	-	
	Sylvia borin	-	-	
	Sylvia cantillans	-	-	
	Sylvia communis	-	-	
	Upupa epops	-	-	
(Zhang *et al.* 2017)	*Calandrella cheleensis*	Q7–Q13 (Q8 not detected)	0.58	Mean ClockpolyQ allele length is positively correlated with the standardized laying date of the first egg.
(Contina *et al.* 2018)	*Passerina ciris*	8	0.60	No association between *Clk*polyQcds polymorphism and fall departure date, migration duration, moult site arrival date, spring arrival date, body mass and wing chord.
Present study	*Limosa lapponica baueri*	Q7–Q12	0.79	No association between *Clk*polyQcds polymorphism and inter-individual migration departure time from the over-wintering sites (N.Z.) or from the stop-over (Eastern Asia).

Figure 2.4 Graph showing that the marginally significant AMOVA-*Fst* in *Clk*polyQ at day 135 divide could be a reflection of underlying weak genetic structure. AMOVA-*Fst* values of *Clk*polyQcds (open dot, *Fst* = 0.054, *P* = 0.07) and of each microsatellite (filled dots) were calculated using Arlequin v.3.5. ** *P* < 0.01; * *P* < 0.05. Note that three microsatellites *Fst* values surpass that of *Clk*polyQcds.



Appendix 3

Table 3.1 Godwits' chronophenotypes (i.e. raw phenotypes and mean chronophenotypes) and individuals' departure locations. Godwits are listed in order from the earliest to the latest departers per site. Note that day 1 = 1 January. *Abbreviations*: indv ID= individuals' ID, which denotes the departure site as the first letter (C = Catlins Coast; F = Firth of Thames; M = Manawatu Estuary) followed by a unique metal band number starting with Y, e.g. Y12699; mdd= mean departure dates or mean chronophenotypes; mdd2= standardized by year per site mean departure dates or standardized mean chronophenotypes. * indicates the individuals that carried geolocators and from which departure dates from Asia was determined.

sample ID	indiv ID	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	mdd	mdd2	Asia_ mdd
NWZ_106801_P 02_WE10	CY12699													57		57	56	
NWZ_106801_P 02_WH01	CY12654												57			57	56	
NWZ_106801_P 02_WA09	CY12674												59	58		59	57	
NWZ_106801_P 03_WH09	CY12630												59			59	58	
NWZ_106801_P 02_WC12	CY12824													60		60	59	
NWZ_106801_P 02_WF05	CY12637												63	57		60	59	
NWZ_106801_P 03_WD06	CY12822													60		60	59	
NWZ_106801_P 02_WD08	CY12802												63	59		61	60	
NWZ_106801_P 02_WB11	CY12693													62		62	61	
NWZ_106801_P 02_WG12	CY12645											65	63	58		62	61	
NWZ_106801_P 02_WD12	CY12643											63	67	59		63	62	
NWZ_106801_P 02_WE09	CY12659												65	61		63	62	

NWZ_106801_P03_WG09	CY12618										65	62	62			63	62	
NWZ_106801_P02_WD11	CY12681												64	63		64	63	
NWZ_106801_P02_WE06	CY12679												64	63		64	63	
NWZ_106801_P02_WH11	CY12810												65	62		64	63	
NWZ_106801_P02_WA10	CY12635											64		64		64	64	
NWZ_106801_P02_WB09	CY12672												64			64	64	
NWZ_106801_P02_WG11	CY12811												64	64		64	64	
NWZ_106801_P03_WB10	CY12646												64			64	64	
NWZ_106801_P03_WG07	CY12636											64				64	64	
NWZ_106801_P01_WB03	CY12616										65	64				65	64	
NWZ_106801_P02_WC08	CY12803												65	64		65	64	
NWZ_106801_P02_WB05	CY12692													65		65	65	
NWZ_106801_P02_WC09	CY12671												65			65	65	
NWZ_106801_P02_WG09	CY12648											64		66		65	65	
NWZ_106801_P02_WH12	CY12638											69	62	64		65	65	
NWZ_106801_P03_WA01	CY12668												65	65		65	65	
NWZ_106801_P03_WA10	CY12617										66	64	65			65	65	
NWZ_106801_P03_WF06	CY12622										64	64	67			65	65	
NWZ_106801_P03_WG04	CY12639											67	64	64		65	65	
NWZ_106801_P03_WH04	CY12825													65		65	65	
NWZ_106801_P01_WH10	CY12619										66	64	63	69		66	65	
NWZ_106801_P02_WE08	CY12812												69	62		66	65	

NWZ_106801_P02_WH07	CY12801												67	64		66	65	
NWZ_106801_P02_WE12	CY12661												65	67		66	66	
NWZ_106801_P02_WF07	CY12807													66		66	66	
NWZ_106801_P02_WG10	CY12696													66		66	66	
NWZ_106801_P02_WH04	CY12821													66		66	66	
NWZ_106801_P03_WA09	CY12386										69	63				66	66	
NWZ_106801_P01_WA11	CY12629										66	65	67	67		66	66	
NWZ_106801_P03_WB06	CY12666											67	63	69		66	66	
NWZ_106801_P02_WG07	CY12806												69	64		67	66	
NWZ_106801_P02_WA08	CY12805												65	69		67	67	
NWZ_106801_P02_WB06	CY12656											64	67	70		67	67	
NWZ_106801_P02_WD06	CY12653											67				67	67	
NWZ_106801_P02_WD10	CY12819													67		67	67	
NWZ_106801_P03_WC01	CY12697													67		67	67	
NWZ_106801_P01_WB06	CY12652											67	64	71		67	67	
NWZ_106801_P02_WD09	CY12662											64	69	69		67	67	
NWZ_106801_P03_WD09	CY12624										69	64	68	70		68	68	
NWZ_106801_P03_WF05	CY12626										69	64	69	69		68	68	
NWZ_106801_P02_WB10	CY12633											71	65	69		68	68	
NWZ_106801_P03_WE06	CY12627										66	73	67	69		69	69	
NWZ_106801_P01_WD06	CY12688													69		69	69	
NWZ_106801_P02_WA05	CY12820													69		69	69	

NWZ_106801_P02_WA11	CY12694												69		69	69		
NWZ_106801_P02_WA12	CY12809											69	69		69	69		
NWZ_106801_P02_WC11	CY12683											69	69		69	69		
NWZ_106801_P02_WE11	CY12691												69		69	69		
NWZ_106801_P02_WF10	CY12817												69		69	69		
NWZ_106801_P02_WH10	CY12695												69		69	69		
NWZ_106801_P03_WA07	CY12698												69		69	69		
NWZ_106801_P03_WD04	CY12700												69		69	69		
NWZ_106801_P03_WH05	CY12823												69		69	69		
NWZ_106801_P03_WE04	CY12663											71	73	65		70	70	
NWZ_106801_P02_WG05	CY12649											69	73	69		70	71	
NWZ_106801_P02_WG08	CY12676												74	67		71	71	
NWZ_106801_P02_WH09	CY12644											73	73	66		71	71	
NWZ_106801_P02_WH05	CY12634											71	73	69		71	71	
NWZ_106801_P02_WH08	CY12675												73	69		71	71	
NWZ_106801_P03_WC03	CY12682												71			71	71	
NWZ_106801_P03_WD07	CY12620										71					71	71	
NWZ_106801_P03_WH03	CY12655											71	73	69		71	71	
NWZ_106801_P01_WC06	CY12814												74	69		72	72	
NWZ_106801_P03_WB04	CY12631										72	73				73	73	
NWZ_106801_P02_WA06	CY12664											73				73	74	
NWZ_106801_P02_WF08	CY12673												73			73	74	

NWZ_106801_P03_WG03	CY12647											73				73	74	
NWZ_106801_P02_WC06	CY12660												74			74	75	
NWZ_106801_P02_WC05	CY12813												74	75		75	75	
NWZ_106801_P02_WF09	CY12665											77	73	75		75	76	
NWZ_106801_P02_WF11	CY12689													75		75	76	
NWZ_106801_P03_WB07	CY12628										76	76	75	76		76	77	
NWZ_106801_P02_WB08	CY12804												73	79		76	77	
NWZ_106801_P02_WE07	CY12808												78			78	79	
NWZ_106801_P02_WF12	CY12621										86					86	88	
NWZ_106801_P03_WG06	CY12826													69		69		
NWZ_106801_P03_WC10	FY13240		61		62											62	60	
NWZ_106801_P01_WG04	BVV											72		64		68	67	
NWZ_106801_P01_WH05	FY12757												67	69		68	67	
NWZ_106801_P01_WH12	FY6602												68			68	67	
NWZ_106801_P02_WA03	FY13172		69	67												68	67	
NWZ_106801_P01_WA06	FY12760												70	69		70	69	
NWZ_106801_P01_WG05	FY12755												70	69		70	69	
NWZ_106801_P01_WE05	FY12794												72	70		71	70	
NWZ_106801_P01_WF04	BUW											73	71	69		71	70	
NWZ_106801_P01_WG01	FY13297			74	70		73							68		71	71	
NWZ_106801_P01_WD05	FY12784												75	69		72	71	
NWZ_106801_P02_WE05	FY13277			72												72	71	

NWZ_106801_P 03_WE08	FY13244		74	70												72	71	
NWZ_106801_P 01_WF09	FY13179		73													73	72	
NWZ_106801_P 02_WC03	FY13180			73												73	72	
NWZ_106801_P 03_WF08	FY13208			73												73	72	
NWZ_106801_P 01_WD12	BUV											77	73	71		74	73	
NWZ_106801_P 01_WG09	FY13237			74												74	74	
NWZ_106801_P 03_WA05	FY13175			74												74	74	
NWZ_106801_P 02_WF04	BVB											78	76	69		74	74	
NWZ_106801_P 01_WE12	BUF											75				75	75	
NWZ_106801_P 01_WF05	FY18601											75	76	77		76	76	
NWZ_106801_P 02_WB03	FY13173			76												76	76	
NWZ_106801_P 02_WC01	FY12788												75	77		76	76	
NWZ_106801_P 02_WE02	FY13104		76	76												76	76	
NWZ_106801_P 01_WA05	FY12776											77	76			77	76	
NWZ_106801_P 03_WE07	FY13176		78	76												77	77	
NWZ_106801_P 03_WF04	FY13183			77												77	77	
NWZ_106801_P 02_WG02	FY13167		78	80			74									77	77	
NWZ_106801_P 01_WH04	BVX												78			78	78	
NWZ_106801_P 02_WB01	FY12774											78	78			78	78	
NWZ_106801_P 02_WG04	FY12754											79	77			78	78	
NWZ_106801_P 01_WB05	FY12780											80	78			79	79	
NWZ_106801_P 01_WD04	BUN											80	77	80		79	79	

NWZ_106801_P03_WD08	FY13343			79												79	79	
NWZ_106801_P03_WD10	FY13162		83	75												79	79	
NWZ_106801_P03_WF07	FY13294				79											79	79	
NWZ_106801_P01_WE04	BUU											84	78	76		79	80	
NWZ_106801_P01_WA10	FY13177			80												80	80	
NWZ_106801_P01_WG12	FY18609											79	82			81	81	
NWZ_106801_P02_WE04	BUM											85		76		81	81	
NWZ_106801_P01_WB01	FY13163		85	76	81											81	81	
NWZ_106801_P01_WC01	FY13227			81												81	81	
NWZ_106801_P02_WD05	FY13274				81		81									81	81	
NWZ_106801_P03_WC08	FY13110	81														81	81	
NWZ_106801_P03_WE03	FY13341			81												81	81	
NWZ_106801_P01_WC04	BTC											86	78	80		81	82	
NWZ_106801_P01_WC05	FY12782											84	79			82	82	
NWZ_106801_P01_WG06	FY13136			82												82	82	
NWZ_106801_P02_WG01	FY12785											85	79			82	82	
NWZ_106801_P03_WH07	FY13181		83		81											82	82	
NWZ_106801_P03_WF10	FY13296			84	81											83	83	
NWZ_106801_P01_WA02	FY13336			83												83	84	
NWZ_106801_P01_WB04	BTC											85	81			83	84	
NWZ_106801_P01_WC10	FY13125	81	85													83	84	
NWZ_106801_P01_WD10	FY13170		83													83	84	

NWZ_106801_P02_WB04	FY13295			83												83	84	
NWZ_106801_P02_WH02	FY13171		85	81												83	84	
NWZ_106801_P01_WF12	BVZ											85	84	81		83	84	
NWZ_106801_P02_WF02	FY13112	84	87	79												83	84	
NWZ_106801_P01_WH01	FY13278			84												84	85	
NWZ_106801_P02_WD02	FY13102	83	85		79			89								84	85	
NWZ_106801_P03_WC04	FY13334			81			88									85	85	
NWZ_106801_P03_WH02	FY13287			84	82		88									85	85	
NWZ_106801_P03_WH08	FY13231		87	86	81											85	85	
NWZ_106801_P01_WH06	FY13137		83	90	82											85	86	
NWZ_106801_P02_WD03	FY13245		89				88									89	90	
NWZ_106801_P01_WA07	FY13279			91												91	93	
*NWZ_106801_P01_WF07	*MY13419					73	63	65								67	66	123
NWZ_106801_P02_WA07	MY12736												69		67	68	67	
NWZ_106801_P03_WE10	MY8861										68	68	69		67	68	67	
NWZ_106801_P01_WH09	MY11438													69	68	69	67	
NWZ_106801_P02_WH06	MY18638												70	69	68	69	68	
NWZ_106801_P03_WB01	MY18617												69	69		69	68	
*NWZ_106801_P01_WH11	*MY13313					71	71	71	68	69	68					70	69	118
NWZ_106801_P01_WE11	MY13471											73	70	69	67	70	69	
*NWZ_106801_P03_WD02	*MY12725											73	70	69	67	70	69	119
*NWZ_106801_P01_WB02	*MY13476											71	69			70	69	119

NWZ_106801_P02_WC10	MY12485					73	71	65	71	73	70	71	70	69	68	70	69	
NWZ_106801_P01_WE02	MY13386					73	72	72	68	69	68	71	70	69	74	71	70	
NWZ_106801_P01_WG08	MY13331					70	71	71								71	70	
NWZ_106801_P01_WE10	MY11440													69	73	71	70	
NWZ_106801_P02_WB07	MY13599										71	73	70	69	73	71	70	
NWZ_106801_P01_WF10	MY12716											73	70	69	73	71	71	
*NWZ_106801_P01_WC03	*MY13585										71	73	70			71	71	130
*NWZ_106801_P01_WH07	*MY13422					81	71	69	69	68	70					71	71	121
NWZ_106801_P01_WB10	MY12723											73	70			72	71	
*NWZ_106801_P01_WG03	*MY13587										72	73	74	71	68	72	71	136
NWZ_106801_P01_WD01	MY18639												73	69	73	72	71	
NWZ_106801_P02_WA02	MY13393					73	72	71	71	73	71					72	71	
NWZ_106801_P01_WD02	MY12743											73	73	69	73	72	71	
NWZ_106801_P01_WF11	MY12721											73	73	69	73	72	71	
NWZ_106801_P02_WC04	MY18635												73	71		72	71	
NWZ_106801_P03_WB02	MY8809												73	69	74	72	71	
NWZ_106801_P03_WC06	MY13583										72	73	73	69	73	72	71	
NWZ_106801_P02_WD01	MY12718											73	74	69	73	72	72	
*NWZ_106801_P01_WE06	*MY13531											73	73	71	73	73	72	130
NWZ_106801_P01_WC12	MY13578								71	73	72	78	73	69	73	73	72	
*NWZ_106801_P03_WG08	*MY8807										75	73	73	71		73	72	123
*NWZ_106801_P01_WA08	*MY13429					81	72	76	71	69	70					73	73	123

*NWZ_106801_P01_WA04	*MY13425					73	74	76	71	76	72	73	74	69	74	73	73	123
NWZ_106801_P03_WC07	MY12705										72	78	70			73	73	
NWZ_106801_P02_WH03	MY8802										75	78		69		74	74	
NWZ_106801_P02_WF03	MY12711										75	73	73			74	74	
NWZ_106801_P01_WH03	MY13582										71	79	73			74	74	
NWZ_106801_P01_WC08	MY13299					78	80	76	75		72	73	73	69	74	74	74	
NWZ_106801_P02_WF01	MY13314					81	72	76	78	79	71	73	74	69	73	75	74	
NWZ_106801_P01_WB11	MY12717											73	73	78		75	74	
NWZ_106801_P01_WG11	MY13323					76	72	76								75	74	
NWZ_106801_P01_WG02	MY13465											78	74	69	78	75	74	
NWZ_106801_P03_WD01	MY13544							76	78	73	75	78	74	71	73	75	74	
*NWZ_106801_P03_WB09	*MY13328					78	74	71	72	79						75	74	122
*NWZ_106801_P03_WB03	*MY13588										79	78	70	71	78	75	75	134
NWZ_106801_P03_WF01	MY13573							76	78	79	72	78	74	71	74	75	75	
NWZ_106801_P02_WE03	MY12707											78	74	72	78	76	75	
NWZ_106801_P01_WB09	MY13373					78	74									76	76	
*NWZ_106801_P01_WC02	*MY13325					78	74	76	78	77	79	78	74	72	74	76	76	121
NWZ_106801_P01_WH02	MY18621												76			76	76	
NWZ_106801_P03_WD03	MY13584										82	71				77	76	
NWZ_106801_P03_WG01	MY13515											73	76	79	78	77	76	
NWZ_106801_P01_WF06	MY13591												74	78	78	77	77	
NWZ_106801_P01_WA09	MY13389					83	74	76	75							77	77	

