Australian
National
University

Unravelling the mystery of migratory behaviour in the Bogong moth *Agrotis infusa* using genomics and novel automated monitoring techniques.

Jesse Rudolf Amenuvegbe Wallace

A thesis presented for the degree of
Doctor of Philosophy

Supervised by:
Prof. Eric Warrant, Prof. Jochen Zeil, Prof. Ryszard Maleszka, Prof. Saul Cunningham

Research School of Biology, The Australian National University

This thesis is an account of research undertaken between June 2017 and March 2022 at the Research School of Biology, the Australian National University, Canberra, Australia.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Jesse Rudolf Amenuvegbe Wallace

March 2022

Abstract

An exceptionally impressive example of animal navigation is presented by the Bogong moth *Agrotis infusa*, that migrates over 1000 km from widely distributed winter breeding grounds to a relatively confined summer range in the Australian Alps, consistently arriving to the same sites as its predecessors, despite never having an opportunity to learn the migratory route, or indeed, the location of its destination. The Bogong moth then waits out the summer in a dormant state known as aestivation, lining the walls of cool cracks and crevices in high altitude granite outcrops, where it forms massive assemblages with an estimated 17000 moths per square metre. Recent and ongoing investigations into the sensory and neurological capabilities of the Bogong moth have revealed that it possesses a "compass sense" that relies on geomagnetic and stellar information. However, since the migratory direction of the Bogong moth varies across its breeding range, a compass is not sufficient on its own for the moth's navigation. How, for instance, does a Bogong moth know—given its starting location—in which direction to migrate? The objective of this thesis is to understand the basis of the Bogong moth migratory direction. Even though this thesis opens as many questions as it answers, significant progress towards achieving this objective is presented (in two parts) herein, primarily through development of the scientific infrastructure for studying Bogong moth biology more generally. Part I introduces a new method for quantitatively measuring Bogong moth activity and abundance using automated camera-based detection, which is then used to model the influence of abiotic factors on Bogong moth behaviour, and to measure the arrival, departure, and population dynamics of the moths in their summer range. In addition to its utility in addressing ethological questions, this new method enables quantitative long-term monitoring of the Bogong moth population, which may prove invaluable for conservation efforts (the Bogong moth has recently been assessed as endangered for the IUCN Red List). In Part II, the annotated sequence of the Bogong moth genome is presented, opening the door to high-throughput molecular research on the moth. Extensive differential gene expression in the sensory and brain tissue of migrating and aestivating moths is observed, along with evidence of epigenomic modification. Finally, the results of re-sequencing the genomes of 77 Bogong moths collected from across their breeding and summer ranges are presented, which show that the Bogong moth population is panmictic, and harbours a vast quantity of rare genetic variants. Interestingly, a small number of variants are highly correlated with migratory direction, indicating promising avenues for further research into the genetic basis of migratory direction.

Acknowledgements

I am grateful for the support of an Australian Government Research Training Program Scholarship, a COVID-19 Extension Scholarship from the ANU, and a Postgraduate Research Scholarship, also from the ANU. I am also grateful for funding from the Nilsson-Ehle Endowments of The Royal Physiographic Society of Lund towards the costs of the sequencing in Chapter 6 (application no. 41704), funding from the European Research Council to my supervisor, Eric Warrant, (Advanced Grant No. 741298), which has enabled much of the research I have done towards this thesis, long-term bioinformatics support from the National Bioinformatics Infrastructure Sweden (NBIS), funded by the Knut and Alice Wallanberg Foundation (KAW), which has been invaluable towards producing the Bogong moth reference genome (Chapter 5), and computation resources provided to Eric and myself by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX (projects SNIC 2018/8-364 and SNIC 2021/23-74), which is partially funded by the Swedish Research Council through grant agreement no. 2018-05973, and which enabled the computational analyses in Chapter 5 and Chapter 6.

I would like to thank my supervisory panel, Prof. Eric Warrant, Prof. Jochen Zeil, Prof. Ryszard Maleszka, and Prof. Saul Cunningham, who not only lent me their formidable combined expertise and experience, but

who also gave me their unwavering support through the many academic and administrative trials and tribulations I have encountered during my time as a research student. I would particularly like to thank my primary supervisor, Eric, for providing me with countless opportunities to explore any and every rabbit hole,[1] and to whole-heartedly use my time as a PhD student to do what a student should: *learn.* To Sylvain Forêt, my Honours supervisor and would-be PhD supervisor, who was taken from us far too early—not a day after I had accepted my position as his PhD student—thank you for instilling in me a spirit for pursuing hard problems and a thirst for understanding the complexity of the natural world. You are missed.

I would also like to thank the people who accompanied me during fieldwork; Dr. Andrea Adden, Dr. Mikkel Brydegaard, Dr. Peter Caley, Claudio Cannone, Dr. David Dreyer, Mary Green, Dr. Zhenhua Hao, Dr. Anna Honkanen, Dale Horne, Dr. Samuel Jansson, Adrien Lefèvre, Benjamin Mathews-Hunter, Lina O'Reilly, Dr. Therese Reber, Mark Rullo, David Szakal, Jochen, Eric, Dzifa Amenuvegbe Wallace, and of course, the irreplaceable and irreverent Dr. Ken Green. You all contributed to a huge number of fond memories, and without you my fieldwork would have been a whole lot harder.

Similarly, I would like to thank my collaborators, and all those who have helped in the lab, or with analyses; Niccy Aitken, Andreas Bachler, Monika Baker, Brendan Beaton, Dr. Kristina Brauburger, Dr. Peter Caley, Dr. Allison Churcher, Dr. David Dreyer, Assoc. Prof. Stanley Heinze, Lana Khaldy, Dr. Robert Kucharski, Dr. Jean-Marc Lassance, Dr. Joanna Maleszka, Mosè Manni, Benjamin Mathews-Hunter, Dr. Estelle Proux-Wéra,

---

[1] moth hole

Dr. Therese Reber, Dr. Lucile Soler, Prof. Doekele Stavenga, Oliver Stuart, and Prof. Evgeny Zdobnov. And my Australian lab-mates, Zoltan Kocsi, Dr. Trevor Murray, Dr. Fiorella Ramirez Esquivel, and Dr. Jack Simpson.

I would also like to extend my gratitude to a few people who gave me important technical, statistical, and/or academic advice which informed my experimental, analytical, and scholarly decisions; Dr. Regan Ashby, Prof. Staffan Bensch, Prof. Justin Borevitz, Assoc. Prof. Conrad Burden, Dr. Rayan Chikhi, Dr. John Clark, Prof. Marie Dacke, Dr. Alistair Drake, Dr. Ted Edwards, Dr. Marianne Horak, Dr. Francis Hui, Prof. Gavin Huttley, Dr. Karl Gordon, Dr. Max Lundberg, Dr. Henrik Lantz, Prof. Dr. Henrik Mouritsen, Dr. Björn Nystedt, Dr. John Taylor, Dr. Wee Tek Tay, and Dr. Tom Walsh.

Thank you also to the anonymous examiners who have agreed to critically review this work. I understand this is a considerable undertaking, and I am grateful for your time and expertise. I hope you find your reading enjoyable.

Finally, I would like to express my gratitude to my family for their tremendous support over the past few years (and before that!), and a big thanks to Papa, Ben, and Peter for proofreading. To Delali; you have no idea how much your impending existence encouraged me to get this thesis finished, so thank you for forcing me to get it across the line. Last but not least, thank you Dzifa. Your love, companionship, teamwork, and support is more than I could ever ask for.

Table of Contents

List of Figures

List of Tables

Papers not included in this thesis

Adden, A., Dreyer, D., Frost, B., Mouritsen, H., Xu, J., Green, K., Whitehouse, M., Chahl, J., **Wallace, J.**, Foster, J., Heinze, S., Warrant, E., 2020a. The starry night sky provides true compass information for long-distance nocturnal navigation in the Australian Bogong moth, in: Adden, A., There and Back Again: The Neural Basis of Migration in the Bogong Moth (PhD Thesis). Lund University, Faculty of Science.

Green, K., Caley, P., Baker, M., Dreyer, D., **Wallace, J.**, Warrant, E., 2021. Australian Bogong moths *Agrotis infusa* (Lepidoptera: Noctuidae), 1951–2020: Decline and crash. Austral Entomology. https://doi.org/10.1111/aen.12517

Stavenga, D.G., **Wallace, J.R.**, Warrant, E.J., 2020. Bogong moths are well camouflaged by effectively decolourized wing scales. Frontiers in Physiology 11, 95.

Chapter 1

Introduction

With change in season comes change in environment. Resources which are abundant in summer may become scarce in winter or *vice versa*. Similarly, an area which is suitable for a particular species during one season may become inhospitable, or at least ecologically unfavourable in another, simply because it is too hot, too cold, too wet, too dry, or perhaps too abundant in natural enemies. Evolution has found many solutions to this problem, but a particularly interesting one is the to-and-fro movement between seasonally suitable habitats, known as migration (Dingle, 2014).

For an animal to successfully migrate, it needs to be able to determine a number of things. First, it needs to know when to leave. Next, the animal needs to be able to pick a direction and stay the course to efficiently transport them away from their starting point, possibly refuelling on the way. For goal-oriented migrants, this includes determining in which direction their destination is, and therefore in which direction to travel. Finally, once at their destination, the animal needs to determine where and when to stop. In species which make the return journey, this process is then repeated. Often, these navigational procedures are not trivial, and require the integration of a multitude of sensory inputs for behavioural control (Buehlmann et al., 2020; Chapman et al., 2015; Durieux and Liedvogel, 2020; Freas and Cheng, 2022; Heinze, 2017). Understanding the solutions evolution has produced to solve such navigational tasks is of interest to biologists and engineers alike, as solutions which work for animals may also work for autonomous machines (Kaushik and Olsson, 2020; Morton et al., 2021; Strydom et al., 2016).

Such complex navigational challenges are well known to be solved by a diversity of animal life (Mouritsen, 2018). Many adult birds are able to navigate over thousands of kilometres to precisely the same nest, year after year (Salewski et al., 2000). Sea turtles and salmon navigate across vast oceans to return to the same areas which they hatched (Lohmann and Lohmann, 2019). Wildebeest cover over 1500 km by hoof annually as they circumnavigate the Serengeti, seeking optimal grazing areas as the seasons change (Torney et al., 2018). And Arctic terns repeatedly navigate over incredibly long distances between the Earth's poles (Egevang et al., 2010).

But these skills are not confined to vertebrates. In fact, owing to their accessibility and reduced neurological complexity (when compared to vertebrates), insects have proven to be valuable models for animal navigation research (Honkanen et al., 2019). Many insects perform impressive migrations (Chapman et al., 2015; Satterfield et al., 2020), although typically these movements occur over multiple generations, and between broad latitudinal zones rather than specific locations (Gao et al., 2020). Two notable exceptions are the eastern-North American populations of Monarch butterfly *Danaus plexippus*, that makes a highly directed diurnal autumn migration from broadly dispersed areas of North America to a specific overwintering destination in central Mexico (over several generations) (Urquhart, 1987), and the mysterious Australian Bogong moth *Agrotis infusa* (Fig. 1.1, *upper left inset*), that performs a nocturnal spring migration from broadly dispersed areas of south-eastern Australia to specific mountainous *oversummering* destinations in the Australian Alps (Common, 1954, 1952). Unlike Monarch butterflies, a single generation of Bogong moths performs the entire round-trip migration.