NWZ_106801_P02_WE01	MY12727											78	76			77	77	
NWZ_106801_P01_WD09	MY13384					81	74	76	78							77	77	
NWZ_106801_P01_WG10	MY12722											79	76			78	77	
NWZ_106801_P02_WA01	MY12726											78	76	78	78	78	77	
NWZ_106801_P02_WA04	MY8851										79	78	76	78	78	78	78	
NWZ_106801_P01_WF02	MY12738											80	76	78		78	78	
NWZ_106801_P03_WE05	MY12720											78				78	78	
*NWZ_106801_P01_WF03	*MY13424					81	81	76	78	80	79	78	76			79	79	124
NWZ_106801_P02_WC07	MY13506										79					79	79	
NWZ_106801_P03_WA04	MY13590											79				79	79	
NWZ_106801_P03_WE09	MY12708										79					79	79	
NWZ_106801_P01_WB08	MY13394					81	81	76	79	79	79					79	79	
*NWZ_106801_P02_WC02	*MY8803											81	78	80	78	79	79	136
NWZ_106801_P03_WD05	MY13329					81	82	76	79	80	79	79				79	80	
NWZ_106801_P02_WD04	MY12701										86	73				80	80	
NWZ_106801_P03_WG02	MY13598										80	79				80	80	
*NWZ_106801_P03_WH06	*MY13592										79		78	79	82	80	80	137
*NWZ_106801_P01_WE03	*MY13600										82	80	77			80	80	139
*NWZ_106801_P01_WB12	*MY13460										79	80	77	79	84	80	80	134
*NWZ_106801_P01_WG07	*MY13426					83	81	85	79	81	72	79	78	78	82	80	80	119
NWZ_106801_P02_WG03	MY12712										79	81	76	79	84	80	80	
NWZ_106801_P02_WF06	MY13505												76	79	85	80	80	

NWZ_106801_P03_WB05	MY13596										80					80	80	
NWZ_106801_P03_WC02	MY12719											80	76	79	85	80	80	
*NWZ_106801_P03_WA03	*MY8855										79	78	82	81	82	80	81	136
*NWZ_106801_P01_WC11	*MY12713											84	78	78	82	81	81	124
NWZ_106801_P03_WF09	MY12706										79	81	78	81	84	81	81	
*NWZ_106801_P01_WD11	*MY12710										83	79				81	81	137
*NWZ_106801_P01_WA01	*MY13576								83	88	82	80	78	79	84	82	82	132
NWZ_106801_P01_WE01	MY13301					82										82	82	
NWZ_106801_P02_WB02	MY13597											79	85			82	82	
NWZ_106801_P03_WC05	MY11427													82	82	82	82	
NWZ_106801_P03_WE02	MY13302					83	81	85	79							82	82	
*NWZ_106801_P02_WD07	*MY13589										85	80	82			82	83	148
NWZ_106801_P03_WF02	MY8854										85	82	82	79	85	83	83	
NWZ_106801_P02_WB12	MY12709											81	86	81		83	83	
*NWZ_106801_P01_WE07	*MY13430					87	89	76	79	88	80	81	85	80	82	83	83	133
NWZ_106801_P01_WH08	MY13375					83										83	84	
*NWZ_106801_P03_WG05	*MY13593											83				83	84	137
*NWZ_106801_P01_WC07	*MY13370					83	89	85	82	81	82	81	87	80	84	83	84	133
NWZ_106801_P02_WG06	MMY8688										82			85	85	84	85	
NWZ_106801_P03_WE01	MY12745												86	81	85	84	85	
NWZ_106801_P03_WA02	MY18629												86	82	85	84	85	
NWZ_106801_P03_WF03	MY12704										85	82	85	85	86	85	85	

*NWZ_106801_P03_WC09	*MY13391					85	89	85	80	88	82					85	86	142
*NWZ_106801_P01_WA12	*MY12724											85				85	86	156
NWZ_106801_P03_WH01	MY18624												85			85	86	
*NWZ_106801_P01_WE09	*MY13381					87	89	87	82		83					86	86	133
NWZ_106801_P01_WC09	MY13388					83	89									86	87	
*NWZ_106801_P01_WE08	*MY13428					87	90	87	82	88	82					86	87	140
NWZ_106801_P03_WB08	MY8814											86				86	87	
*NWZ_106801_P01_WD07	*MY13421					86	89	87	82	88						86	87	124
*NWZ_106801_P01_WA03	*MY12715											82	90	85	89	87	87	132
NWZ_106801_P01_WF01	MY13320					87	90	88	89	89	86	81				87	88	
NWZ_106801_P03_WA06	MY13318					87	90	87								88	89	
NWZ_106801_P03_WA08	MY13333					93	90	87	83							88	89	
NWZ_106801_P01_WB07	MY13300					73	72	74								73	-	-
NWZ_106801_P01_WF08	MY13383					81	81									81	-	-
NWZ_106801_P01_WD03	MY13581										75	73	74	71		73	-	-
NWZ_106801_P01_WD08	MY13391					85	89	85	80	88	82					85	-	-

Table 3.2 List of the candidate genes selected from bibliography and classified in 4 categories by their functional group: 1) Genes/ proteins of the central circadian oscillator (CCO); 2) genes described as targeting outputs of CCO elements; 3) components of the Hypothalamic-Pituitary-Gonadal, 4) fat/energy metabolism or 5) genes identified as perhaps associated with migration timing or biological rhythms. First part of the table shows the initial candidate genes list. Second part of the table lists the "new" candidate genes, which were added to the study based on having high *Fst* values in the preliminary genome-wide comparison of 9 extreme early birds versus 10 extreme late birds. 120 genes were used for the analyses (i.e. passing the genotyping process and the filtering steps) are indicated with a number in the first column. Underlined and with * (*per* gene) is indicating that this gene has any SNP in the 124 SNPs subset used for the population structure analyses. Candidate genes that were not found in the godwit genome are listed at the end of the table. Initial candidate genes are organized by relative importance (that is, personal consideration of its importance as candidate gene based on how much previous studies support a relation between protein function and regulation of circadian/circannual rhythms). "New" acquisitions are classified from the highest to lowest *Fst* value (i.e. indicating more skew between the 9 extreme 'early' and 10 extreme 'late). Scaffolds# are indicated for each candidate gene. Details of the references are at the end of this document. A description of candidate genes' function described for the human orthologue can be found at http://www.genecards.org/.

	Initial candidate genes				
Genes used	**Scaffold#**	**Protein name**	**Full name**	**Function group**	**References**
1	**scaffold03386**	CLOCK	Circadian Locomotor Output Cycles Kaput Protein	Genes / proteins on the CCO	(Yoshimura *et al.* 2000; Dawson *et al.* 2001; Fidler & Gwinner 2003; Sumova *et al.* 2003; Leder *et al.* 2006; Bozek *et al.* 2009; Albrecht 2012; Ueshima *et al.* 2012)
2	**scaffold00037**	BMAL1/ ARNTL/ MOP3	Aryl Hydrocarbon Receptor Nuclear Translocator-Like	Genes / proteins on the CCO	(Yoshimura *et al.* 2000; Dawson *et al.* 2001; Haque *et al.* 2003; Yasuo *et al.* 2003; Johnsen *et al.* 2007; Bozek *et al.* 2009; Cassone & Westneat 2012)
3	**scaffold06242**	BMAL2/ MOP9/ ARNTL2	Aryl Hydrocarbon Receptor Nuclear Translocator-Like 2	Genes / proteins on the CCO	(Cermakian *et al.* 2000; Hogenesch *et al.* 2000; Schoenhard *et al.* 2003)
4	**scaffold00050**	CRY1	Cryptochrome Circadian Clock 1	Genes / proteins on the CCO	(Fu *et al.* 2002; Chong *et al.* 2003; Yasuo *et al.* 2003; Feillet *et al.* 2006; Barrett & Hoekstra 2011; Cassone & Westneat 2012)
5	**scaffold03154**	CRY2	Cryptochrome Circadian Clock 2	Genes / proteins on the CCO	(Fu *et al.* 2002; Helfer *et al.* 2006; Albrecht 2012; Cassone & Westneat 2012; Peek *et al.* 2012)

6	**scaffold01501**	PER3	Period Circadian Clock 3	Genes / proteins on the CCO	(Yasuo *et al.* 2003; Feillet *et al.* 2006; Helfer *et al.* 2006; Albrecht 2012; Cassone & Westneat 2012; Ikegami & Yoshimura 2012; Peek *et al.* 2012)
7	**scaffold00959**	CK1α (Csnk1a1)	Casein Kinase 1, Alpha	Genes / proteins on the CCO	(Lowrey *et al.* 2000; Lee *et al.* 2011; Zheng *et al.* 2014; Lam *et al.* 2015)
8	**scaffold00376**	REV-ERBβ/ NR1D2	Nuclear Receptor Subfamily 1, Group D, Member 2	Genes / proteins on the CCO	(Preitner *et al.* 2002; Burris 2008; Zhang & Kay 2010; Bugge *et al.* 2012; Cho *et al.* 2012; Yang *et al.* 2013)
9	**scaffold01473**	OPN4	Opsin 4	Genes / proteins on the CCO	(Berson *et al.* 2002; Halford *et al.* 2009; Zhang & Kay 2010; Shinomiya *et al.* 2014)
10	**scaffold03142**	FBXL3	F-Box And Leucine-Rich Repeat Protein 3	Genes / proteins on the CCO	(Ko & Takahashi 2006; Virshup & Forger 2007; Yoo *et al.* 2013; Hirano *et al.* 2017)
11	**scaffold08092**	NOCT	Circadian Deadenylase NOC	Genes / proteins on the CCO	(Douris & Green 2008; Li *et al.* 2008)
12	**scaffold00009**	CSNK1E	Casein Kinase 1, Epsilon	Genes / proteins on the CCO	(Reppert & Weaver 2002; Gachon *et al.* 2004; Steinmeyer *et al.* 2009; Bass & Takahashi 2010; Kategaya *et al.* 2012; Steinmeyer *et al.* 2012; Storz *et al.* 2013)

13	**scaffold05082**	CK1δ (Csnk1d)	Casein Kinase 1, Delta	Genes / proteins on the CCO	(Xu *et al.* 2005)
14	**scaffold02191**	Dec-02	Basic Helix-Loop-Helix Family, Member E41	Genes / proteins on the CCO	(Honma *et al.* 2002; Nakashima *et al.* 2008; Zhang & Kay 2010)
15	**scaffold02192**	SFPQ	Splicing Factor Proline And Glutamine Rich	Genes / proteins on the CCO	
16	**scaffold00237**	NPAS2/MOP4	Neural pas protein domain 2	Genes / proteins on the CCO	(Gachon *et al.* 2004; Steinmeyer *et al.* 2009; Haque *et al.* 2010; Zhang & Kay 2010; Ruegg *et al.* 2014)
17	**scaffold03247**	PPARGC1A	Peroxisome Proliferator-Activated Receptor Gamma, Coactivator 1	Genes / proteins on the CCO	(Zhang & Kay 2010)
18	**scaffold05859**	PK2	Prokineticin 2	Genes / proteins that form 'outputs' of the CCO.	(Cheng *et al.* 2002; Cheng *et al.* 2005; Li *et al.* 2006)
19	**scaffold00942**	RORalpha1	Retinoic acid receptor-Related Orphan Receptor A	Genes / proteins on the CCO	(Dardente *et al.* 2010)
20	**scaffold00662**	PASD1	Circadian Clock Protein PASD1	Genes / proteins on the CCO	(van der Veen *et al.* 2017)
21	**scaffold00007**	TPH1	Tryptophan 5-hydroxylase/ Tryptophan Hydroxylase 1	Genes / proteins that form 'outputs' of the CCO.	(Malek *et al.* 2004)
22	**scaffold06435**	MEL1A	Melatonin Receptor 1 A	Genes / proteins that form 'outputs' of the CCO.	(Barrett *et al.* 2003)
23	**scaffold00068**	ASMT	Hydroxyindolo-o-methyl transferase (HIOMT)/ Acetylserotonin O-Methyltransferase	Genes / proteins that form 'outputs' of the CCO.	(Gauer & Craft 1996)

24	**scaffold00488**	VIP	Vasoactive Intestinal Peptide	Genes / proteins of the HPG axis and take part in regulation of expression of CCO	(Romijn *et al.* 1996; Colwell *et al.* 2003; Maywood *et al.* 2007; Ciarleglio *et al.* 2009; Loh *et al.* 2011; Evans *et al.* 2015)
25	**scaffold02723**	VIPR1	Pituitary Adenylase Cyclase-Activating polypeptide Type II	Genes / proteins of the HPG axis and take part in regulation of expression of CCO	(Romijn *et al.* 1996; Colwell *et al.* 2003; Maywood *et al.* 2007; Ciarleglio *et al.* 2009; Loh *et al.* 2011; Evans *et al.* 2015)
26	**scaffold00330**	VIPR2	Pituitary Adenylate Cyclase-Activating Polypeptide Type III	Genes / proteins of the HPG axis and take part in regulation of expression of CCO	(Romijn *et al.* 1996; Colwell *et al.* 2003; Maywood *et al.* 2007; Ciarleglio *et al.* 2009; Loh *et al.* 2011; Evans *et al.* 2015)
27	**scaffold01032**	FSHβ	Follicle Stimulating Hormone, Beta Polypeptide	Genes / proteins of the HPG axis	(Yoo *et al.* 2013)
28	**scaffold03952**	FSHR	Follicle Stimulating Hormone Receptor	Genes / proteins of the HPG axis	(Yoo *et al.* 2013)
29	**scaffold03484**	SIX1	SIX Homeobox 1	Genes / proteins of the HPG axis	(Helm *et al.* 2013)
30	**scaffold00012**	SIX2	SIX Homeobox 2	Genes / proteins of the HPG axis	(Helm *et al.* 2013)
31	**scaffold01801**	SIX4	SIX Homeobox 4	Genes / proteins of the HPG axis	(Helm *et al.* 2013)
32	**scaffold04454**	LHR	Luteinizing Hormone Receptor/ Luteinizing hormone/choriogonadotropin receptor	Genes / proteins of the HPG axis	

33	**scaffold01956**	GnRH receptor II	Gonadotropin-Releasing Hormone Receptor (Type 2)	Genes / proteins of the HPG axis	(Halford *et al.* 2009; Williams *et al.* 2017)
34	**scaffold13635**	EYA1	EYA Transcriptional Coactivator And Phosphatase 1	Genes / proteins of the HPG axis	
35	**scaffold00565**	EYA2	EYA Transcriptional Coactivator And Phosphatase 2	Genes / proteins of the HPG axis	
36	**scaffold12955**	EYA3	EYA Transcriptional Coactivator And Phosphatase 3	Genes / proteins of the HPG axis	(Dardente *et al.* 2010; Dupre *et al.* 2010; Ikegami & Yoshimura 2012; Williams *et al.* 2017)
37	**scaffold08021**	TSHR	Thyroid Stimulating Hormone Receptor	Genes / proteins of the HPG axis	(de Miera *et al.* 2013; Dardente *et al.* 2014)
38	**scaffold00360**	TSHβ	Thyroid Stimulating Hormone, Beta	Genes / proteins of the HPG axis	(Barrett & Bolborea 2012; Dardente *et al.* 2014; Wood & Loudon 2014; Williams *et al.* 2017)
39	**scaffold00586**	TRH/ TRF/ PRH	Thyrotrophin- releasing hormone	Genes / proteins of the HPG axis	(Bolborea & Dale 2013; Dardente *et al.* 2014; Shinomiya *et al.* 2014)
40	**scaffold00002**	TRHR	Thyrotrophin- releasing hormone Receptor	Genes / proteins of the HPG axis	Bolborea & Dale 2013; Dardente *et al.* 2014; Shinomiya *et al.* 2014)
41	**scaffold01316**	Mel1c/GPR 50	Melatonin Receptor C	Genes / proteins of the HPG axis	(Barrett *et al.* 2003; Barrett & Bolborea 2012; Bolborea & Dale 2013)

42	**scaffold06762**	CRH receptor	Corticotropin-releasing hormone Receptor 1	Genes / proteins of the HPG axis	
43	**scaffold00318**	ACTH	Adrenocorticotropic hormone/ Melanocortin Receptor 2	Genes / proteins of the HPG axis	(Cockrem 2013)
44	**scaffold00876**	TG	Thyroglobulin	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	(Dietrich *et al.* 2012)
45	**scaffold00265**	DIO2	Deiodinase, Iodothyronine, Type II	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	(Peeters *et al.* 1999; Nakahara *et al.* 2002; Nowak & Zawilska 2003; Csernus *et al.* 2004; Faluhelyi *et al.* 2004; Moore *et al.* 2005; Yasuo *et al.* 2006; Racz *et al.* 2008; Steinmeyer *et al.* 2009; Mueller *et al.* 2011; Dardente *et al.* 2014)
46	**scaffold02963**	SLC16A2	Solute Carrier Family 16, Member 2 (Thyroid Hormone Transporter)	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	(Barrett & Bolborea 2012)
47	**scaffold06584**	SLC2A2/ GLUT2	Solute carrier Family 2 , member 2	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	(Jones *et al.* 2012)
48	**scaffold07849**	SLC2A3/ GLUT3	Solute carrier Family 2 , member 3	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	(Jones *et al.* 2012)
49	**scaffold00169**	TTR	Transthyretin-like	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012; Bolborea & Dale 2013)
50	**scaffold13235**	GLUC1/SLC2A1	Solute carrier family 2 facilitated glucose transporter member 1	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)