The migration of the Bogong moth (Fig. 1.1) is a wonderful example of one of nature's solutions to these navigational challenges, which is repeatedly executed by new, naïve generations of moths in an extraordinarily precise manner, year after year (reviewed by Warrant et al., 2016). Having developed through the winter (Fig. 1.1, *upper right inset, outer circle*) dispersed across southern Queensland, New South Wales (NSW), South Australia, and western Victoria, young adult Bogong moths escape the coming dry summer heat of their breeding grounds by flying to the cool caves and crevices of the Australian Alps (Fig. 1.1, *white areas*), where they gregariously aestivate (Fig. 1.1, *inset, lower right*), camouflaging themselves against the rock walls (Stavenga et al., 2020) until it's time to return to breed in autumn (Green, 2010a). This journey can be over 1000 km. The moths use relatively few caves, and they have been using the same particular caves for *at least* thousands of years, and the equivalent number of generations (Keaney et al., 2016; Stephenson et al., 2020). Moreover, the moths come from a wide range of directions (Fig. 1.1, *arrows*), with breeding grounds to the north, arcing right through to the south-west (Green, 2008). This adds complexity to the impressive navigational feat they perform, especially given how the moths manage to find their cave without ever having been there before.

In addition to having a unique life history characterised by a remarkable round-trip migration, the Bogong moth is accessible to neuroethological (Adden et al., 2020b) and behavioural (Dreyer et al., 2021, 2018) experimentation, and is typically highly abundant—it is even considered a minor pest in the low-lying agricultural areas in which it breeds (Common, 1954; Farrow and McDonald, 1987). Thus, the Bogong moth makes for an almost ideal system to study the ecological, sensory, neural, and—thanks to the work presented in Part II of this thesis—molecular basis of long-distance navigation (Heinze and Warrant, 2016).

Indeed, fruitful research has already recently been done on the navigation systems of the Bogong moth. For instance, we now know that the Bogong moth possesses a so-called "compass sense" that relies on geomagnetic (Dreyer et al., 2018) and stellar (Adden et al., 2020a) information, and we have a growing understanding of how this information is processed in the brain (Adden, 2020). However, we still don't know the molecular mechanism for magnetoreception in the Bogong moth (or, for that matter, any long-distance navigator) but promising progress has been made in songbirds (Xu et al., 2021). And we still don't know how the Bogong moth pinpoints its migratory destination, which is one of the most important open mechanistic questions in long-distance navigation research (Mouritsen, 2018).

And so begins this thesis, which, as the title suggests, aims to unravel the mystery of migratory behaviour in the Bogong moth using genomics and automated monitoring techniques...

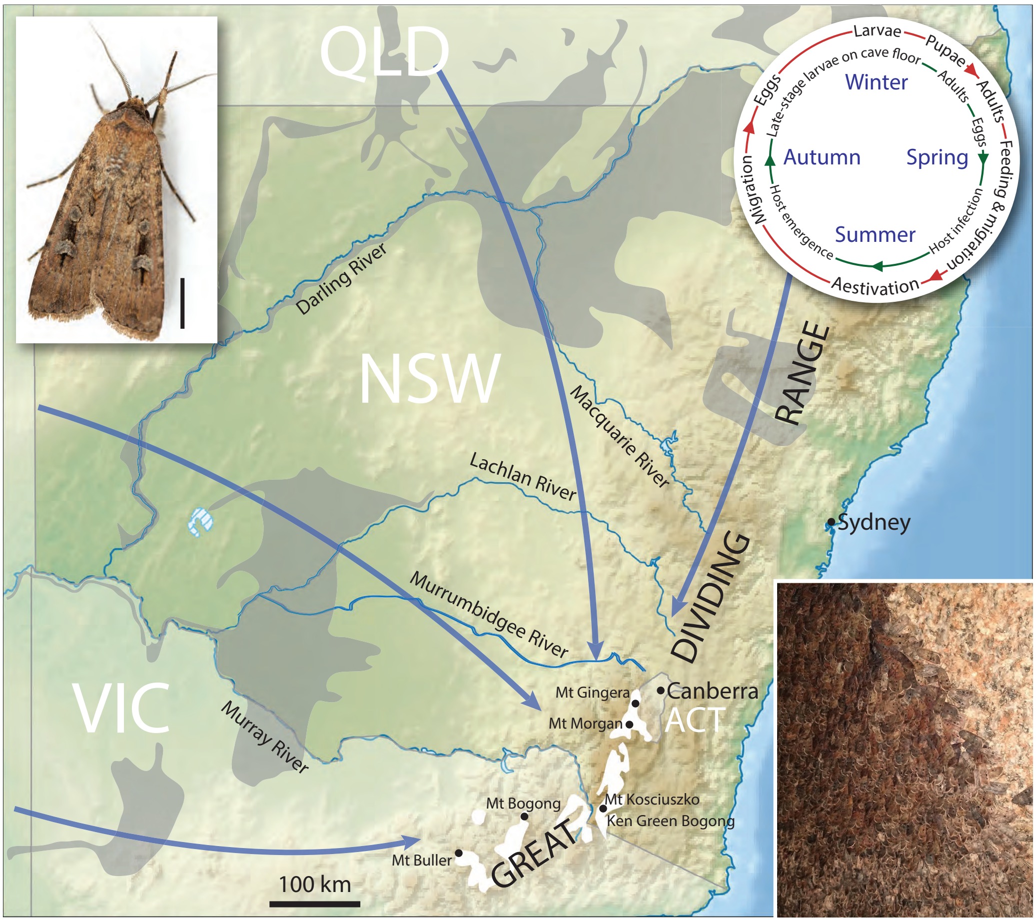


Figure 1.1: The Bogong moth. **Inset, upper left:** A male Bogong moth (*Agrotis infusa*). Scale bar = 5 mm. Photo courtesy of Dr. Ajay Narendra, Macquarie University, Australia. **Main:** Likely migratory routes (*arrows*) of moths during spring to alpine regions in south-eastern Australia. Autumn migration occurs in the reverse directions. Areas of grey cracking clays—favoured soils for Bogong moth winter development—are shown in *grey*. The *white areas* represent elevations above 1500 m, where all known summer aestivation sites are located. **Inset, upper right:** The life cycles of the Bogong moth (*outer red circle*) and the parasitic mermithid nematodes *Amphimermis bogongae* and *Hexamermis cavicola* (*inner green circle*). The nematode life cycle occurs entirely within the Bogong moth aestivation cave. Bogong moths undergo a spring migration to escape the increasingly warm conditions of the breeding grounds. Derived from information given in Common (1954) and Welch (1963). **Inset, lower right:** Around 17,000 moths/m^2 undergo a summer aestivation of up to four months on the walls of specific caves in the Australian Alps before making the return migration in autumn.

But things are never quite that simple. I will now briefly digress to discuss the plight of the Bogong moth. In the spring of 2017, the year I started this project, I set off on a month-long trip around NSW, light-trapping on most nights using a powerful searchlight and a white bed sheet (typically strung up under a branch, or between two trees). The goal of this trip was to collect migrating Bogong moth samples for transcriptional profiling (Chapter 5), and to collect Bogong moths from across their breeding grounds, so that I could later perform population genetics on them (Chapter 6). One year earlier, Profs. Eric Warrant and Barrie Frost had made a similar (albeit shorter) trip, and had reported catching 83 Bogong moths in just three nights while in the breeding grounds. In the five years prior to that (2011-2015), hundreds of Bogong moths were caught per night during spring on Mt. Kaputar, in northern NSW, which is along the Bogong moths' migratory route.

My experience in spring 2017 was quite different—in fact, it took two entire weeks of light-trapping every evening to catch just 44 Bogong moths in the breeding grounds, and at Mt. Kaputar, it took ten days to catch just 70. Later that year, a massive reduction of arrivals of Bogong moths to the mountains was observed (Mansergh et al., 2019). It was a similar experience the following year. In spring 2018, by the end of about 30 nights of light trapping in the breeding grounds (this time in western NSW and western Victoria), we caught on average just one Bogong moth per night. Meanwhile, at Mt. Kaputar, about 100 moths were caught by Dr. David Dreyer in a two-week period in October. With the support of field notes and an assortment of data going back decades, we concluded that the Bogong moths had undergone a dramatic population crash following decades of slow decline (Green et al., 2021). Ultimately, this led to the Bogong moth being listed as endangered by the IUCN Red List, following an assessment to which I contributed (Warrant et al., 2021).

Naturally, the population crash of the Bogong moth increased the challenge of studying its migratory behaviour. However, it also provided an acute motivation for studying the biology of the Bogong moth more generally, as the stakes became not just developing an understanding of how Bogong moths navigate, but also the conservation of this iconic animal and the ecosystems of which it is part. It is clear that the patchy (albeit deeply impressive) type of monitoring evidence we provided to infer the Bogong moths' population decline (Green et al., 2021) will not be sufficient if we are to conserve the Bogong moth migration going forward, and a more quantitative approach is required (Wintle et al., 2021). This—along with a multitude of questions about the Bogong moths' navigation—prompted the development of a method for monitoring the moths over long periods, which is presented and implemented in Part I of this thesis.

This novel method takes advantage of a peculiar behaviour of otherwise dormant, aestivating Bogong moths—a behaviour which Common (1954) described in some detail, but was unable to provide a completely satisfactory explanation for. Namely, on most nights throughout summer, some portion of the aestivating Bogong moths take to the air. On warmer nights, the number of moths which take flight is rather large—and the result is spectacular: hundreds of thousands of moths fly in a chaotic frenzy, colliding with each other and crashing into the granite rock faces surrounding their aestivation caves, filling the air with a thick cloud of allergenic lepidopteran scales. This occurs just after sunset, and lasts for about an hour, after which the moths return to the relative safety of their caves. In our research group, we admiringly refer to this phenomenon as “the Bogong maelstrom”—and witnessing it is something of a rite of passage. The purpose of these flights is unknown, although drinking moths have been observed (Warrant et al., 2016). Whatever the case, the flights are almost certainly important, as otherwise this behaviour would presumably be strongly selected against, owing to the risk of predation by bats while in the air (Common, 1954).

It transpires that this evening flight behaviour is useful for inferring the presence of otherwise inaccessible Bogong moths (which prefer deep, small cracks in the granite complexes they occupy during the summer months), as the abundance of *flying* moths can readily be measured (Chapter 2). Having established this technique, we can then take such measurements each day across the entire Bogong moth aestivation season, allowing us to regress the intensity of the behaviour against daily weather factors, and infer far more robust estimates of the arrival and departure of the moths than was previously possible (Chapter 3). For reference, the previous method for monitoring the arrival of the Bogong moths to the mountains—which was used for the last 45 years—was for one man to simply ski about all winter and spring until he spotted one (Green et al., 2021). Moreover, the new method can be used to start to disentangle the possible purposes of the Bogong moths’ evening flights, and even provide evidence that the actual purpose has something to do with the very thing that makes the Bogong moth so interesting: navigation (Chapter 4).

But of course, there is more to the Bogong moth than *just* its navigational abilities. For instance, the Bogong moth is of great cultural importance. For thousands of years (Keaney et al., 2016; Stephenson et al., 2020), Bogong moths were an important food source for Aboriginal people from the areas surrounding the Australian Alps, who would converge on the mountains between November and February to undertake a variety of cultural practices, including collection of, and feasting upon, Bogong moths (Flood, 1996, 1980). Although this practice ended during the historical period of genocide and dispossession of Aboriginal people which followed the arrival of Europeans at the turn of the 19th century, the Bogong moth maintains its cultural significance, and is the subject of the annual *Ngan Girra Festival* held in Albury, NSW (Love, 2010) and of numerous publicly-displayed artworks (e.g. Davys, 2014; Foley, 2001; Knox, 2012; Rennie, 2017; Tsuri, 2017; Williams, 2003; Williams and Harding, 2001).