51	**scaffold00775**	SLC1A2	Glial High Affinity glutamate transporter member 3	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)
52	**scaffold01723**	SLC1A4	Glial High Affinity glutamate transporter member 4	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)
53	**scaffold01890**	HSPA5	Heat shock 70KDa prot 5	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)
54	**scaffold00052**	HSP90AA1	Heat shock prot 90KDa alpha	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)
55	**scaffold01457**	PARL	Presenilin Associated, Rhomboid-Like	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)
56	**scaffold00554**	NPVF	Neuropeptide VG Precursor	Found association to seasonality and reproduction (in mammals)	(Simonneaux *et al.* 2013)
57	**scaffold00822**	PRL	Prolactin	Found association to seasonality and reproduction (in mammals)	(Prendergast *et al.* 2002)
58	**scaffold01818**	DRD1	Dopamine Receptor D1	References of association with exploratory migration	(Fidler & Gwinner 2003; Korsten *et al.* 2010; Liedvogel *et al.* 2011; Mueller *et al.* 2011)
59	**scaffold00238**	DRD2	Dopamine Receptor D2	References of association with exploratory migration	(Fidler & Gwinner 2003; Korsten *et al.* 2010; Liedvogel *et al.* 2011; Mueller *et al.* 2011)
60	**scaffold01266**	DRD3	Dopamine Receptor D3	References of association with exploratory migration	(Fidler & Gwinner 2003; Korsten *et al.* 2010; Liedvogel *et al.* 2011; Mueller *et al.* 2011)
61	**scaffold00565**	MAPK1	Mitogen-Activated Protein Kinase 1	Regulates stability of several CCO proteins	(Akashi & Nishida 2000; Hirota & Fukada 2004;

					Sanada *et al.* 2004; Vitalini *et al.* 2007; Fuchikawa *et al.* 2010)
62	**scaffold00197**	PHB2	Prohibitin 2	Regulates transcription of CCO genes	(Knippschild *et al.* 2005; Ko & Takahashi 2006; Kategaya *et al.* 2012)
63	**scaffold00973**	ESR2	Estrogen Receptor 2 (ERβ)	Regulates transcription of CCO genes	(Dawson *et al.* 2001; Bozek *et al.* 2009)
64	**scaffold00116**	ESRRB	Estrogen-Related Receptor Beta	Regulates transcription of CCO genes	(Dawson *et al.* 2001; Bozek *et al.* 2009)
65	**scaffold00009**	TRIOBP	Trio-Associated Repeat On Actin	Described with a role in brain physiology modifications during migration	(Johnston *et al.* 2016)
66	**scaffold00118**	NEK2	Serine/Threonine-Protein Kinase Nek2	Described with a role in brain physiology modifications during migration	(Jones *et al.* 2012)
67	**scaffold00022**	SUMO1	Ubiquitin-Homology Domain Protein PIC1	Involved in transcriptional regulation	(Bellet & Sassone-Corsi 2010)
68	**scaffold02090**	SCTR (paralog_of_adcyap1R1)	Secretin Receptor	Paralog of adcyap1Receptor	
69	**scaffold00103**	MBOAT/GOAT	O-Acyltransferase (Membrane Bound) Domain Containing	Fat metabolism	(Goymann *et al.* 2017)
70	**scaffold08506**	GHSR	Ghrelin Receptor	Fat metabolism	(Goymann *et al.* 2017)
71	**scaffold00082**	RYR3	Brain Ryanodine Receptor-Calcium Release Channel	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)

72	**scaffold00884**	PDE3A	Cyclic GMP-Inhibited Phosphodiesterase A	Fat metabolism	(Kim *et al.* 2007; Plano *et al.* 2012; Stuber *et al.* 2013)
73	**scaffold00852**	GREB1	Gene Regulated By Estrogen In Breast Cancer	Described to be associated to migratory timing in salmons.	(Hess *et al.* 2016)
74	**scaffold00573**	GREB1L	Growth Regulation By Estrogen In Breast Cancer 1 Like	Described to be associated to migratory timing in salmons.	(Hess *et al.* 2016)
75	**scaffold00039**	FGF12a	Fibroblast Growth Factor Homologous Factor 12a	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
76	**scaffold04623**	FGF12b	Fibroblast Growth Factor Homologous Factor 12b	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
77	**scaffold00213**	CPNE1	Copine I	Described with a role in brain physiology modifications during migration	(Jones *et al.* 2008; Ruegg *et al.* 2014)
78	**scaffold01439**	CRABP1	Cellular Retinoic Acid Binding Protein 1	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
79	**scaffold00197**	EPHA1	Tyrosine-Protein Kinase Receptor EPH 1	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
80	**scaffold00234**	EPHB10	Tyrosine-Protein Kinase Receptor EPH 10	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
81	**scaffold01190**	EPHB2	Tyrosine-Protein Kinase Receptor EPH 2	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)

82	**scaffold00523**	EPHA3	Tyrosine-Protein Kinase Receptor EPH 3	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
83	**scaffold10364**	EPHB3	Tyrosine-Protein Kinase Receptor EPH 3	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
84	**scaffold00110**	EPHB4	Tyrosine-Protein Kinase Receptor EPH 4	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
85	**scaffold00197**	EPHB6	Tyrosine-Protein Kinase Receptor EPH 6	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
86	**scaffold00310**	PAK3	P21 Protein (Cdc42/Rac)-Activated Kinase 3	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
87	**scaffold00713**	PDGFA	Platelet Derived Growth Factor Subunit A	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
88	**scaffold05635**	RASL10B	Ras-Like Protein Family Member 10B	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
89	**scaffold00030**	ADAM10	A Disintegrin And Metalloproteinase Domain	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
90	**scaffold02440**	ARPC1B	Actin Related Protein 2/3 Complex Subunit 1B	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)

91	**scaffold00191**	GLI3	DNA-Binding Protein	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
92	**scaffold00139**	FGF9	Fibroblast Growth Factor Homologous Factor 1	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
93	**scaffold06360**	IWS1	IWS1-Like Protein	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neu ral plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
94	**scaffold00023**	LAMA4	Laminin, Alpha 4	Cell function, involved in several processes	(Johnston *et al.* 2016)
95	**scaffold00620**	HTR1F	5-Hydroxytryptamine Receptor 1F	Perhaps involved in brain physiology modifications during migration	
96	**scaffold05739**	HTR1E	Serotonin Receptor 1E	Perhaps involved in brain physiology modifications during migration	
97	**scaffold00274**	RDX	Radixin	Perhaps involved in brain physiology modifications during migration	
98	**scaffold00018**	HTR7	Serotonin Receptor 7	Perhaps involved in brain physiology modifications during migration	
99	**scaffold00627**	FNDC3A	Fibronectin Type-III Domain-Containing Protein 3A	Perhaps involved in brain physiology modifications during migration	
100	**scaffold00655**	MTUS2	Cardiac Zipper Protein	Cell function, involved in several processes	
101	**scaffold00565**	PER2*	Period Circadian Clock 2	Genes / proteins on the CCO	(Yasuo *et al.* 2003; Feillet *et al.* 2006; Helfer *et al.* 2006; Albrecht 2012; Cassone & Westneat 2012; Ikegami & Yoshimura 2012; Peek *et al.* 2012)

-	**scaffold00753**	RORβ	Nuclear receptor, Related Orphan Receptor Beta	Genes / proteins on the CCO	(Dardente *et al.* 2010)
-	**scaffold15614**	REV-ERBα/ NR1D1	Nuclear Receptor Subfamily 1 Group D Member 1	Genes / proteins on the CCO	(Triqueneaux *et al.* 2004; Herzog 2007; Zhang & Kay 2010)
-	**scaffold10203**	MEL1B	Melatonin Receptor 1 B	Genes / proteins that form 'outputs' of the CCO.	(Barrett *et al.* 2003)
-	**scaffold01875**	HSPA2	Heat shock 70KDa prot 2	Migration-related transcripts according to Jones *et al.* 2008	(Jones *et al.* 2012)
-	**scaffold00276**	PRLR	Prolactin Receptor Precursor	Found association to seasonality and reproduction (in mammals)	(Prendergast *et al.* 2002)
-	**scaffold01105**	MC1R	Melanocortin 1 receptor (Alpha Melanocyte Stimulating Hormone receptor)	Variation in the pigmentation of birds when they migrate	(Baiao & Parker 2012)
-	**scaffold00446**	RHOJ	Ras Homolog Family Member J	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)
-	**scaffold09876**	SBK1	Serine/Threonine-Protein Kinase SBK1	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Baerwald *et al.* 2016; Johnston *et al.* 2016)
-	**scaffold00321**	TLN1	Talin	Perhaps involved in brain physiology modifications during migration	
-	**scaffold00197**	ZYX	Zyxin	Found to be differentially expressed seasonally in a migratory bird species. Its function is related to neural plasticity associated with seasonal changes in the brain.	(Johnston *et al.* 2016)

	"New" candidate genes				*Fst*_9vs10
102	**scaffold00258**	HDAC1	Histone Deacetylase 1	Genes / proteins of the HPG axis and take part in regulation of expression of CCO	0.5848
103	**scaffold01265**	SLC9A8	Solute carrier Family 9 , member 8	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	0.5493
104	**scaffold00377**	PINOPSIN	Pineal gland-specific opsin	Photoreceptor	0.5694
105	**scaffold03029**	NAALAD2	Glutamate Carboxypeptidase III	Energy metabolism	0.6115
106	**scaffold03413**	GBF1	Golgi Brefeldin A Resistant Guanine Nucleotide Exchange Factor 1	Energy metabolism	0.8216
107	**scaffold00021**	APOB	Apolipoprotein B	Fat metabolism	0.5492
108	**scaffold02927**	ACACA	Acetyl-CoA Carboxylase Alpha	Fat metabolism	0.5493
109	**scaffold01066**	GCN1	General Control Of Amino-Acid Synthesis 1-Like Protein 1	Fat metabolism	0.619
110	**scaffold05702**	AGRN	Agrin	Neuromuscular function	0.5493
111	**scaffold00497**	PDGFRA	Alpha-Type Platelet-Derived Growth Factor Receptor	Described with a role in brain physiology modifications	0.5815
112	**scaffold01802**	CDH1	Cadherin 1	Described with a role in brain physiology modifications	0.5815
113	**scaffold00955**	CDH13	Cadherin 13	Described with a role in brain physiology modifications	0.5905
114	**scaffold00015**	MYH2	Myosin Heavy Chain 2	Described with a role in brain physiology modifications	0.6097
115	**scaffold02232**	CNKSR1	CNK Homolog Protein 1	Cell function, involved in several processes	0.5593

116	**scaffold00130**	GALNT18	Polypeptide N-Acetylgalactosaminyltransferase 18	Cell function, involved in several processes	0.5481
117	**scaffold08300**	UBE4B	Ubiquitin-Fusion Degradation Protein 2	Cell function, involved in several processes	0.5848
118	**scaffold00008**	LTBP1	Latent Transforming Growth Factor Beta Binding Protein 1	Cell function, involved in several processes	0.6003
119	**scaffold01877**	NDUFB2	NADH:Ubiquinone Oxidoreductase Subunit B2	Cell function, involved in several processes	0.6097
120	**scaffold00444**	RELN	Reelin	Cell function, involved in several processes	0.6192
-	**scaffold00015**	MYO10	Myosin X	Described with a role in brain physiology modifications	0.6212
-	**scaffold04262**	RP1	Retinitis Pigmentosa 1	Related to photoreceptors in the retina	0.6301
-	**scaffold00173**	CNDP1	Glutamate Carboxypeptidase-Like Protein 2	Fat metabolism	0.5527
-	**scaffold00170**	FER	FER Tyrosine Kinase	Cell function, involved in several processes	0.5493
-	**scaffold07582**	TGF-B	Transforming Growth Factor Beta 1	Genes / proteins of the HPG axis	0.5493
-	**scaffold00910**	ELP4	Elongator Acetyltransferase Complex Subunit 4	Cell function, involved in several processes	0.5593
-	**scaffold00065**	PTPN11	Protein Tyrosine Phosphatase, Non-Receptor Type 11	Cell function, involved in several processes	0.5601
-	**scaffold04437**	EPG5	Ectopic P-Granules Autophagy Protein 5 Homolog	Cell function, involved in several processes	0.603

	Not found in the godwit genome				References
-	-	AANAT	Aralkylamine N-Acetyltransferase	Genes / proteins that form 'outputs' of the central circadian oscillator (CCO).	(Foulkes *et al.* 1997; Chong *et al.* 2000; Barrett *et al.* 2003; Leder *et al.* 2006; Toller *et al.* 2006; Steinmeyer *et al.* 2009; Haque *et al.* 2010)
-	-	NFIL3	Nuclear factor interleukin 3	Regulates transcription of CCO genes	(Yasuo *et al.* 2003; Herzog 2007)
-	-	ADCYAP1	Adenylate Cyclase Activating Polypeptide 1 (Pituitary)	Genes / proteins of the HPG axis	(Yasuo *et al.* 2005; Nakao *et al.* 2008; Ikegami & Yoshimura 2012; de Miera *et al.* 2013; Herwig *et al.* 2013; Yoshimura 2013; Ruegg *et al.* 2014)
-	-	SIX3	SIX Homeobox 3	Genes / proteins of the HPG axis	(Helm *et al.* 2013)
-	-	DIO3	Deiodinase, iodothyronine type III	Genes / proteins responding to products of the HPG axis - in particular the thyroid.	(Yoshimura 2013; Williams *et al.* 2017)
-	-	DRD4	Dopamine Receptor D4	References of association with exploratory migration	(Fidler & Gwinner 2003; Korsten *et al.* 2010; Liedvogel *et al.* 2011; Mueller *et al.* 2011)
-	-	ISCA1	Iron-Sulfur Cluster Assembly 1	Related to navigation molecular mechanisms for migration	(Mandilaras & Missirlis 2012)
-	-	EGR1	Early Growth Response 1	Found association with circadian rhythms (in mammals)	(Barrett & Bolborea 2012)

Table 3.3 List of the 20 individuals selected from the extremes of the behavioural continua used for the preliminary work from which we obtained the SNPs that were subsequently genotyped for further investigation (please see text for more details). Indiv. ID indicates individuals' ID, which denotes the departure site as the first letter (C = Catlins Coast; F = Firth of Thames; M = Manawatu Estuary) followed by a unique metal band number starting with Y, e.g. Y12699; mdd= mean departure dates or mean chronophenotypes; mdd2= standardized by year per site mean departure dates or standardized mean chronophenotypes. The column with the labels indicates to which chronophenotipic category (i.e. E = 'early' vs. L = 'late') birds' were assigned to calculate *Fst* per SNP. Note that individuals 10E was discarded for the analyses since read coverage of this genome indicated some problems in its library pooling step.

Indiv. ID	Label	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	mdd	mdd2
MY13425	**1E**	73	74	76	71	76	72	73	74	69	74	73	73
MY8861	**2E**						68	68	69		67	68	68
CY12616	**3E**						65	64				65	65
FY10320	**4E**							67				67	67
FY18612	**5E**							62	70			66	66
MY12705	**6E**						72	78	70			73	73
MY13587	**7E**						72	73	74	71	68	72	72
MY13582	**8E**						71	79	73			74	74
MY13583	**9E**						72	73	73	69	73	72	72
CY12634	**10E**							71	73	69		71	71
MY12710	**1L**						83	79				81	81
CY12621	**2L**						86					86	86
FY12782	**3L**							84	79			82	82
FY12785	**4L**							85	79			82	82
FY12752	**5L**							91				91	91
MY8814	**6L**							86				86	86
MY13428	**7L**	87	90	87	82	88	82					86	86

MY12724	**8L**							85				85	85
MY13596	**9L**						80					80	80

Figure 3.1 Results of the simulated analysis to assess power of MLM to detect associated SNPs. (a) Q-Q plot showing expected –log10(p-value) versus observed –log10(p-value) resultant from the simulated analysis to assess power of LMM to detect associated SNPs. Lambda (λ) = 1.05. Three SNPs reported to be associated with migration timing (Hess *et al.* 2016) (open squares) were included in our dataset (3,412 SNPs, filled squares) to assess the power of detection using LMM. Note that the three SNPs taken from Hess *et al.* 2016 (open squares and noted as: 54772_22, 52458_16 and 47080_53) showed strong association (as expected), in contrast to our candidate SNPs which remained not significant and fitting the expected p-value. The dashed line indicate the Bonferroni corrected alpha level of 0.05 (Log10(P) = -7.2).