The Bogong moth also happens to be of great ecological importance, and is something of a keystone species to the ecosystem of the Australian Alps, transporting some 4.9 Tj of energy, 7.2 t of nitrogen, and 0.97 t of phosphorus from its breeding grounds into the alpine ecosystem each year (Green, 2011). In doing so, it provides food for many species, including antechinuses *Antechinus swainsonii* (Green, 1989), rats *Rattus fuscipes* (Carron et al., 1990), bats *Chalinolobus gouldii* and *Tadarida australis* (Mitchell and Chick, 2002), ravens *Corvus mellori* (Green, 2011), foxes *Vulpes vulpes* (Green, 2003), wild pigs *Sus scrofa* (Caley and Welvaert, 2018), the endangered mountain pygmy possum *Burramys parvus* (Smith and Broome, 1992), ants (personal observations),[1] and two species of obligate Bogong moth-

---

[1] I am unsure of the species, but I have often found that these tiny black ants make stopping for lunch outside a Bogong moth aestivation cave rather unpleasant. They get in *everything*. Bogong moth dust may be allergenic, but at least moths don't bite!

Figure 1.2: A pool of water at the entrance of a Bogong moth aestivation cave on Mt. Kosciuszko, in late February, shortly after rain. The pool contains hundreds of Bogong moth-parasitic nematodes, *Amphimermis bogongae*, which can grow up to 20 cm long (Welch, 1963). Normally, the nematodes would be under the soil, but apparently they are drawn to the surface when it rains (image credit: Jesse Wallace). **Inset:** Close-up of a "bundle" of nematodes (Bogong moth for scale—image credit: Australian National Insect Collection).

parasitic mermithid nematodes, *Hexamermis cavicola* and *Amphimermis bogongae* (Fig. 1.2) (Welch, 1963).

This latter relationship is particularly interesting, especially since these two species of mermithid nematodes are only found within Bogong moth aestivation caves (Common, 1954; Welch, 1963), and they therefore depend on the repeated annual arrival of Bogong moths to those exact caves. The nematodes overwinter as free-living final-stage larvae, buried deep within moist layers of cave-floor detritus left by centuries of Bogong moth generations. At the beginning of spring the nematodes moult into sexually mature adults and subsequently lay their eggs (Fig. 1.1, *inset, upper right*). Then, coinciding with the arrival of the Bogong moths, the eggs hatch, and the resulting larvae begin to infect many of the now aestivating Bogong moths, which provide a rich food supply for the growing nematodes. The larvae eventually reach a length of up to 20 cm (Welch, 1963), filling the infected moths' body cavities and killing the moths by the time the nematodes exits in their final larval stage in late summer (after which they return to the cave floor in preparation for the coming winter, Fig. 1.2). Mermithid nematode infection of *adult* insects is incredibly rare (*larval* infection is the norm), which underscores the remarkable evolutionary coupling of the life cycles of these two nematodes with that of their adult Bogong moth host.

Now, returning to the problem of Bogong moth navigation. We know that Bogong moths occupy precisely the same sites for their aestivation each year. And from archaeological evidence, we know that this has been the case for at least two millennia (Stephenson et al., 2020). But the co-evolution of the Bogong moth with *A. bogongae* and *H. cavicola* indicates that the Bogong moths' oversummering site-fidelity has existed for *much* longer than that. This navigational precision has somehow been maintained over such a long period by the Bogong moth despite there being no overlap in migratory generations, and no repeat migrations, meaning there is no opportunity for any Bogong moth to learn its migratory destination from others, or from its own experience (as opposed to e.g. site-fidelitous songbirds, Salewski et al., 2000).

How does the Bogong moth achieve this incredible feat? Of course—as mentioned above—the moth has a compass. But a compass is only one part of a series of tools and pieces of information that one needs in order to navigate. In particular, a compass can tell a Bogong moth which way it *is* heading, but the Bogong moth also needs to know in which direction it *ought* to head. As we have established, this information cannot be *learned* by the moth—it simply has no opportunity to do so. Therefore, it must be *inherited*.

In Part II, we embark on a search for the source of this heritability, through the previously uncharted territory of the Bogong moth genome. On the way, we make a number of interesting—and sometimes unexpected—discoveries. Some notable examples are that the Bogong moth slows its genetic machinery and metabolism, but up-regulates its immune system during its aestivation, and that the genome of the Bogong moth has large amounts of cytosine methylation, even within so-called "non-CpG" contexts, contradicting recent claims by Mendoza et al. (2021) that this type of modification is a vertebrate invention, and therefore not present in any insect (Chapter 5). In Chapter 6, we discover that the Bogong moth population is, surprisingly, approximately panmictic, meaning that there is no detectable differentiation between moths from different parts of the breeding grounds at the whole-genome level. Remarkably, in spite of panmixia, there are a small number of genetic variants which are correlated with inferred migratory direction, suggesting a possible molecular basis for migratory direction inheritance, a discovery that lays the foundation for promising lines of future research.

In summary, the Bogong moth is a magnificent long-distance migrant, an accessible emerging model for animal navigation research, a keystone of the Australian Alps ecosystem (with a fascinating and unusual evolutionary relationship with two species of nematode), and an endangered cultural icon. Despite this, there are many gaps in our understanding of its biology: What are the proximate triggers of the Bogong moth migration? What is the purpose of the evening flights undertaken by a portion of aestivating Bogong moths each night during summer? How do Bogong moths know when they have arrived, and can stop their migration, and how does the Bogong moth find its final *specific* aestivation site? What is the molecular basis of the geomagnetic sense in the Bogong moth? Are there specific genetic drivers of the moth's heritable migratory direction? What is the impact of the moth's close association with the parasitic nematode on both genomic and epigenomic systems? These questions—broad as they are—underpin the research that has resulted in this thesis. And perhaps this thesis doesn't quite answer any of these questions fully but instead raises a whole lot more. Nonetheless, it is my hope that the reader will find this thesis a worthwhile contribution to our collective understanding of the biology of the Bogong moth, and that the research infrastructure built herein—in the form of a novel method for monitoring an endangered insect and the elucidation of the reference genome of the same—will provide a foundation for the work that solves these problems in years to come.

And so begins this thesis…

Part I

Camfi: Observations of Bogong moth migration and aestivation using wildlife cameras

This part is made up of three chapters, each structured as a research paper. In these chapters, I, along with my co-authors, introduce and implement a novel method for the long-term, automated monitoring of Bogong moths, which uses nothing but unmodified, off-the-shelf, inexpensive wildlife cameras, and a few hours of computation time—on not-so-inexpensive graphical processing units (GPUs). This method, which is named “Camfi” (for **C**amera-based **A**utomated **M**onitoring of **F**lying **I**nsects), is the culmination of months of algorithm development, programming, and fieldwork.

In Chapter 2, the monitoring method is introduced, and validated against a relatively small set of data obtained from a boulder field near Cabramurra, in the Snowy Mountains, NSW. This boulder field, and others like it, are habitat for the critically endangered mountain pygmy possum *Burramys parvus* (Hawke et al., 2019). It is also a stop-over site for the migration of the Bogong moth.

In Chapter 3, cameras were deployed over two summers, close to known Bogong moth aestivation sites in the Australian Alps, NSW. Here, we demonstrate the utility of Camfi for long-term monitoring, and make the case for its broader adoption, as efforts to conserve the dwindling Bogong moth population take off. In the process, we make a number of novel observations relating to the migratory and aestivation behaviour of the Bogong moth. These include unprecedented observations of a cluster of aestivating Bogong moths inside a cave on the day a major bushfire came within 1 km of the site. The results of analysing the rich, quantitative data produced by our method demonstrate that Bogong moth flight behaviour heavily depends on the weather and other yet-to-be-discovered processes, highlighting the importance of the collection of dense monitoring data to further our understanding of the dynamics of the Bogong moth population.

In Chapter 4, Camfi is extended to allow for the tracking of multiple flying insects in short video clips. This enables us to collect quantitative data on various characteristics of Bogong moth flight in the wild, including the direction of displacement of the moths. When used to detect Bogong moths on two elevation transects on Mt. Kosciuszko, the method produces an enormous amount of moth flight data. Analyses of these data demonstrate that Bogong moths fly relative to visual landmarks in the wild, and use their evening flights to gradually adjust their elevation, possibly while calibrating their internal navigation systems to the visual panorama, the setting sun, and geomagnetic field.

Our results—and the novel methods we used to produce them—open the door to many promising avenues of experimentation, ecological observation and long-term monitoring, which will undoubtedly lead to further important discoveries concerning the behaviour, ecology, and population dynamics of the iconic and endangered Bogong moth.

Chapter 2

Inexpensive monitoring of flying insect activity and abundance using wildlife cameras

Jesse R A Wallace[1,2], Therese Reber[2], Brendan Beaton[1], David Dreyer[2] and Eric J Warrant[1,2]

[1]Research School of Biology, Australian National University. [2]Lund Vision Group, Department of Biology, Lund University, Sweden.

Abstract

1. The ability to measure flying insect activity and abundance is important for ecologists, conservationists and agronomists alike. However, existing methods are laborious and produce data with low temporal resolution (e.g. trapping and direct observation), or are expensive, technically complex, and require vehicle access to field sites (e.g. radar and lidar entomology). 2. We propose a method called “camfi” for long-term non-invasive monitoring of the activity and abundance of low-flying insects using images obtained from inexpensive wildlife cameras, which retail for under USD$100 and are simple to operate. We show that in certain circumstances, this method facilitates measurement of wingbeat frequency, a diagnostic parameter for species identification. To increase usefulness of our method for very large monitoring programs, we have developed and implemented a tool for automatic detection and annotation of flying insect targets based on the popular Mask R-CNN framework. This tool can be trained to detect and annotate insects in a few hours, taking advantage of transfer learning. 3. We demonstrate the utility of the method by measuring activity levels and wingbeat frequencies in Australian Bogong moths *Agrotis infusa* in the Snowy Mountains of New South Wales, and find that these moths have log-normally distributed wingbeat frequencies (mean = 49.4 Hz, std = 5.25 Hz), undertake dusk flights in large numbers, and that the intensity of their dusk flights is modulated by daily weather factors. Validation of our tool for automatic image annotation gives baseline performance metrics for comparisons with future annotation models. The tool performs well on our test set, and produces annotations which can be easily modified by hand if required. Training completed in less than 2 h on a single machine, and inference took on average 1.15 s per image on a laptop. 4. Our method will prove invaluable for ongoing efforts to understand the behaviour and ecology of the iconic Bogong moth, and can easily be adapted to other flying insects. The method is particularly suited to studies on low-flying insects in remote areas, and is suitable for very large-scale monitoring programs, or programs with relatively low budgets.

2.1 Introduction

The ability to measure flying insect activity and abundance is important for ecologists, conservationists and agronomists alike. Traditionally, this is done using tedious and invasive methods including nets (e.g. Drake and Farrow, 1985), window traps (e.g. Knuff et al., 2019), light traps (e.g. Beck et al., 2006; Infusino et al., 2017), and pheromone traps (e.g. Athanassiou et al., 2004; Laurent and Frérot, 2007), with the latter being favoured by agronomists for its specificity. The WWII development of radar led to the introduction of radar ornithology (Eastwood, 1967; Gauthreaux Jr and Belser, 2003), and ultimately radar entomology (Drake and Reynolds, 2012; Riley, 1989), which facilitated non-invasive remote sensing of insects flying up to a couple of kilometres above the ground, and became extremely important for understanding the scale and dynamics of insect migration (Chapman et al., 2011). More recently, entomological lidar has been introduced, which benefits from a number of advantages over radar, in particular the ability to measure insects flying close to the ground, without suffering from ground clutter (Brydegaard et al., 2017; Brydegaard and Jansson, 2019). However, both entomological radar and entomological lidar systems are relatively large (requiring vehicle access to study sites), bespoke, expensive, and require expertise to operate, reducing their utility and accessibility to field biologists.