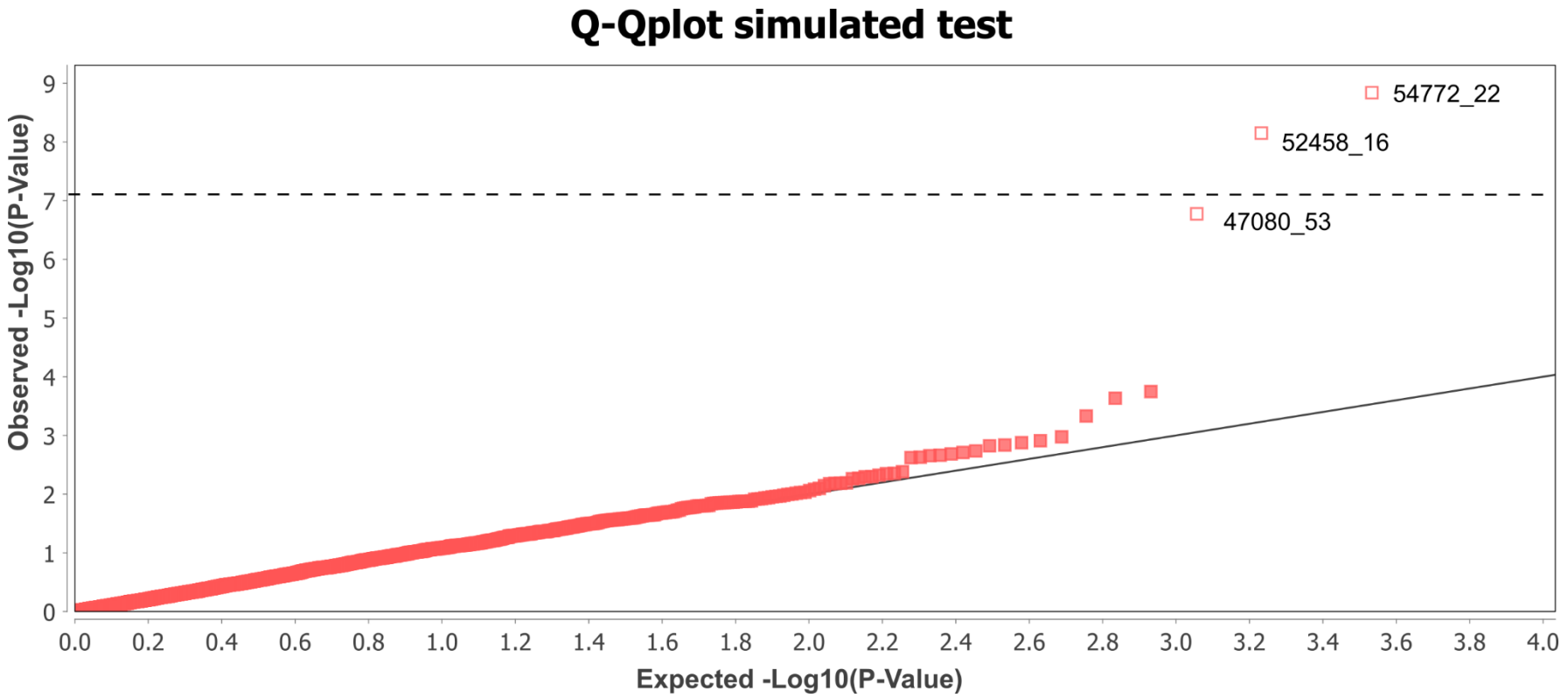


Table 3.4 Mean read depth coverage and genome coverage assessment of the 10 ‘early’ 9 ‘late’ individuals for the preliminary work (see text for more details). *Abbreviations*: CI = Interval of confidence.

Sample	Genome Coverage	Mean Read Depth	95% CI
1E	83%	4.85	1–10
2E	82%	4.76	1–10
3E	80%	3.58	1–8
4E	82%	4.54	1–10
5E	82%	4.55	1–10
6E	79%	3.26	1–7
7E	81%	3.76	1–8
8E	79%	3.42	1–7
9E	82%	4.26	1–9
10E	80%	3.54	1–8
1L	82%	4.1	1–9
2L	77%	2.91	1–7
3L	79%	3.24	1–7
4L	78%	3.12	1–7
5L	79%	3.3	1–7
6L	80%	3.67	1–8
7L	77%	2.98	1–7
8L	75%	2.74	1–6
9L	75%	2.67	1–6

Figure 3.2 FastStructure plot of individual ancestry at $K = 2$. (a) 3,412 candidate SNPs, (b) 319 SNPs separated at least by 10 Kb. Individuals were sorter by their chronophenotype (x-axis) and organized by site, which are separated with dash lines (i.e. Manawatu, Firth of Thames and Catlins).

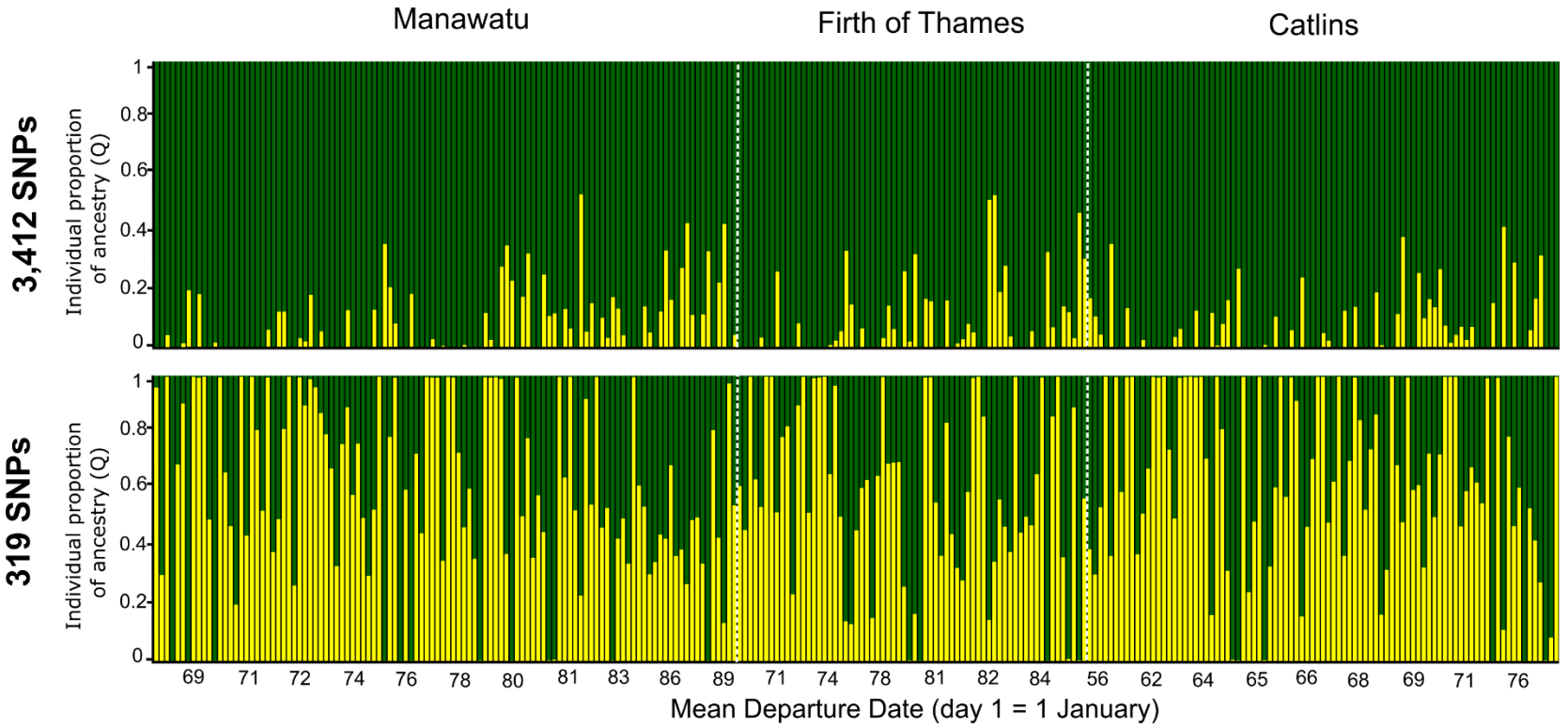


Figure 3.3 Correlation between individual ancestry (Q) and mean standardized chronophenotype for the three sites: dots (Catlins), triangles (Firth of Thames) and squares (Manawatu), (a) for candidate 124 SNPs separated by at least 100 Kb; (b) for 27 microsatellites used in Chapter 1. Interval of confidence are shown in shadow and r^2 are indicated only for significant correlations. These analyses were done using common individuals (n = 141: 15 from Catlins, 41 from Firth of Thames and 91 from Manawatu) between the present study and Chapter 2.

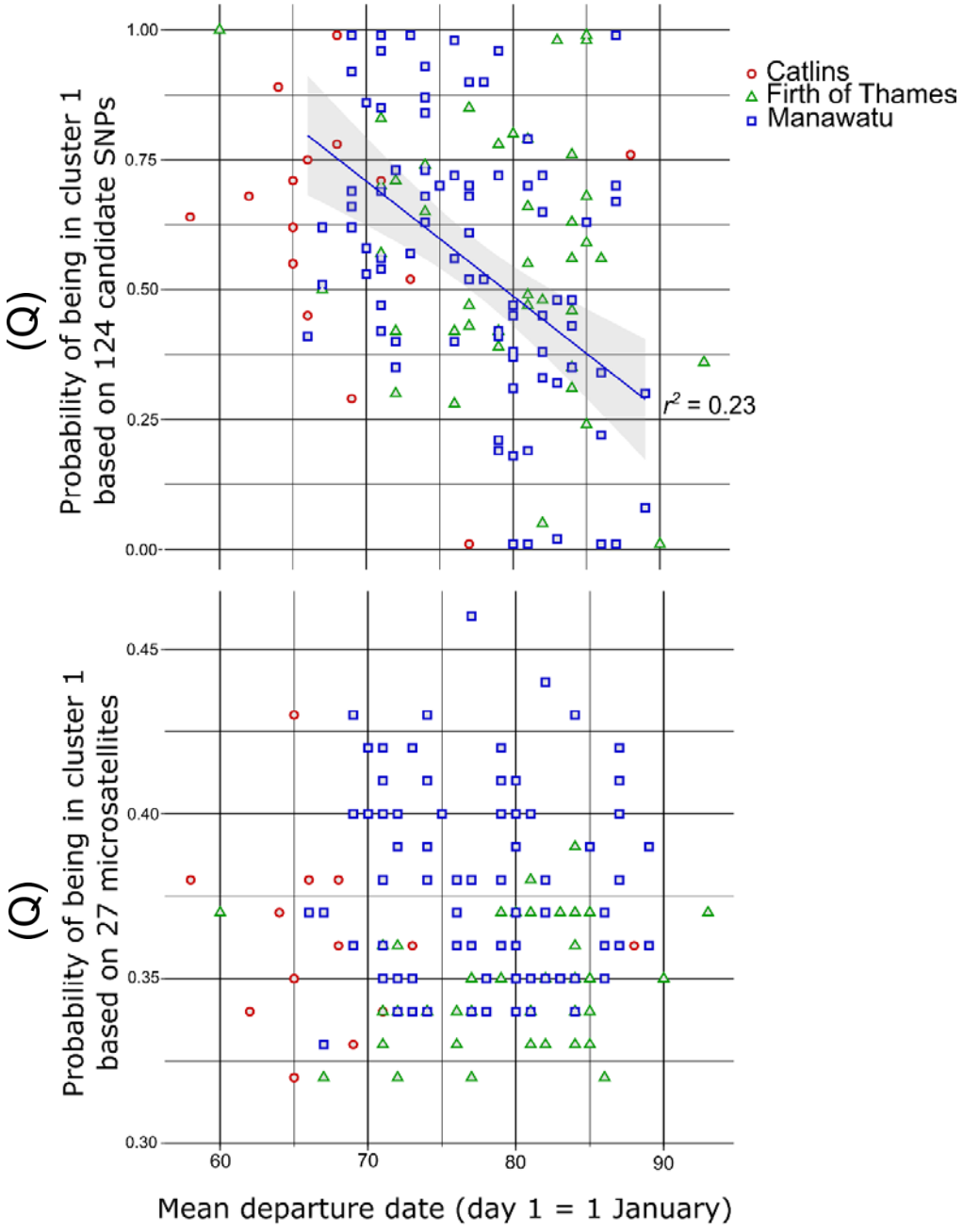


Figure 3.4 QQ-plot of the Linear Mixed Model using departure dates from Asia (n = 34 individuals) and 3,412 SNPs. Note that no SNP showed to be significant after correction for multiple testing.

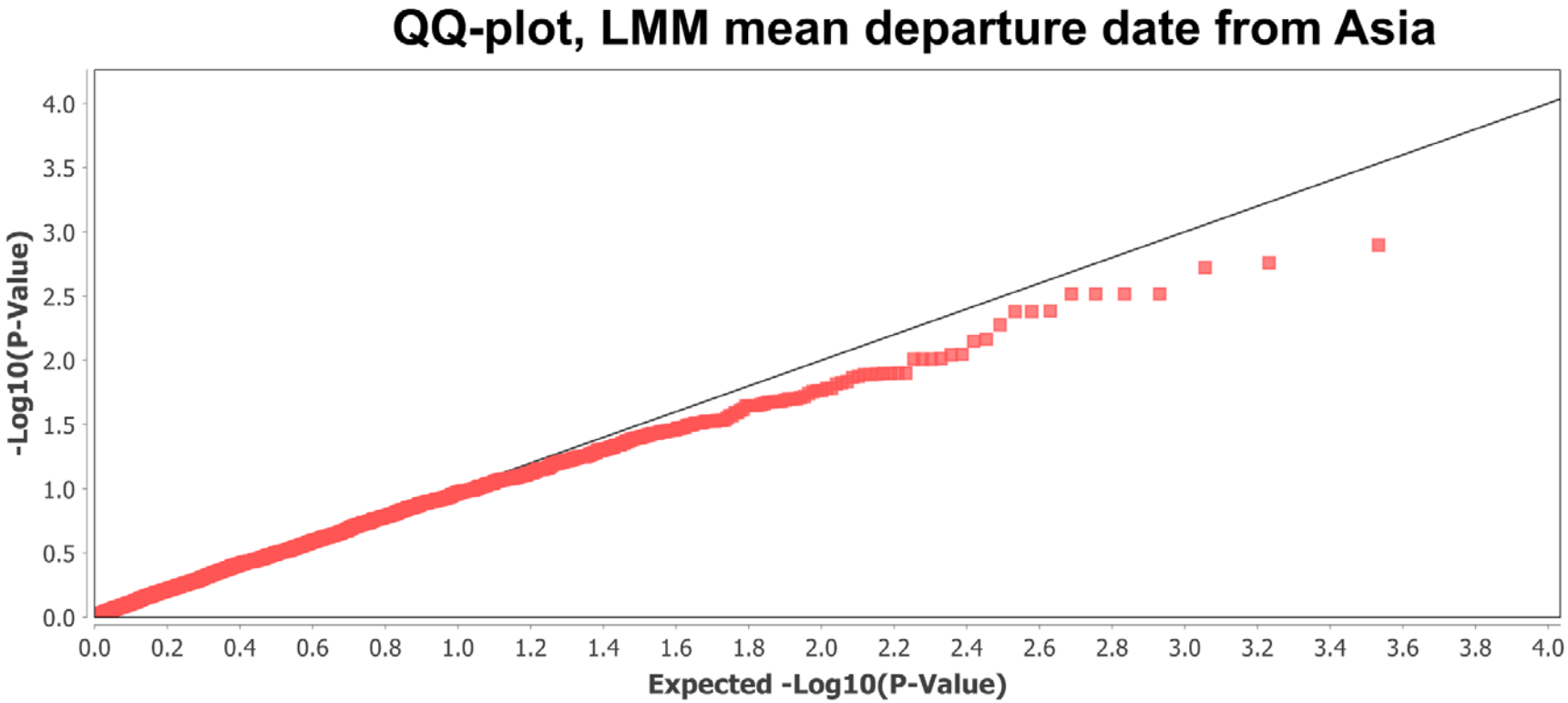


Figure 3.5 Differences in *Fst* between 10 ‘early’ versus 9 ‘late’ phenotypic genomes. Plot shows non-overlapping 10 kb windows along the godwit reference genome, with scaffolds ranked by their order in the genome reference (i.e. not in true genomic order). Grey and blue points show observed distributions of *Fst* values per SNP (total = 8,941,974), with blue points highlighting outliers with the top 0.5%. Red squares indicate genomic regions where outliers are observed across at least 3 adjacent windows (i.e. over 30 kb or more of the reference genome). The vertical placement of these red squares has no meaning.

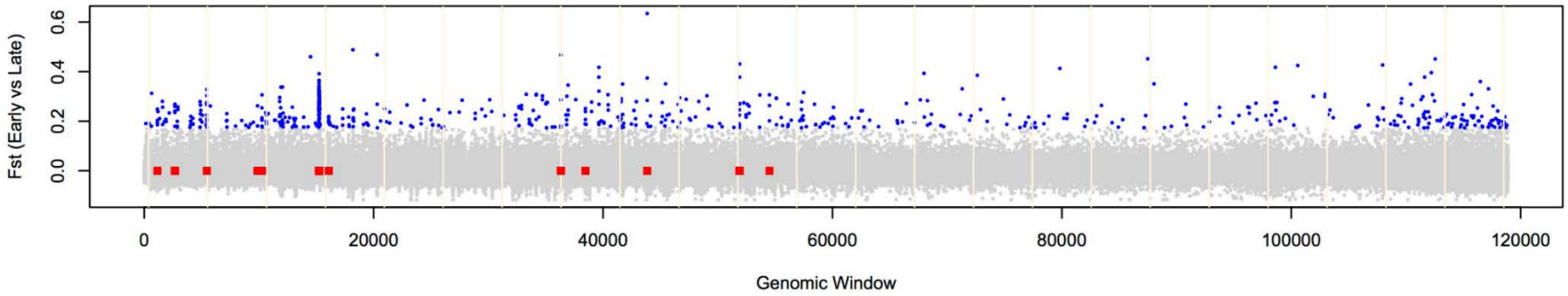
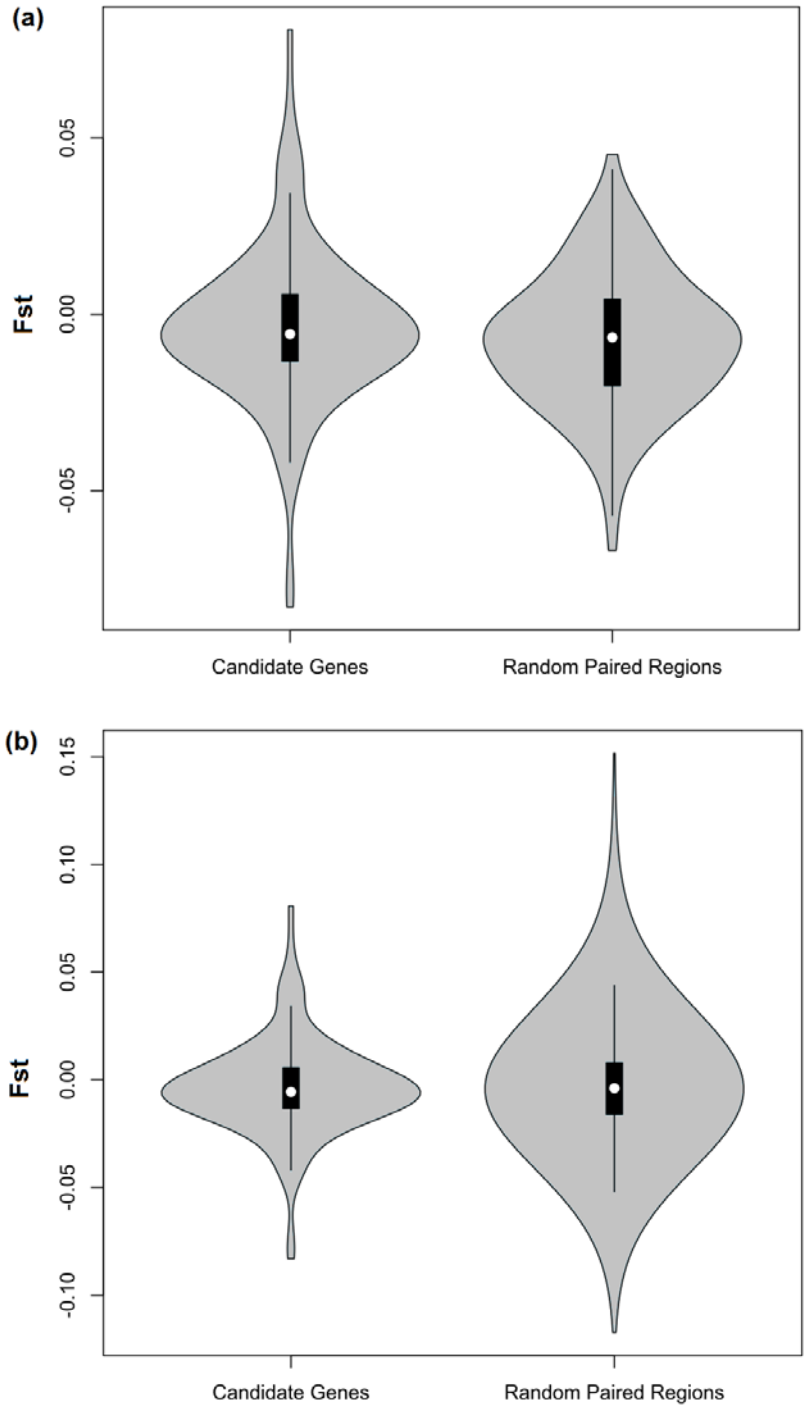


Figure 3.6 Plot showing the distribution of the *Fst* mean values of candidate gene regions (n = 120) compared to random regions of the same length as the candidate gene regions using the 19 individuals re-sequenced. (a) 120 random regions were compared. (b) 1000 random regions were compared. White dots indicate means for each group of regions (i.e. candidate and random regions), and width of the grey area indicates density.



Godwits in a sunset. Landscape after a hard – and fun – day of fieldwork at the Manawatu River estuary (November 2014).

References

Abraham U, Albrecht U, Brandstatter R (2003) Hypothalamic circadian organization in birds. II. *Clock* gene expression. *Chronobiology International* **20**, 657-669.

Abraham U, Albrecht U, Gwinner E, Brandstatter R (2002) Spatial and temporal variation of passer Per2 gene expression in two distinct cell groups of the suprachiasmatic hypothalamus in the house sparrow (*Passer domesticus*). *European Journal of Neuroscience* **16**, 429-436.

Akashi M, Nishida E (2000) Involvement of the MAP kinase cascade in resetting of the mammalian circadian clock. *Genes & Development* **14**, 645-649.

Albrecht U (2012) Timing to perfection: The biology of central and peripheral circadian clocks. *Neuron* **74**, 246-260.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Aschoff J (1960) Exogenous and endogenous components in circadian rhythms. *Cold Spring Harbor Symposia on Quantitative Biology* **25**, 11-28.

Aschoff J, Fatranska M, Giedke H, *et al.* (1971) Human circadian rhythms in continuous darkness: Entrainment by social cues. *Science* **171**, 213-215.

Backstrom N, Ovarnstrom A, Gustafsson L, Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* **2**, 435-438.

Baerwald MR, Meek MH, Stephens MR, *et al.* (2016) Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. *Molecular Ecology* **25**, 1785-1800.

Baiao PC, Parker PG (2012) Evolution of the melanocortin-1 receptor (MC1R) in Boobies and Gannets (Aves, Suliformes). *Journal of Heredity* **103**, 322-329.

Bao R, Huang L, Andrade J, *et al.* (2014) Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Informatics* **13**, 67-82.

Barrett P, Bolborea M (2012) Molecular pathways involved in seasonal body weight and reproductive responses governed by melatonin. *Journal of Pineal Research* **52**, 376-388.

Barrett P, Schuster C, Mercer J, Morgan PJ (2003) Sensitization: A mechanism for melatonin action in the pars tuberalis. *Journal of Neuroendocrinology* **15**, 415-421.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* **12**, 767-780.

Barrett RK, Page TL (1989) Effects of light on circadian pacemaker development. I. The freerunning period. *Journal of Comparative Physiology A* **165**, 41-49.

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11-21.

Bass J, Takahashi JS (2010) Circadian integration of metabolism and energetics. *Science* **330**, 1349-1354.

Battley PF (2006) Consistent annual schedules in a migratory shorebird. *Biology Letters* **2**, 517-520.

Battley PF (2007) Plumage and timing of migration in bar-tailed godwits: a comment on Drent et al. (2003). *Oikos* **116**, 349-352.

Battley PF, Conklin JR (2017) Geolocator wetness data accurately detect periods of migratory flight in two species of shorebird. *Wader Study* **124**, 112-119.

Battley PF, Piersma T (2005) Body composition and flight ranges of bar-tailed Godwits (*Limosa lapponica baueri*) from New Zealand. *Auk* **122**, 922-937.
Battley PF, Schuckard R, Melville DS (2011) Movements of bar-tailed godwits and red knots within New Zealand. Science for conservation 315. New Zealand Department of Conservation, Wellington. 56 p.
Battley PF, Warnock N, Tibbitts TL, *et al.* (2012) Contrasting extreme long-distance migration patterns in bar-tailed godwits *Limosa lapponica*. *Journal of Avian Biology* **43**, 21-32.
Bauer S, Gienapp P, Madsen J (2008) The relevance of environmental conditions for departure decision changes en route in migrating geese. *Ecology* **89**, 1953-1960.
Bazzi G, Ambrosini R, Caprioli M, *et al.* (2015) *Clock* gene polymorphism and scheduling of migration: a geolocator study of the barn swallow *Hirundo rustica*. *Scientific Reports* **5**, 12443.
Bazzi G, Cecere JG, Caprioli M, *et al.* (2016a) *Clock* gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Molecular Ecology* **25**, 6077–6091.
Bazzi G, Galimberti A, Hays QR, *et al.* (2016b) *Adcyap1* polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the wilson's warbler (*Cardellina pusilla*). *Ecology and Evolution* **6**, 3226-3239.
Bazzi G, Podofillini S, Gatti E, *et al.* (2016c) Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird. *Current Zoology*, **zow103**.
Bell-Pedersen D, Cassone VM, Earnest DJ, *et al.* (2005) Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nature Reviews Genetics* **6**, 544-556.
Bellet MM, Sassone-Corsi P (2010) Mammalian circadian clock and metabolism – the epigenetic link. *Journal of Cell Science* **123**, 3837-3848.
Bendesky A, Bargmann CI (2011) Genetic contributions to behavioural diversity at the gene–environment interface. *Nature Reviews Genetics* **12**, 809-820.
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - A practical and powerful approach to multipe testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300.
Benowitz KM, McKinney EC, Cunningham CB, Moore AJ (2017) Relating quantitative variation within a behavior to variation in transcription. *Evolution* **71**, 1999-2009.
Berson DM, Dunn FA, Takao M (2002) Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070.
Berthold P (1999) A comprehensive theory for the evolution, control and adaptability of avian migration. *Ostrich* **70**, 1-11.
Berthold P (2001) *Bird migration: A general survey* Oxford; New York: Oxford University Press, c2001; 2nd ed.
Berthold P, Fiedler W, Querner U (2000) Die Zugunruhe bei Vögeln — eine Darstellung nach Videoaufnahmen bei Infrarotlichtbeleuchtung. *Journal Fur Ornithologie* **141**, 285-299.
Berthold P, Helbig AJ (1992) The genetics of bird migration: Stimulus, timing, and direction. *Ibis* **134**, 35-40.
Berthold P, Querner U (1981) Genetic basis of migratory behavior in European warblers. *Science* **212**, 77-79.
Biebach H, Falk H, Krebs JR (1991) The effect of constant light and phase shifts on a learned time- place association in garden warblers (*Sylvia borin*): hourglass or circadian clock? *Journal of Biological Rhythms* **6**, 353.

Bolborea M, Dale N (2013) Hypothalamic tanycytes: potential roles in the control of feeding and energy balance. *Trends in Neurosciences* **36**, 91-100.
Boss J, Liedvogel M, Lundberg M, *et al.* (2016) Gene expression in the brain of a migratory songbird during breeding and migration. *Movement Ecology* **4**, 4.
Both C, Bouwhuis S, Lessells CM, Visser ME (2006) Climate change and population declines in a long-distance migratory bird. *Nature* **441**, 81-83.
Both C, Van Turnhout CAM, Bijlsma RG, *et al.* (2010) Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. *Proceedings of the Royal Society B-Biological Sciences* **277**, 1259-1266.
Both C, Visser ME (2001) Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. *Nature* **411**, 296-298.
Bourret A, Garant D (2015) Candidate gene–environment interactions and their relationships with timing of breeding in a wild bird population. *Ecology and Evolution* **5**, 3628-3641.
Bozek K, Relogio A, Kielbasa SM, *et al.* (2009) Regulation of clock-controlled genes in mammals. *Plos One* **4**, e4882.
Bradbury PJ, Zhang Z, Kroon DE, *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.
Bradshaw WE, Holzapfel CM (2008) Genetic response to rapid climate change: it's seasonal timing that matters. *Molecular Ecology* **17**, 157-166.
Brandstatter R (2003) Encoding time of day and time of year by the avian circadian system. *Journal of Neuroendocrinology* **15**, 398-404.
Brandstatter R, Abraham U, Albrecht U (2001a) Initial demonstration of rhythmic Per gene expression in the hypothalamus of a non-mammalian vertebrate, the house sparrow. *Neuroreport* **12**, 1167-1170.
Brandstatter R, Kumar V, Van't Hof TJ, Gwinner E (2001b) Seasonal variations of in vivo and *in vitro* melatonin production in a passeriform bird, the house sparrow (*Passer domesticus*). *Journal of Pineal Research* **31**, 120-126.
Brieuc MSO, Ono K, Drinan DP, Naish KA (2015) Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* **24**, 2729-2746.
Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004-1006, 1008-1010.
Bugge A, Feng D, Everett LJ, *et al.* (2012) Rev-erbα and Rev-erbβ coordinately protect the circadian clock and normal metabolic function. *Genes & Development* **26**, 657-667.
Burris TP (2008) Nuclear hormone receptors for heme: REV-ERBα and REV-ERBβ are ligand-regulated components of the mammalian clock. *Molecular Endocrinology* **22**, 1509-1520.
Buskirk J, Mulvihill RS, Leberman RC (2012) Phenotypic plasticity alone cannot explain climate-induced change in avian migration timing. *Ecology and Evolution* **2**, 2430-2437.
Canal-Corretger MM, Vilaplana J, Cambras T, Diez-Noguera A (2001) Functioning of the rat circadian system is modified by light applied in critical postnatal days. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **280**, R1023-1030.

Caprioli M, Ambrosini R, Boncoraglio G, *et al.* (2012) *Clock* gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *Plos One* **7**, e35140.

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* **5**, 618.

Casagrande S, Dell'Omo G, Costantini D, Tagliavini J (2006) Genetic differences between early- and late-breeding Eurasian kestrels. *Evolutionary Ecology Research* **8**, 1029-1038.

Cassone VM (2014) Avian circadian organization: A chorus of clocks. *Frontiers in Neuroendocrinology* **35**, 76-88.

Cassone VM, Westneat DF (2012) The bird of time: Cognition and the avian biological clock. *Front Mol Neurosci* **5**, 32.

Cermakian N, Whitmore D, Foulkes NS, Sassone-Corsi P (2000) Asynchronous oscillations of two zebrafish CLOCK partners reveal differential clock control and function. *Proceedings of the National Academy of Sciences* **97**, 4339-4344.

Chakarov N, Jonker R, Boerner M, Hoffman J, Kruger O (2013) Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. *Molecular Ecology* **22**, 5430-5440.

Charmantier A, Gienapp P (2014) Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evolutionary Applications* **7**, 15-28.

Charmantier A, McCleery RH, Cole LR, *et al.* (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800-803.

Cheng MY, Bittman EL, Hattar S, Zhou Q-Y (2005) Regulation of prokineticin 2 expression by light and the circadian clock. *BMC Neuroscience* **6**, 17-17.

Cheng MY, Bullock CM, Li C, *et al.* (2002) Prokineticin 2 transmits the behavioural circadian rhythm of the suprachiasmatic nucleus. *Nature* **417**, 405.

Chernetsov N, Berthold P, Querner U (2004) Migratory orientation of first-year white storks (*Ciconia ciconia*): inherited information and social interactions. *Journal of Experimental Biology* **207**, 937-943.

Cho H, Zhao X, Hatori M, *et al.* (2012) Regulation of circadian behaviour and metabolism by REV-ERB-alpha and REV-ERB-beta. *Nature* **485**, 123-127.

Chong NW, Bernard M, Klein DC (2000) Characterization of the chicken serotonin N-acetyltransferase gene - Activation via clock gene heterodimer/e box interaction. *Journal of Biological Chemistry* **275**, 32991-32998.

Chong NW, Chaurasia SS, Haque R, Klein DC, Iuvone PM (2003) Temporal-spatial characterization of chicken clock genes: circadian expression in retina, pineal gland, and peripheral tissues. *Journal of Neurochemistry* **85**, 851-860.

Ciarleglio CM, Axley JC, Strauss BR, Gamble KL, McMahon DG (2011) Perinatal photoperiod imprints the circadian clock. *Nature Neuroscience* **14**, 25-27.

Ciarleglio CM, Gamble KL, Axley JC, *et al.* (2009) Population encoding by circadian clock neurons organizes circadian behavior. *Journal of Neuroscience* **29**, 1670-1676.

Cockrem JF (2013) Corticosterone responses and personality in birds: Individual variation and the ability to cope with environmental changes due to climate change. *General and Comparative Endocrinology* **190**, 156-163.

Colwell CS, Michel S, Itri J, *et al.* (2003) Disrupted circadian rhythms in VIP- and PHI-deficient mice. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **285**, R939-R949.

Conklin JR (2011) *Extreme migration and the annual cycle: Individual strategies in New Zealand bar-tailed godwits*. Publication based, Massey University.
Conklin JR, Battley PF (2011) Impacts of wind on individual migration schedules of New Zealand bar-tailed godwits. *Behavioral Ecology* **22**, 854-861.
Conklin JR, Battley PF (2012) Carry-over effects and compensation: Late arrival on non-breeding grounds affects wing moult but not plumage or schedules of departing bar-tailed godwits *Limosa lapponica baueri*. *Journal of Avian Biology* **43**, 252-263.
Conklin JR, Battley PF, Potter MA (2013) Absolute consistency: Individual versus population variation in annual-cycle schedules of a long-distance migrant bird. *Plos One* **8**, e54535.
Conklin JR, Battley PF, Potter MA, Fox JW (2010) Breeding latitude drives individual schedules in a trans-hemispheric migrant bird. *Nature Communications* **1**, 67.
Conklin JR, Battley PF, Potter MA, Ruthrauff DR (2011) Geographic variation in morphology of Alaskan-breeding bar-tailed godwits (*Limosa lapponica*) is not maintained on their nonbreeding grounds in New Zealand. *Auk* **128**, 363-373.
Conklin JR, Reneerkens J, Verkuil YI, *et al.* (2016) Low genetic differentiation between Greenlandic and Siberian sanderling populations implies a different phylogeographic history than found in red knots. *Journal of Ornithology* **157**, 325-332.
Contina A, Bridge ES, Ross JD, Shipley JR, Kelly JF (2018) Examination of *Clock* and *Adcyap1* gene variation in a neotropical migratory passerine. *Plos One* **13**, e0190859.
Coppack T, Tindemans I, Czisch M, *et al.* (2008) Can long-distance migratory birds adjust to the advancement of spring by shortening migration distance? The response of the pied flycatcher to latitudinal photoperiodic variation. *Global Change Biology* **14**, 2516-2522.
Cornelius JM, Boswell T, Jenni-Eiermann S, Breuner CW, Ramenofsky M (2013) Contributions of endocrinology to the migration life history of birds. *General and Comparative Endocrinology* **190**, 47-60.
Costa R, Peixoto AA, Barbujani G, Kyriacou CP (1992) A latitudinal cline in a Drosophila *Clock* gene. *Proceedings of the Royal Society B-Biological Sciences* **250**, 43-49.
Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *Bmc Bioinformatics* **11**, 485.
Crews D (2011) Epigenetic modifications of brain and behavior: Theory and practice. *Hormones and Behavior* **59**, 393-398.
Csernus V, Jozsa R, Reglodi D, *et al.* (2004) The effect of PACAP on rhythmic melatonin release of avian pineals. *General and Comparative Endocrinology* **135**, 62-69.
Csernus VJ (2006) The avian pineal gland. *Chronobiology International* **23**, 329-339.
D'Urban Jackson J, dos Remedios N, Maher KH, *et al.* (2017) Polygamy slows down population divergence in shorebirds. *Evolution* **71**, 1313-1326.
Daan S (2000) Colin Pittendrigh, Jurgen Aschoff, and the natural entrainment of circadian systems. *Journal of Biological Rhythms* **15**, 195-207.
Daan S, Merrow M, Roenneberg T (2002) External time - Internal time. *Journal of Biological Rhythms* **17**, 107-109.
Danecek P, Auton A, Abecasis G, *et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
Dardente H, Hazlerigg DG, Ebling FJ (2014) Thyroid hormone and seasonal rhythmicity. *Frontiers in Endocrinology* **5**.