We propose a method for long-term non-invasive monitoring of the activity and abundance of low-flying insects using inexpensive wildlife cameras, which retail for under USD$100 and are simple to operate. We show that in certain circumstances, this method facilitates the measurement of wingbeat frequency, a diagnostic parameter for species identification. We demonstrate the utility of the method by measuring activity levels and wingbeat frequencies in Australian Bogong moths *Agrotis infusa* that were photographed flying across a boulder field near Cabramurra in the Snowy Mountains of New South Wales. The Bogong moth is an important source of energy and nutrients in the fragile Australian alpine ecosystem (Green, 2011), and is a model species for studying directed nocturnal insect migration and navigation (Adden et al., 2020b; Dreyer et al., 2018; Warrant et al., 2016). A dramatic drop in the population of Bogong moths has been observed in recent years (Green et al., 2021; Mansergh et al., 2019), adding it to the growing list of known invertebrate species whose populations are declining (Sánchez-Bayo and Wyckhuys, 2019). The present method will prove invaluable for ongoing efforts to understand the behaviour and ecology, and monitor the population of this iconic species. Our method can easily be adapted to other flying insects, and is particularly suited to large-scale monitoring programs with limited resources.

2.2 Methods

The methods outlined below summarise the use of wildlife cameras for monitoring flying insects, and detail specific methods employed in this study. With the exception of the camera set-up (in Section 2.2.1 and Appendix A.7), and the manual image annotation (Section 2.2.2), each of the methods described below have been automated in our freely available software. A full practical step-by-step guide for using this method along with complete documentation of the latest version of the code is provided at https://camfi.readthedocs.io/. A PDF version of the documentation is provided at https://camfi.readthedocs.io/_/downloads/en/latest/pdf/.

2.2.1 Image collection

A total of ten wildlife cameras (BlazeVideo, model SL112) were mounted in various orientations at an alpine boulder field near Cabramurra, NSW (35°57’03S 148°23’50E, circa 1520 m elevation). The location was chosen as it is a known stopover point for Bogong moths on their forward migration. During the study period, light trapping was also being done in the area as part of other work, and an overwhelming majority of insects caught in these traps each night were Bogong moths (Linda Broome, pers. comm.). This fact assists our analysis by validating the assumption that most moths observed by the cameras were a single species (i.e. Bogong moths). The locations of four of the cameras are shown in Fig. 2.1a.

The cameras were set to capture photographs on a timer, capturing one photo every ten minutes, between the hours of 19:00 and 07:00 each night (Australian Eastern Daylight Time). Each camera was equipped with a 38-LED infra-red (940 nm wavelength) flash for unobtrusive night-time photography. Capture settings, such as ISO, exposure time, and whether to use the flash, were automatically selected by the camera before each capture, based on the ambient light levels. The cameras employ a fixed-focus lens.

The cameras were deployed on 14 November 2019 and collected on 26 November 2019, for a total of 11 nights of captures, resulting in a total of 8640 recorded images.

2.2.2 Image annotation

All images were manually annotated for flying moths using VIA (Dutta and Zisserman, 2019). It was noted that for many of the night-time shots, the exposure time was relatively long, which resulted in considerable motion blur from flying moths. In cases when this motion blur was completely contained within the frame of the camera, a polyline annotation from tip to tail of the motion blur was made in VIA, following the curved or straight path of the motion blur (illustrated in Fig. 2.1b). In cases where the motion blur was not completely within the frame of the camera, or where the motion blur was short with respect to the wingbeat, either a circular or point annotation was used instead (illustrated in Fig. 2.1c). Image metadata, including date and time of capture, and exposure time, were extracted from each of the images and incorporated into the output data file from VIA to enable downstream analyses, using our newly developed Python program, named "camfi".

2.2.3 Automated annotation using Mask R-CNN

Although the process of manually annotating the images is simple to undertake, it is also time-consuming, particularly for large volumes of images. For large-scale studies, it may be desirable to use automated annotation, either by itself or in conjunction with manual annotation. To that end, we have developed an automatic annotation tool, which is included with camfi, and used by running `camfi annotate` from the command-line. The automatic annotation relies on Mask R-CNN (He et al., 2017), a state-of-the-art deep learning framework for object instance segmentation. The tool