Dardente H, Wyse CA, Birnie MJ, *et al.* (2010) A molecular switch for photoperiod responsiveness in mammals. *Current Biology* **20**, 2193-2198.
Darlington TK, Wager-Smith K, Ceriani MF, *et al.* (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science* **280**, 1599-1603.
Dawson A (2002) Photoperiodic control of the annual cycle in birds and comparison with mammals. *Ardea* **90**, 355-367.
Dawson A (2007) Seasonality in a temperate zone bird can be entrained by near equatorial photoperiods. *Proceedings of the Royal Society B: Biological Sciences* **274**, 721.
Dawson A, King VM, Bentley GE, Ball GF (2001) Photoperiodic control of seasonality in birds. *Journal of Biological Rhythms* **16**, 365-380.
De Coursey PJ (1960) Daily light sensitivity rhythm in a rodent. *Science* **131**, 33-35.
de Miera CS, Hanon EA, Dardente H, *et al.* (2013) Circannual variation in thyroid hormone deiodinases in a short-day breeder. *Journal of Neuroendocrinology* **25**, 412-421.
Delmore KE, Huebner S, Kane NC, *et al.* (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology* **24**, 1873-1888.
Delmore KE, Toews DPL, Germain RR, Owens GL, Irwin DE (2016) The genetics of seasonal migration and plumage color. *Current Biology* **26**, 2167-2173.
Dietrich JW, Landgrafe G, Fotiadou EH (2012) TSH and thyrotropic agonists: Key actors in thyroid homeostasis. *Journal of Thyroid Research* **2012**, 29 pages.
Dingemanse NJ, Wolf M (2010) Recent models for adaptive personality differences: a review. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3947-3958.
Dingemanse NJ, Wolf M (2013) Between-individual differences in behavioural plasticity within populations: causes and consequences. *Animal Behaviour* **85**, 1031-1039.
Dominoni DM, Akesson S, Klaassen R, Spoelstra K, Bulla M (2017) Methods in field chronobiology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **372**.
Dor R, Cooper CB, Lovette IJ, *et al.* (2011a) *Clock* gene variation in *Tachycineta* swallows. *Ecology and Evolution* **2**, 95-105.
Dor R, Lovette IJ, Safran RJ, *et al.* (2011b) Low variation in the polymorphic *Clock* gene poly-Q region despite population genetic structure across barn swallow (*Hirundo rustica*) populations. *Plos One* **6**, e28843.
Douris N, Green CB (2008) NOC out the fat: A short review of the circadian deadenylase Nocturnin. *Annals of Medicine* **40**, 622-626.
Dray S, Dufour A-B (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**, 1-20.
Drent R, Both C, Green M, Madsen J, Piersma T (2003) Pay-offs and penalties of competing migratory schedules. *Oikos* **103**, 274-292.
Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* **96**, 271-290.
Dunlap JC, Loros JJ, DeCoursey PJ (2004) *Chronobiology: Biological Timekeeping* Sinauer Associates, Incorporated.
Dupre SM, Miedzinska K, Duval CV, *et al.* (2010) Identification of *eya3* and TAC1 as long-day signals in the sheep pituitary. *Current Biology* **20**, 829-835.
Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Easton A, Arbuzova J, Turek FW (2003) The circadian *Clock* mutation increases exploratory activity and escape-seeking behavior. *Genes Brain and Behavior* **2**, 11-19.
Edwards SV, Dillon M (2004) Hitchhiking and recombination in birds: Evidence from Mhc-linked and unlinked loci in red-winged blackbirds (*Agelaius phoeniceus*). *Genetics Research* **84**, 175-192.
Eikenaar C, Klinner T, Szostek KL, Bairlein F (2014) Migratory restlessness in captive individuals predicts actual departure in the wild. *Biology Letters* **10**, 20140154.
Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1-15.
Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications* **7**, 1026-1042.
Ellegren H, Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics* **8**, 689-698.
Evans JA, Leise TL, Castanon-Cervantes O, Davidson AJ (2015) Neural correlates of individual differences in circadian behaviour. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150769.
Evens R, Beenaerts N, Witters N, Artois T (2017) Repeated migration of a juvenile European nightjar *Caprimulgus europaeus*. *Journal of Ornithology* **158**, 881-886.
Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**, 47-50.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.
Faluhelyi N, Reglodi D, Lengvari I, Csernus V (2004) Development of the circadian melatonin rhythm and the effect of PACAP on melatonin release in the embryonic chicken pineal gland. An in vitro study. *Regulatory Peptides* **123**, 23-28.
Feillet CA, Albrecht U, Challet E (2006) "Feeding time" for the brain: A matter of clocks. *Journal of Physiology-Paris* **100**, 252-260.
Fidler AE, Gwinner E (2003) Comparative analysis of avian BMALI and CLOCK protein sequences: A search for features associated with owl nocturnal behaviour. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **136**, 861-874.
Fidler AE, van Oers K, Drent PJ, *et al.* (2007) *Drd4* gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings of the Royal Society B-Biological Sciences* **274**, 1685-1691.
Fitzpatrick MJ, Ben-Shahar Y, Smid HM, *et al.* (2005) Candidate genes for behavioural ecology. *Trends in Ecology & Evolution* **20**, 96-104.
Foulkes NS, Whitmore D, Sassone-Corsi P (1997) Rhythmic transcription: The molecular basis of circadian melatonin synthesis. *Biology of the Cell* **89**, 487-494.
Franchini P, Irisarri I, Fudickar A, *et al.* (2017) Animal tracking meets migration genomics: Transcriptomic analysis of a partially migratory bird species. *Molecular Ecology* **26**, 3204-3216.
Fridolfsson AK, Ellegren H (1999) A simple and universal method for molecular sexing of non-ratite birds. *Journal of Avian Biology* **30**, 116-121.
Friesen VL, Smith AL, Gomez-Diaz E, *et al.* (2007) Sympatric speciation by allochrony in a seabird. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18589-18594.

Fu Z, Inaba M, Noguchi T, Kato H (2002) Molecular cloning and circadian regulation of cryptochrome genes in Japanese quail (*Coturnix coturnix japonica*). *J Biol Rhythms* **17**, 14-27.

Fuchikawa T, Sanada S, Nishio R, *et al.* (2010) The clock gene cryptochrome of *Bactrocera cucurbitae* (Diptera: Tephritidae) in strains with different mating times. *Heredity* **104**, 387-392.

Fudickar AM, Greives TJ, Atwell JW, Stricker CA, Ketterson ED (2016) Reproductive allochrony in seasonally sympatric populations maintained by differential response to photoperiod: Implications for population divergence and response to climate change. *The American Naturalist* **187**, 436-446.

Gachon F, Nagoshi E, Brown SA, Ripperger J, Schibler U (2004) The mammalian circadian timing system: From gene expression to physiology. *Chromosoma* **113**, 103-112.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv* **1207.3907**.

Gauer F, Craft CM (1996) Circadian regulation of hydroxyindole-O-methyltransferase mRNA levels in rat pineal and retina. *Brain Research* **737**, 99-109.

Gaut BS, Long AD (2003) The lowdown on Linkage Disequilibrium. *The Plant Cell* **15**, 1502-1506.

Gekakis N, Staknis D, Nguyen HB, *et al.* (1998) Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564-1569.

Gerkema MP, Davies WIL, Foster RG, Menaker M, Hut RA (2013) The nocturnal bottleneck and the evolution of activity patterns in mammals. *Proceedings of the Royal Society B-Biological Sciences* **280**, 20130508.

Gienapp P, Laine VN, Mateman AC, van Oers K, Visser ME (2017) Environment-dependent genotype-phenotype associations in avian breeding time. *Frontiers in Genetics* **8**, 102.

Gill RE, Jr., Tibbitts TL, Douglas DC, *et al.* (2009) Extreme endurance flights by landbirds crossing the Pacific Ocean: ecological corridor rather than barrier? *Proceedings of the Royal Society B-Biological Sciences* **276**, 447-458.

Gnerre S, MacCallum I, Przybylski D, *et al.* (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* **108**, 1513-1518.

Goddard KAB, Hopkins PJ, Hall JM, Witte JS (2000) Linkage Disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *The American Journal of Human Genetics* **66**, 216-234.

Goldbeter A (1995) A model for circadian oscillations in the Drosophila period protein (PER). *Proceedings of the Royal Society B-Biological Sciences* **261**, 319-324.

Goymann W, Helm B, Jensen W, Schwabl I, Moore IT (2012) A tropical bird can use the equatorial change in sunrise and sunset times to synchronize its circannual clock. *Proceedings of the Royal Society B-Biological Sciences* **279**, 3527-3534.

Goymann W, Lupi S, Kaiya H, Cardinale M, Fusani L (2017) Ghrelin affects stopover decisions and food intake in a long-distance migrant. *Proceedings of the National Academy of Sciences* **114**, 1946-1951.

Greives TJ, Kingma SA, Beltrami G, Hau M (2012) Melatonin delays clutch initiation in a wild songbird. *Biology Letters* **8**, 330-332.

Gunnarsson TG, Gill JA, Sigurbjornsson T, Sutherland WJ (2004) Pair bonds - Arrival synchrony in migratory birds. *Nature* **431**, 646-646.

Gwinner E (1986) *Circannual rhythms* Heidelberg, Germany: Springer.

Gwinner E (1996) Circadian and circannual programmes in avian migration. *The journal of Experimental Biology* **199**, 39-48.
Gwinner E (2003) Circannual rhythms in birds. *Current Opinion in Neurobiology* **13**, 770-778.
Gwinner E, Brandstatter R (2001) Complex bird clocks. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **356**, 1801-1810.
Gwinner E, Helm B (2003) Circannual and circadian contributions to the timing of avian migration. In: *Avian Migration*, pp. 81-95.
Gwinner E, Helm B, Koolhaas A, *et al.* (2012) Chapter 26 - Avian migration: Temporal multitasking and a case study of melatonin cycles in waders. In: *Progress in Brain Research* (eds. Andries Kalsbeek MMTR, Russell GF), pp. 457-479. Elsevier.
Halford S, Pires SS, Turton M, *et al.* (2009) VA opsin-based photoreceptors in the hypothalamus of birds. *Current Biology* **19**, 1396-1402.
Hallworth MT, Marra PP (2015) Miniaturized GPS tags identify non-breeding territories of a small breeding migratory songbird. *Scientific Reports* **5**, 11069.
Hands S, Sinadinos C, Wyttenbach A (2008) Polyglutamine gene function and dysfunction in the ageing brain. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1779**, 507-521.
Haque R, Ali FG, Biscoglia R, *et al.* (2010) CLOCK and NPAS2 have overlapping roles in the circadian oscillation of arylalkylamine N-acetyltransferase mRNA in chicken cone photoreceptors. *Journal of Neurochemistry* **113**, 1296-1306.
Haque R, Chaurasia SS, Chong NW, Klein DC, Iuvone PM (2003) Circadian expression of *Bmal1* in chicken retina, pineal gland, and peripheral tissues. *Investigative Ophthalmology & Visual Science* **44**, U70-U70.
Hardin PE (2005) The circadian timekeeping system of Drosophila. *Current Biology* **15**, R714-R722.
Hastings MH (2000) Circadian clockwork: Two loops are better than one. *Nature Reviews Neuroscience* **1**, 143-146.
Hau M, Gwinner E (1994) Melatonin facilitates synchronization of sparrow circadian rhythms to light. *Sensory, Neural and Behavioral Physiology* **175**, 343-347.
Hau M, Perfito N, Moore IT (2008) Timing of breeding in tropical birds: Mechanisms and evolutionary implications. *Ornitologia Neotropical* **19**, 39-59.
Hazlerigg DG, Ebling FJP, Johnston JD (2005) Photoperiod differentially regulates gene expression rhythms in the rostral and caudal SCN. *Current Biology* **15**, R449-R450.
Hazlerigg DG, Lincoln GA (2011) Hypothesis: Cyclical histogenesis is the basis of circannual timing. *Journal of Biological Rhythms* **26**, 471-485.
Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8.
Hecht BC, Thrower FP, Hale MC, Miller MR, Nichols KM (2012) Genetic architecture of migration-related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*. *G3: Genes|Genomes|Genetics* **2**, 1113-1127.
Hejjas K, Vas J, Topal J, *et al.* (2007) Association of polymorphisms in the dopamine D4 receptor gene and the activity-impulsivity endophenotype in dogs. *Animal Genetics* **38**, 629-633.
Helfer G, Fidler AE, Vallone D, Foulkes NS, Brandstaetter R (2006) Molecular analysis of clock gene expression in the avian brain. *Chronobiology International* **23**, 113-127.

Helm B, Ben-Shlomo R, Sheriff MJ, *et al.* (2013) Annual rhythms that underlie phenology: Biological time-keeping meets environmental change. *Proceedings of the Royal Society B-Biological Sciences* **280**, 20130016.
Helm B, Gwinner E (2005) Carry-over effects of day length during spring migration. *Journal of Ornithology* **146**, 348-354.
Helm B, Gwinner E (2006) Migratory restlessness in an equatorial nonmigratory bird. *Plos Biology* **4**, 611-614.
Helm B, Gwinner E, Trost L (2005) Flexible seasonal timing and migratory behavior: Results from Stonechat breeding programs. *Annals of the New York Academy of Sciences* **1046**, 216-227.
Helm B, Schwabl I, Gwinner E (2009) Circannual basis of geographically distinct bird schedules. *Journal of Experimental Biology* **212**, 1259-1269.
Helm B, Visser ME, Schwartz W, *et al.* (2017) Two sides of a coin: Ecological and chronobiological perspectives of timing in the wild. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160246.
Helm BR, Davidowitz G (2012) Do resource thresholds play a role in the onset of maturation-a key life history transition? *Integrative and Comparative Biology* **52**, E76-E76.
Hendry AP, Berg OK, Quinn TP (1999) Condition dependence and adaptation-by-time: breeding date, life history, and energy allocation within a population of salmon. *Oikos* **85**, 499-514.
Hendry AP, Day T (2005) Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology* **14**, 901-916.
Herwig A, de Vries EM, Bolborea M, *et al.* (2013) Hypothalamic ventricular ependymal thyroid hormone deiodinases are an important element of circannual timing in the siberian hamster (*Phodopus sungorus*). *Plos One* **8**, e62003.
Herzog ED (2007) Neurons and networks in daily rhythms. *Nat Rev Neurosci* **8**, 790-802.
Hess JE, Zendt JS, Matala AR, Narum SR (2016) Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society B- Biological Sciences* **283**, 20153064.
Hirano A, Braas D, Fu Y-H, Ptáček LJ (2017) FAD regulates CRYPTOCHROME protein stability and circadian clock in mice. *Cell Reports* **19**, 255-266.
Hirota T, Fukada Y (2004) Resetting mechanism of central and peripheral circadian clocks in mammals. *Zoological Science* **21**, 359-368.
Hogenesch JB, Gu YZ, Moran SM, *et al.* (2000) The basic helix-loop-helix-PAS protein MOP9 is a brain-specific heterodimeric partner of circadian and hypoxia factors. *Journal of Neuroscience* **20**, Rc83.
Holberton RL, Able KP (1992) Persistence of circannual cycles in a migratory bird held in constant dim light. *Journal of Comparative Physiology A* **171**, 477-481.
Holliday JA, Wang T, Aitken S (2012) Predicting adaptive phenotypes from multilocus genotypes in sitka spruce (*Picea sitchensis*) using random forest. *G3: Genes|Genomes|Genetics* **2**, 1085.
Holmes C, Smith H, Ganderton R, *et al.* (2001) Psychosis and aggression in Alzheimer's disease: The effect of dopamine receptor gene variation. *J Neurol Neurosurg Psychiatry* **71**, 777-779.
Holmes RT (1971) Latitudinal differences in the breeding and molt schedules of Alaskan red-backed sandpipers (*Calidris alpina*). *The Condor* **73**, 93-99.
Honma S, Kawamoto T, Takagi Y, *et al.* (2002) Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature* **419**, 841-844.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.
Hurley JM, Loros JJ, Dunlap JC (2016) Circadian oscillators: Around the transcription–translation feedback loop and on to output. *Trends in Biochemical Sciences* **41**, 834-846.
Husby A, Kawakami T, Ronnegard L, *et al.* (2015) Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proceedings of the Royal Society B-Biological Sciences* **282**, 20150156.
Hut RA, Beersma DGM (2011) Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. *Philosophical Transactions of the Royal Society B-Biological Sciences* **366**, 2141-2154.
Hut RA, Paolucci S, Dor R, Kyriacou CP, Daan S (2013) Latitudinal clines: An evolutionary view on biological rhythms. *Proceedings of the Royal Society B-Biological Sciences* **280**, 20130433.
Ikegami K, Yoshimura T (2012) Circadian clocks and the measurement of daylength in seasonal reproduction. *Molecular and Cellular Endocrinology* **349**, 76-81.
Jenni L, Kery M (2003) Timing of autumn bird migration under climate change: advances in long-distance migrants, delays in short-distance migrants. *Proceedings of the Royal Society B-Biological Sciences* **270**, 1467-1471.
Johnsen A, Fidler AE, Kuhn S, *et al.* (2007) Avian *Clock* gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Molecular Ecology* **16**, 4867-4880.
Johnston JD, Ebling FJP, Hazlerigg DG (2005) Photoperiod regulates multiple gene expression in the suprachiasmatic nuclei and pars tuberalis of the Siberian hamster (*Phodopus sungorus*). *European Journal of Neuroscience* **21**, 2967-2974.
Johnston RA, Paxton KL, Moore FR, Wayne RK, Smith TB (2016) Seasonal gene expression in a migratory songbird. *Molecular Ecology* **25**, 5680-5691.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.
Jones FC, Grabherr MG, Chan YF, *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.
Jones S, Pfister-Genskow M, Cirelli C, Benca RM (2008) Changes in brain gene expression during migration in the white-crowned sparrow. *Brain Research Bulletin* **76**, 536-544.
Jonzen N, Linden A, Ergon T, *et al.* (2006) Rapid advance of spring arrival dates in long-distance migratory birds. *Science* **312**, 1959-1961.
Kardos M, Husby A, McFarlane SE, Qvarnström A, Ellegren H (2016) Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources* **16**, 727-741.
Kategaya LS, Hilliard A, Zhang L, *et al.* (2012) Casein kinase 1 proteomics reveal prohibitin 2 function in molecular clock. *Plos One* **7**.
Kawakami T, Backström N, Burri R, *et al.* (2014) Estimation of linkage disequilibrium and interspecific gene flow in Ficedula flycatchers by a newly developed 50k single-nucleotide polymorphism array. *Molecular Ecology Resources* **14**, 1248-1260.
Kendall MG (1948) *Rank correlation methods* Griffin, Oxford, England.

Ketterson ED, Fudickar AM, Atwell JW, Greives TJ (2015) Seasonal timing and population divergence: when to breed, when to migrate. *Current Opinion in Behavioral Sciences* **6**, 50-58.
Kim JS, Bailey MJ, Ho AK, *et al.* (2007) Daily rhythm in pineal phosphodiesterase (PDE) activity reflects adrenergic/3',5'-cyclic adenosine 5'-monophosphate induction of the PDE4B2 variant. *Endocrinology* **148**, 1475-1485.
Kissner KJ, Weatherhead PJ, Francis CM (2003) Sexual size dimorphism and timing of spring migration in birds. *Journal of Evolutionary Biology* **16**, 154-162.
Knief U, Schielzeth H, Backstrom N, *et al.* (2017) Association mapping of morphological traits in wild and captive zebra finches: Reliable within, but not between populations. *Molecular Ecology* **26**, 1285-1305.
Knippschild U, Gocht A, Wolff S, *et al.* (2005) The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. *Cell Signal* **17**, 675-689.
Knudsen E, Linden A, Both C, *et al.* (2011) Challenging claims in the study of migratory birds and climate change. *Biological Reviews* **86**, 928-946.
Ko CH, Takahashi JS (2006) Molecular components of the mammalian circadian clock. *Human Molecular Genetics* **15**, R271-R277.
Kölzsch A, Bauer S, de Boer R, *et al.* (2015) Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore. *Journal of Animal Ecology* **84**, 272-283.
Korsten P, Mueller JC, Hermannstadter C, *et al.* (2010) Association between DRD4 gene polymorphism and personality variation in great tits: A test across four wild populations. *Molecular Ecology* **19**, 832-843.
Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **9**, 29.
Kruk M, Noordervliet MAW, Ter Keurs WJ (1998) Natal philopatry in the black-tailed godwit *Limosa limosa L.* and its possible implications for conservation. *Ringing and Migration* **19**, 13-16.
Kuhn K, Schwenk K, Both C, *et al.* (2013) Differentiation in neutral genes and a candidate gene in the pied flycatcher: Using biological archives to track global climate change. *Ecology and Evolution* **3**, 4799-4814.
Kumar V, Singh BP, Rani S (2004) The bird clock: A complex, multi-oscillatory and highly diversified system. *Biological Rhythm Research* **35**, 121-144.
Kumar V, Wingfield JC, Dawson A, *et al.* (2010) Biological clocks and regulation of seasonal reproduction and migration in birds. *Physiological and Biochemical Zoology* **83**, 827-835.
Kwon JM, Goate AM (2000) The candidate gene approach. *Alcohol Res Health* **24**, 164-168.
Lam V, Li Y, Murphy K, Kwok R, Chiu J (2015) Casein kinase 1 alpha (CK1a) promotes nuclear entry of PERIOD and represses CLOCK transcriptional activity in the *Drosophila melanogaster* circadian clock. *The FASEB Journal* **29**.
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359.
Laurie CC, Doheny KF, Mirel DB, *et al.* (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology* **34**, 591-602.
Lawrence E (2005) *Henderson's dictionary of biology.* Pearson education.
Le Rouzic A, Álvarez-Castro JM, Carlborg Ö (2008) Dissection of the genetic architecture of body weight in chicken reveals the impact of epistasis on domestication traits. *Genetics* **179**, 1591-1599.

Leder EH, Danzmann RG, Ferguson MM (2006) The candidate gene, *Clock*, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity* **97**, 74-80.
Ledon-Rettig CC, Richards CL, Martin LB (2013) Epigenetics for behavioral ecologists. *Behavioral Ecology* **24**, 311-324.
Lee HM, Chen R, Kim H, *et al.* (2011) The period of the circadian oscillator is primarily determined by the balance between casein kinase 1 and protein phosphatase 1. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16451-16456.
Lee W-P, Stromberg MP, Ward A, *et al.* (2014) MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. *Plos One* **9**, e90581.
Lehtonen PK, Laaksonen T, Artemyev AV, *et al.* (2012) Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity* **108**, 431-440.
Lewis RD, Warman GR, Saunders DS (1997) Simulations of free-running rhythms, light entrainment and the light-pulse phase response curves for the locomotor activity rhythm in period mutant of *Drosophila melanogaster*. *Journal of Theoretical Biology* **185**, 503-510.
Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843-2851.
Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
Li JD, Hu WP, Boehmer L, *et al.* (2006) Attenuated circadian rhythms in mice lacking the prokineticin 2 gene. *Journal of Neuroscience* **26**, 11615-11623.
Li MH, Merila J (2010) Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**, 600-610.
Li R, Yue J, Zhang Y, *et al.* (2008) CLOCK/BMAL1 regulates human nocturnin transcription through binding to the E-box of nocturnin promoter. *Mol Cell Biochem* **317**, 169-177.
Liedvogel M, Akesson S, Bensch S (2011) The genetics of migration on the move. *Trends in Ecology & Evolution* **26**, 561-569.
Liedvogel M, Cornwallis CK, Sheldon BC (2012) Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *Journal of Evolutionary Biology* **25**, 813-823.
Liedvogel M, Lundberg M (2014) The genetics of migration. In: *Animal movements across scales* (eds. Hansson LA, Akesson S), pp. 219-231. Oxford University Press, Oxford.
Liedvogel M, Sheldon BC (2010) Low variability and absence of phenotypic correlates of *Clock* gene variation in a great tit *Parus major* population. *Journal of Avian Biology* **41**, 543-550.
Liedvogel M, Szulkin M, Knowles SCL, Wood MJ, Sheldon BC (2009) Phenotypic correlates of *Clock* gene variation in a wild blue tit population: Evidence for a role in seasonal timing of reproduction. *Molecular Ecology* **18**, 2444-2456.
Lipka A, Tian F, Wang Q, *et al.* (2012) GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **28**, 2397-2399.
Loh DH, Dragich JM, Kudo T, *et al.* (2011) Effects of vasoactive intestinal peptide genotype on circadian gene expression in the suprachiasmatic nucleus and peripheral organs. *Journal of Biological Rhythms* **26**, 200-209.

Lowrey PL, Shimomura K, Antoch MP, *et al.* (2000) Positional syntenic cloning and functional characterization of the mammalian circadian mutation *tau*. *Science* **288**, 483-492.
Lugo Ramos JS, Delmore KE, Liedvogel M (2017) Candidate genes for migration do not distinguish migratory and non-migratory birds. *Journal of Comparative Physiology A* **203**, 383-397.
Lundberg M, Liedvogel M, Larson K, *et al.* (2017) Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters* **1**, 155-168.
Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**, 565.
Majumdar G, Trivedi AK, Gupta NJ, Kumar V (2015) Circadian synchronization determines critical day length for seasonal responses. *Physiology & Behavior* **147**, 282-290.
Malek ZS, Pévet P, Raison S (2004) Circadian change in tryptophan hydroxylase protein levels within the rat intergeniculate leaflets and raphe nuclei. *Neuroscience* **125**, 749-758.
Mandilaras K, Missirlis F (2012) Genes for iron metabolism influence circadian rhythms in *Drosophila melanogaster*. *Metallomics* **4**, 928-936.
Manichaikul A, Mychaleckyj JC, Rich SS, *et al.* (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873.
Manolio TA, Collins FS, Cox NJ, *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747-753.
Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**, 209-220.
Marra PP, Francis CM, Mulvihill RS, Moore FR (2005) The influence of climate on the timing and rate of spring bird migration. *Oecologia* **142**, 307-315.
Matschiner M, Salzburger W (2009) TANDEM: Integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**, 1982-1983.
Maywood ES, O'Neill JS, Chesham JE, Hastings MH (2007) Minireview: The circadian clockwork of the Suprachiasmatic nuclei - Analysis of a cellular oscillator that drives endocrine rhythms. *Endocrinology* **148**, 5624-5634.
Mettke-Hofmann C, Gwinner E (2003) Long-term memory for a life on the move. *Proceedings of the National Academy of Sciences* **100**, 5863-5866.
Mettler R, Segelbacher G, Schaefer HM (2015) Interactions between a candidate gene for migration (*adcyap1*), morphology and sex predict spring arrival in blackcap populations. *Plos One* **10**, e0144587.
Miyatake T (2002) Pleiotropic effect, clock genes, and reproductive isolation. *Population Ecology* **44**, 201-207.
Møller AP (2001) Heritability of arrival date in a migratory bird. *Proceedings of the Royal Society B-Biological Sciences* **268**, 203-206.
Møller AP, Antonov A, Stokke BG, *et al.* (2011) Isolation by time and habitat and coexistence of distinct host races of the common cuckoo. *Journal of Evolutionary Biology* **24**, 676-684.
Moore JP, Burger LL, Dalkin AC, Winters SJ (2005) Pituitary adenylate cyclase activating polypeptide messenger RNA in the paraventricular nucleus and anterior pituitary during the rat estrous cycle. *Biology of Reproduction* **73**, 491-499.
Mueller F, Taylor PD, Sjoberg S, *et al.* (2016) Towards a conceptual framework for explaining variation in nocturnal departure time of songbird migrants. *Movement Ecology* **4**, 1-12.

Mueller JC, Partecke J, Hatchwell BJ, Gaston KJ, Evans KL (2013a) Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Molecular Ecology* **22**, 3629-3637.

Mueller JC, Pulido F, Kempenaers B (2011) Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B-Biological Sciences* **278**, 2848-2856.

Mueller T, O'Hara RB, Converse SJ, Urbanek RP, Fagan WF (2013b) Social learning of migratory performance. *Science* **341**, 999-1002.

Murphy KM, Cooper A, Tobias ES (2014) Chapter 3 - The human genome, gene regulation, and genomic variation A2 - Padmanabhan, Sandosh. In: *Handbook of Pharmacogenomics and Stratified Medicine*, pp. 41-56. Academic Press, San Diego.

Nagy AD, Csernus VJ (2007) The role of PACAP in the control of circadian expression of clock genes in the chicken pineal gland. *Peptides* **28**, 1767-1774.

Nakahara K, Abe Y, Murakami T, Shiota K, Murakami N (2002) Pituitary adenylate cyclase-activating polypeptide (PACAP) is involved in melatonin release via the specific receptor PACAP-r1, but not in the circadian oscillator, in chick pineal cells. *Brain Research* **939**, 19-25.

Nakao N, Ono H, Yamamura T, *et al.* (2008) Thyrotrophin in the pars tuberalis triggers photoperiodic response. *Nature* **452**, 317-323.

Nakashima A, Kawamoto T, Honda KK, *et al.* (2008) DEC1 modulates the circadian phase of clock gene expression. *Molecular and Cellular Biology* **28**, 4080-4092.

Nakashima H, Onai K (1996) The circadian conidiation rhythm in *Neurospora crassa*. *Seminars in Cell & Developmental Biology* **7**, 765-774.

Narum SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics* **7**, 783-787.

Nisimura T, Numata H (2001) Endogenous timing mechanism controlling the circannual pupation rhythm of the varied carpet beetle *Anthrenus verbasci*. *Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology* **187**, 433-440.

Nowak JZ, Zawilska JB (2003) PACAP in avians: Origin, occurrence, and receptors--pharmacological and functional considerations. *Curr Pharm Des* **9**, 467-481.

Nussey DH, Wilson AJ, Brommer JE (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology* **20**, 831-844.

O'Malley KG, Banks MA (2008) A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) Clock gene: evidence for selection on PolyQ length variants. *Proceedings of the Royal Society B-Biological Sciences* **275**, 2813-2821.

Okano T, Fukada Y (2001) Photoreception and circadian clock system of the chicken pineal gland. *Microscopy Research and Technique* **53**, 72-80.

Panda S, Hogenesch JB, Kay SA (2002) Circadian rhythms from flies to human. *Nature* **417**, 329-335.

Peakall R, Smouse PE (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537-2539.

Peek CB, Ramsey KM, Marcheva B, Bass J (2012) Nutrient sensing and the circadian clock. *Trends in Endocrinology and Metabolism* **23**, 312-318.

Peeters K, Gerets HHJ, Princen K, Vandesande F (1999) Molecular cloning and expression of a chicken pituitary adenylate cyclase-activating polypeptide receptor. *Molecular Brain Research* **71**, 244-255.

Pegoraro M, Gesto JS, Kyriacou CP, Tauber E (2014) Role for circadian clock genes in seasonal timing: Testing the Bünning hypothesis. *Plos Genetics* **10**, e1004603.
Perneger TV (1998) What's wrong with Bonferroni adjustments. *BMJ : British Medical Journal* **316**, 1236-1238.
Peterson MP, Abolins-Abols M, Atwell JW, *et al.* (2015) Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research* **2**, 2-115.
Pfenning AR, Hara E, Whitney O, *et al.* (2014) Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**.
Pigllucci M (1996) How organisms respond to environmental changes: From phenotypes to molecules (and vice versa). *Trends in Ecology & Evolution* **11**, 168-173.
Plano SA, Agostino PV, de la Iglesia HO, Golombek DA (2012) cGMP-Phosphodiesterase inhibition enhances photic responses and synchronization of the biological circadian clock in rodents. *Plos One* **7**, e37121.
Poelstra JW, Ellegren H, Wolf JBW (2013) An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity* **111**, 467-473.
Porlier M, Charmantier A, Bourgault P, *et al.* (2012) Variation in phenotypic plasticity and selection patterns in blue tit breeding time: between- and within-population comparisons. *Journal of Animal Ecology* **81**, 1041-1051.
Powledge TM (2011) Behavioral epigenetics: How nurture shapes nature. *Bioscience* **61**, 588-592.
Pravosudov VV, Kitaysky AS, Omanska A (2006) The relationship between migratory behaviour, memory and the hippocampus: An intraspecific comparison. *Proceedings of the Royal Society B-Biological Sciences* **273**, 2641-2649.
Preitner N, Damiola F, Lopez-Molina L, *et al.* (2002) The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**, 251-260.
Prendergast BJ, Nelson RJ, Zucker I (2002) 19 - Mammalian Seasonal Rhythms: Behavior and Neuroendocrine Substrates A2 - Pfaff, Donald W. In: *Hormones, Brain and Behavior* (eds. Arnold AP, Fahrbach SE, Etgen AM, Rubin RT), pp. 93-156. Academic Press, San Diego.
Price AL, Patterson NJ, Plenge RM, *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904.
Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nature reviews. Genetics* **11**, 459-463.
Price JL (1997) Insights into the molecular mechanisms of temperature compensation from the Drosophila period and timeless mutants. *Chronobiology International* **14**, 455-468.
Puechmaille SJ (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* **16**, 608-627.
Pulido F (2007a) The genetics and evolution of avian migration. *Bioscience* **57**, 165-174.
Pulido F (2007b) Phenotypic changes in spring arrival: Evolution, phenotypic plasticity, effects of weather and condition. *Climate Research* **35**, 5-23.
Pulido F, Berthold P (2003) Quantitative genetic analysis of migratory behaviour. In: *Avian Migration*, pp. 53-77.

Pulido F, Berthold P, Mohr G, Querner U (2001) Heritability of the timing of autumn migration in a natural bird population. *Proceedings of the Royal Society B-Biological Sciences* **268**, 953-959.
Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.
Qin H, Morris N, Kang SJ, *et al.* (2010) Interrogating local population structure for fine mapping in genome-wide association studies. *Bioinformatics* **26**, 2961-2968.
R Core Team (2014) R: A language and evironment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., URL http://www.R-project.org/.
Racz B, Horvath G, Faluhelyi N, *et al.* (2008) Effects of PACAP on the circadian changes of signaling pathways in chicken pinealocytes. *Journal of Molecular Neuroscience* **36**, 220-226.
Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**, 573-589.
Rajaratnam SMW, Redman JR (1998) Entrainment of activity rhythms to temperature cycles in diurnal palm squirrels. *Physiology & Behavior* **63**, 271-277.
Ramenofsky M, Agatsuma R, Barga M, *et al.* (2003) Migratory behavior: New insights from captive studies. In: *Avian Migration* (eds. Berthold P, Gwinner E, Sonnenschein E), pp. 97-111. Springer Berlin Heidelberg, Berlin, Heidelberg.
Rani S, Kumar V (2013) Avian circannual systems: Persistence and sex differences. *General and Comparative Endocrinology* **190**, 61-67.
Réale D, McAdam AG, Boutin S, Berteaux D (2003) Genetic and plastic responses of a northern mammal to climate change. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 591-596.
Reich DE, Cargill M, Bolk S, *et al.* (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199-204.
Reischl S, Kramer A (2011) Kinases and phosphatases in the mammalian circadian clock. *FEBS Letters* **585**, 1393-1399.
Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. *Nature* **418**, 935-941.
Resuehr D, Wildemann U, Sikes H, Olcese J (2007) E-box regulation of gonadotropin-releasing hormone (GnRH) receptor expression in immortalized gonadotrope cells. *Molecular and Cellular Endocrinology* **278**, 36-43.
Rey G, Cesbron F, Rougemont J, *et al.* (2011) Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol* **9**, e1000595.
Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.
Ripperger JA, Shearman LP, Reppert SM, Schibler U (2000) CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. *Genes & Development* **14**, 679-689.
Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* **405**, 847-856.
Rolshausen G, Segelbacher G, Hobson KA, Schaefer HM (2009) Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. *Current Biology* **19**, 2097-2101.
Romijn HJ, Sluiter AA, Pool CW, Wortel J, Buijs RM (1996) Differences in colocalization between Fos and PHI, GRP, VIP and VP in neurons of the rat

suprachiasmatic nucleus after a light stimulus during the phase delay versus the phase advance period of the night. *Journal of Comparative Neurology* **372**, 1-8.
Rönkä A, Kvist L, Karvonen J, *et al.* (2008) Population genetic structure in the Temminck's stint *Calidris temminckii*, with an emphasis on Fennoscandian populations. *Conservation Genetics* **9**, 29-37.
Rönnegård L, McFarlane SE, Husby A, *et al.* (2016) Increasing the power of genome wide association studies in natural populations using repeated measures – evaluation and implementation. *Methods in Ecology and Evolution* **7**, 792-799.
Rosenberg NA (2004) distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.
Rowan W (1932) Experiments in bird migration. *Zoology* **18**, 639- 654.
Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**, 4757-4769.
Rund SSC, Lee SJ, Bush BR, Duffield GE (2012) Strain- and sex-specific differences in daily flight activity and the circadian clock of *Anopheles gambiae* mosquitoes. *Journal of Insect Physiology* **58**, 1609-1619.
Saino N, Ambrosini R, Albetti B, *et al.* (2017) Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports* **7**, 45412.
Saino N, Bazzi G, Gatti E, *et al.* (2015) Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology* **24**, 1758-1773.
Saino N, Romano M, Caprioli M, *et al.* (2013) Timing of molt of barn swallows is delayed in a rare *Clock* genotype. *Peerj* **1**, e17.
Saleem Q, Anand A, Jain S, Brahmachari SK (2001) The polyglutamine motif is highly conserved at the Clock locus in various organisms and is not polymorphic in humans. *Human Genetics* **109**, 136-142.
Sanada K, Harada Y, Sakai M, Todo T, Fukada Y (2004) Serine phosphorylation of mCRY1 and mCRY2 by mitogen-activated protein kinase. *Genes to Cells* **9**, 697-708.
Santure AW, De Cauwer I, Robinson MR, *et al.* (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Molecular Ecology* **22**, 3949-3962.
Santure AW, Poissant J, De Cauwer I, *et al.* (2015) Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology* **24**, 6148-6162.
Sawyer SL, Mukherjee N, Pakstis AJ, *et al.* (2005) Linkage disequilibrium patterns vary substantially among populations. *European Journal Of Human Genetics* **13**, 677.
Schaap J, Albus H, Vanderleest HT, *et al.* (2003) Heterogeneity of rhythmic suprachiasmatic nucleus neurons: Implications, for circadian waveform and photoperiodic encoding. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15994-15999.
Scheiermann C, Kunisaki Y, Frenette PS (2013) Circadian control of the immune system. *Nature Reviews Immunology* **13**, 190-198.
Scheiner SM (1993) Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics* **24**, 35-68.
Schmaljohann H, Kämpfer S, Fritzsch A, Kima R, Eikenaar C (2015) Start of nocturnal migratory restlessness in captive birds predicts nocturnal departure time in free-flying birds. *Behavioral Ecology and Sociobiology* **69**, 909-914.

Schoenhard JA, Smith LH, Painter CA, *et al.* (2003) Regulation of the PAI-1 promoter by circadian clock components: differential activation by BMAL1 and BMAL2. *Journal of Molecular and Cellular Cardiology* **35**, 473-481.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Schwartz WJ, Helm B, Gerkema MP (2017) Wild clocks: preface and glossary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20170211.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69**, 82-90.

Sham PC, Purcell SM (2014) Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics* **15**, 335.

Shinomiya A, Shimmura T, Nishiwaki-Ohkawa T, Yoshimura T (2014) Regulation of seasonal reproduction by hypothalamic activation of thyroid hormone. *Frontiers in Endocrinology* **5**, 12.

Silva CNS, McFarlane SE, Hagen IJ, *et al.* (2017) Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity* **119**, 197-205.

Simonneaux V, Ancel C, Gauer F, Poirel V-J (2013) Kisspeptins and RFRP-3 act in concert to synchronize rodent reproduction with seasons. *Frontiers in Neuroscience* **7**, 22.

Slate J, Santure AW, Feulner PGD, *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics* **26**, 275-284.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.

Smith PA, Gilchrist HG, Forbes MR, Martin J-L, Allard K (2010) Inter-annual variation in the breeding chronology of arctic shorebirds: effects of weather, snow melt and predators. *Journal of Avian Biology* **41**, 292-304.

Spottiswoode C, Saino N (2010) Sexual selection and climate change. *Oxford University Press*.

Stapley J, Birkhead TR, Burke T, Slate J (2010a) Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Res* **20**, 496-502.

Stapley J, Reger J, Feulner PGD, *et al.* (2010b) Adaptation genomics: The next generation. *Trends in Ecology & Evolution* **25**, 705-712.

Steinmeyer C, Kempenaers B, Mueller JC (2012) Testing for associations between candidate genes for circadian rhythms and individual variation in sleep behaviour in blue tits. *Genetica* **140**, 219-228.

Steinmeyer C, Mueller JC, Kempenaers B (2009) Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits. *Genetica* **136**, 109-117.

Stephan FK (2002) The "other" circadian system: Food as a Zeitgeber. *Journal of Biological Rhythms* **17**, 284-292.

Storz SS, Tovin A, Mracek P, *et al.* (2013) Casein kinase 1 delta activity: A key element in the Zebrafish circadian timing system. *Plos One* **8**, e54189.

Stuber EF, Baumgartner C, Dingemanse NJ, Kempenaers B, Mueller JC (2016) Genetic correlates of individual differences in sleep behavior of free-living great tits (*Parus major*). *G3: Genes|Genomes|Genetics* **6**, 599-607.

Stuber EF, Verpeut J, Horvat-Gordon M, Ramachandran R, Bartell PA (2013) Differential regulation of adipokines may influence migratory behavior in the white-throated sparrow (*Zonotrichia albicollis*). *Plos One* **8**, e59097.

Styrsky JD, Berthold P, Robinson WD (2004) Endogenous control of migration and calendar effects in an intratropical migrant, the yellow-green vireo. *Animal Behaviour* **67**, 1141-1149.
Sumova A, Jac M, Sladek M, Sauman I, Illnerova H (2003) Clock gene daily profiles and their phase relationship in the rat suprachiasmatic nucleus are affected by photoperiod. *Journal of Biological Rhythms* **18**, 134-144.
Sutton JT, Robertson BC, Jamieson IG (2011) Dye shift: A neglected source of genotyping error in molecular ecology. *Molecular Ecology Resources* **11**, 514-520.
Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews Genetics* **3**, 391.
Tarka M, Hansson B, Hasselquist D (2015) Selection and evolutionary potential of spring arrival phenology in males and females of a migratory songbird. *Journal of Evolutionary Biology* **28**, 1024-1038.
Tauber E, Roe H, Costa R, Hennessy JM, Kyriacou CP (2003) Temporal mating isolation driven by a behavioral gene in Drosophila. *Current Biology* **13**, 140-145.
Taylor RS, Friesen VL (2017) The role of allochrony in speciation. *Molecular Ecology* **26**, 3330-3342.
Teplitsky C, Mills JA, Alho JS, Yarrall JW, Merilä J (2008) Bergmann's rule and climate change revisited: Disentangling environmental and genetic responses in a wild bird population. *Proceedings of the National Academy of Sciences* **105**, 13492-13496.
Terada A, Yamada R, Tsuda K, Sese J (2016) LAMPLINK: Detection of statistically significant SNP combinations from GWAS data. *Bioinformatics* **32**, 3513-3515.
Thibert-Plante X, Hendry AP (2011) The consequences of phenotypic plasticity for ecological speciation. *Journal of Evolutionary Biology* **24**, 326-342.
Thorup K, Bisson I-A, Bowlin MS, *et al.* (2007) Evidence for a navigational map stretching across the continental U.S. in a migratory songbird. *Proceedings of the National Academy of Sciences* **104**, 18115-18119.
Toller GL, Nagy E, Horvath RA, Klausz B, Rekasi Z (2006) Circadian expression of *Bmal1* and serotonin-N-acetyltransferase mRNAs in chicken retina cells and pinealocytes *in vivo* and *in vitro*. *Journal of Molecular Neuroscience* **28**, 143-150.
Tøttrup AP, Rainio K, Coppack T, *et al.* (2010) Local temperature fine-tunes the timing of spring migration in birds. *Integrative and Comparative Biology* **50**, 293-304.
Triqueneaux G, Thenot S, Kakizawa T, *et al.* (2004) The orphan receptor Rev-erbα gene is a target of the circadian clock pacemaker. *Journal of Molecular Endocrinology* **33**, 585-608.
Tschirren B, Bensch S (2010) Genetics of personalities: No simple answers for complex traits. *Molecular Ecology* **19**, 624-626.
Ueshima T, Kawamoto T, Honda KK, *et al.* (2012) Identification of a new clock-related element EL-box involved in circadian regulation by BMAL1/CLOCK and HES1. *Gene* **510**, 118-125.
Usui T, Butchart SHM, Phillimore AB (2017) Temporal shifts and temperature sensitivity of avian spring migratory phenology: A phylogenetic meta-analysis. *Journal of Animal Ecology* **86**, 250-261.
Våge J, Wade C, Biagi T, *et al.* (2010) Association of dopamine- and serotonin-related genes with canine aggression. *Genes, Brain and Behavior* **9**, 372-378.
van der Veen DR, Riede SJ, Heideman PD, *et al.* (2017) Flexible clock systems: Adjusting the temporal programme. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **372**, 20160254.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418-426.
van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes* **6**, 255-256.
VanRaden PM (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414-4423.
Verboven N, Visser ME (1998) Seasonal variation in local recruitment of great tits: The importance of being early. *Oikos* **81**, 511-524.
Verkuil YI, Trimbos K, Haddrath O, Baker AJ, Piersma T (2009) Characterization of polymorphic microsatellite DNA markers in the black-tailed godwit (*Limosa limosa*: Aves). *Molecular Ecology Resources* **9**, 1415-1418.
Virshup DM, Forger DB (2007) After hours keeps clock researchers CRYing overtime. *Cell* **129**, 857-859.
Visscher PM (2008) Sizing up human height variation. *Nature Genetics* **40**, 489.
Visser ME, Caro SP, van Oers K, Schaper SV, Helm B (2010) Phenology, seasonal timing and circannual rhythms: Towards a unified framework. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 3113-3127.
Visser ME, Gienapp P, Husby A, *et al.* (2015) Effects of spring temperatures on the strength of selection on timing of reproduction in a long-distance migratory bird. *Plos Biology* **13**, e1002120-e1002120.
Vitalini MW, de Paula RM, Goldsmith CS, *et al.* (2007) Circadian rhythmicity mediated by temporal regulation of the activity of p38 MAPK. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18223-18228.
Voorman A, Lumley T, McKnight B, Rice K (2011) Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *Plos One* **6**, e19416.
Wang J (2017) The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources* **17**, 981-990.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
Wetzel DP, Hatch MI, Westneat DF (2015) Genetic sources of individual variation in parental care behavior. *Behavioral Ecology and Sociobiology* **69**, 1933-1943.
Whitfield CW, Cziko A-M, Robinson GE (2003) Gene expression profiles in the brain predict behavior in individual honey bees. *Science* **302**, 296-299.
Wikelski M, Martin LB, Scheuerlein A, *et al.* (2008) Avian circannual clocks: Adaptive significance and possible involvement of energy turnover in their proximate control. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 411-423.
Williams CT, Klaassen M, Barnes BM, *et al.* (2017) Seasonal reproductive tactics: Annual timing and the capital-to-income breeder continuum. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160250.
Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by Fst do not necessarily require large sample sizes when using many SNP markers. *Plos One* **7**, e42649.
Wood S, Loudon A (2014) Clocks for all seasons: Unwinding the roles and mechanisms of circadian and interval timers in the hypothalamus and pituitary. *Journal of Endocrinology* **222**, R39-59.
Wyse C, Hazlerigg D (2009) Seasonal biology: Avian photoreception goes deep. *Current Biology* **19**, R685-R687.

Xu Y, Padiath QS, Shapiro RE, *et al.* (2005) Functional consequences of a CKI delta mutation causing familial advanced sleep phase syndrome. *Nature* **434**, 640-644.
Yang F, Inoue I, Kumagai M, *et al.* (2013) Real-time analysis of the circadian oscillation of the Rev-Erb beta promoter. *J Atheroscler Thromb* **20**, 267-276.
Yang JA, Benyamin B, McEvoy BP, *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-U131.
Yasuo S, Nakao N, Ohkura S, *et al.* (2006) Long-day suppressed expression of type 2 deiodinase gene in the mediobasal hypothalamus of the saanen goat, a short-day breeder: Implication for seasonal window of thyroid hormone action on reproductive neuroendocrine. *Endocrinology* **147**, 432-440.
Yasuo S, Watanabe M, Nakao N, *et al.* (2005) The reciprocal switching of two thyroid hormone-activating and -inactivating enzyme genes is involved in the photoperiodic gonadal response of Japanese quail. *Endocrinology* **146**, 2551-2554.
Yasuo S, Watanabe M, Okabayashi N, Ebihara S, Yoshimura T (2003) Circadian clock genes and photoperiodism: Comprehensive analysis of clock gene expression in the mediobasal hypothalamus, the Suprachiasmatic nucleus, and the pineal gland of Japanese quail under various light schedules. *Endocrinology* **144**, 3742-3748.
Yoo S-H, Mohawk Jennifer A, Siepka Sandra M, *et al.* (2013) Competing E3 ubiquitin ligases govern circadian periodicity by degradation of CRY in nucleus and cytoplasm. *Cell* **152**, 1091-1105.
Yoshimura T (2010) Neuroendocrine mechanism of seasonal reproduction in birds and mammals. *Animal Science Journal* **81**, 403-410.
Yoshimura T (2013) Thyroid hormone and seasonal regulation of reproduction. *Frontiers in Neuroendocrinology* **34**, 157-166.
Yoshimura T, Suzuki Y, Makino E, *et al.* (2000) Molecular analysis of avian circadian clock genes. *Molecular Brain Research* **78**, 207-215.
Yoshitane H, Fukada Y (2009) CIPC-dependent phosphorylation of CLOCK and NPAS2 in circadian clockwork. *Sleep and Biological Rhythms* **7**, 226-234.
Yu J, Pressoir G, Briggs W, *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**, 203-208.
Zeman M, Herichova I (2011) Circadian melatonin production develops faster in birds than in mammals. *General and Comparative Endocrinology* **172**, 23-30.
Zhan S, Zhang W, Niitepold K, *et al.* (2014) The genetics of monarch butterfly migration and warning colouration. *Nature* **514**, 317-321.
Zhang EE, Kay SA (2010) Clocks not winding down: Unravelling circadian networks. *Nature Reviews Molecular Cell Biology* **11**, 764-776.
Zhang G, Jarvis ED, Gilbert MTP (2014) A flock of genomes. *Science* **346**, 1308.
Zhang L, Lin Q, Ding H, Wang D, Cai Y (2012) Promoter methylation of seven clock genes in total leukocytes from different-aged healthy subjects. *Biological Rhythm Research* **43**, 631-637.
Zhang S, Xu X, Wang W, Yang W, Liang W (2017) *Clock* gene is associated with individual variation in the activation of reproductive endocrine and behavior of Asian short toed lark. *Scientific Reports* **7**, 15002.
Zhao N, Han JG, Shyu C-R, Korkin D (2014) Determining effects of non-synonymous SNPs on protein-protein interactions using supervised and semi-supervised learning. *PLOS Computational Biology* **10**, e1003592.
Zheng X, Sowcik M, Chen D, Sehgal A (2014) Casein kinase 1 promotes synchrony of the circadian clock network. *Molecular and Cellular Biology* **34**, 2682-2694.

Zhou X, Carbonetto P, Stephens M (2013) Polygenic modeling with bayesian sparse linear mixed models. *Plos Genetics* **9**, e1003264.
Zhou X, Stephens M (2012) Genome-wide efficient mixed model analysis for association studies. *Nature Genetics* **44**, 821-824.
Zhu M, Zhao S (2007) Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences* **3**, 420-427.
Zhu X, Zhang S, Zhao H, Cooper RS (2002) Association mapping, using a mixture model for complex traits. *Genetic Epidemiology* **23**, 181-196.

Calendar drawn in 2014 for fun and inspired by my new life in New Zealand and its biodiversity, with a special representation of the New Zealand godwit (or kuaka).