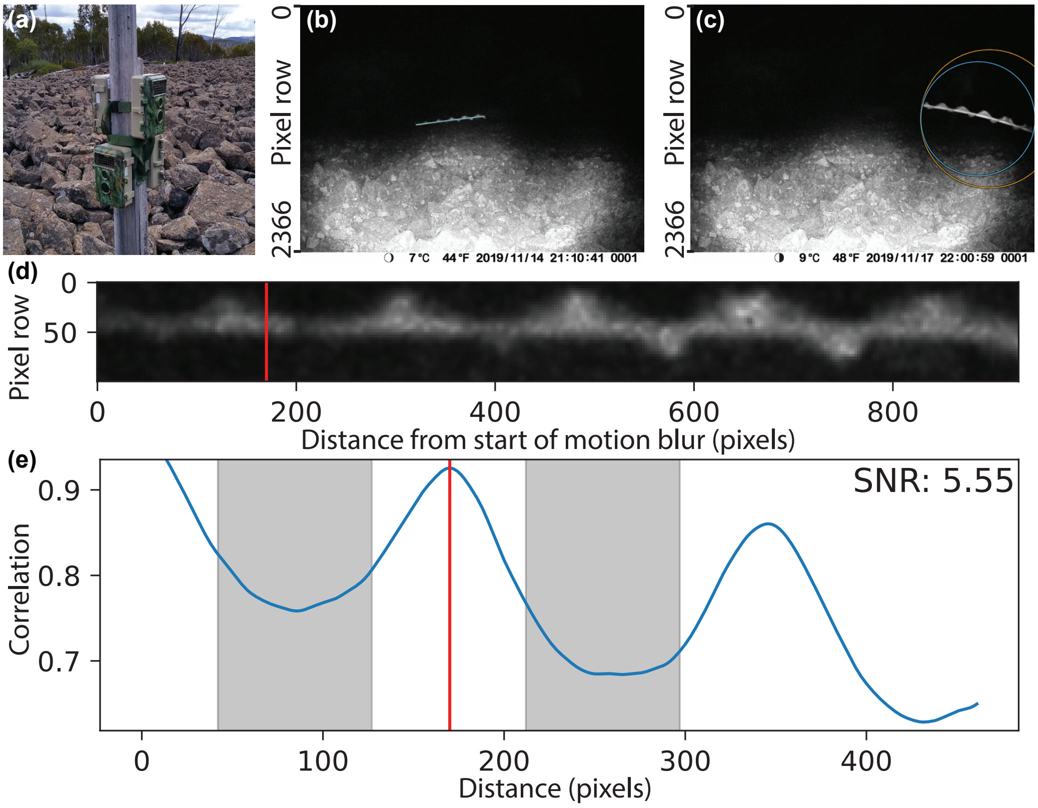


Figure 2.1: Example images showing data collection procedures used in this study. **(a)** Ten wildlife cameras (BlazeVideo, model SL112) were set to capture still photos on timers, and were deployed at the study site in a boulder field near Cabramurra, NSW in November 2019 (four cameras shown). **(b)** Motion blurs of moths captured by the cameras were marked with a polyline annotation. Manual annotation made in VIA (Dutta and Zisserman, 2019) is shown in *orange*, and the annotation made by our automated procedure is shown in *blue* (although since both annotations are very similar, they overlap and only the *blue* annotation is visible). **(c)** Circular or point annotations were used for images of moths whose motion blurs were not fully contained within the frame of the camera, or where the length of the motion blur was too short to see the moth's wingbeat (latter case not shown). Manual annotation made in VIA (Dutta and Zisserman, 2019) is shown in *orange*, and the annotation made by our automated procedure is shown in *blue*. **(d)** Straightened and cropped "region-of-interest image" of moth motion blur, taken from image shown in b. Red vertical line shows periodicity along the axis of the motion blur as calculated by our algorithm. **(e)** Autocorellation of region-of-interest image (shown in d) along the axis of the motion blur. *Red line* shows peak periodicity as calculated by our algorithm. Signal-to-noise ratio (SNR) is calculated as the Z-score of the correlation at the peak, if drawn from a normal distribution with mean and variance equal to those of the correlation values within the *shaded regions*, defined by the intervals $\left(\frac{1}{4}P^*, \frac{3}{4}P^*\right) \cup \left(\frac{5}{4}P^*, \frac{7}{4}P^*\right)$, where P^* is the pixel period.

operates on VIA project files, allowing it to serve as a drop-in replacement for manual annotation. The tool also allows the annotations it generates to be loaded into VIA and manually edited if required.

2.2.3.1 Training:

To simplify training of the model to target other species, we have implemented a tool which automates the training process, and this is described below. This tool is packaged with camfi, and is used by running `camfi train` from the command-line. We have also included the model which we trained with camfi, so for species whose appearance is similar to that of Bogong moths while in flight, re-training the model may not be necessary.

We adopted the Mask R-CNN model architecture (He et al., 2017) with a Feature Pyramid Network (FPN) backbone (Lin et al., 2017). In the Mask R-CNN framework, the "head" of the model, which is the part of the model which generates the output, can include a variety of branches (corresponding to each type of output). We included the "class", "bounding box", and "instance segmentation" branches provided by the Mask R-CNN framework. We assigned two target classes; one for moths and another for background. We initialised the FPN backbone with a model which was pre-trained on the Common Objects in Context (COCO) object segmentation dataset (Lin et al., 2014) and employed the method of transfer learning to train the head and fine-tune the model. The data used for training were the set of images of flying moths we had previously annotated manually which contained at least one moth, as well as the corresponding manual annotations. This set contained 928 images (the remaining 7712 images from the full set of 8640 images had no moths in them). We reserved 50 randomly selected images from this set for testing, which were not seen by the model during training. Therefore, 878 images were used for training. The training data were augmented by creating new images by horizontal reflection of random individual images within the set of 878. In each iteration of training, a batch of five images[1] were used. The model was trained for 15 epochs[1] (full traversals of training data), for a total of 2640 iterations.

The manual annotations we used were polylines, points, and circles. However, Mask R-CNN operates on bounding boxes and segmentation masks, so some pre-processing of the annotations is required. These pre-processing steps are performed by our software directly on the output of the manual annotation process in VIA. For training the model, bounding boxes and segmentation masks are calculated on-the-fly from the coordinates of the manually annotated polylines, circles, and points. The bounding boxes are simply taken as the smallest bounding box of all coordinates in an annotation, plus a constant margin of ten pixels.[1] The masks are produced by initialising a mask array with zeros, then setting the coordinates of the annotation in the mask array to one, followed by a morphological dilation of five pixels.[1] For polyline annotations, all points along each of the line segments are set to one, whereas for point or circle annotations, just the pixel at the centre of the annotation is set.

We have made our annotation model available as part of the camfi software, and is the default model used by `camfi annotate`. We expect it to work out-of-the box for target species which are similar to the Bogong moth.

---

[1] Configurable parameter when running Camfi.

2.2.3.2 Inference:

Automation of the inference steps described in this section is implemented in the `camfi annotate` command-line tool, included with camfi. In inference mode, the Mask R-CNN model outputs candidate annotations for a given input image as a set of bounding boxes, class labels, segmentation masks (with a score from 0 to 1 for each pixel belonging to a particular object instance), and prediction scores (also from 0 to 1). Non-maximum suppression on candidate annotations is performed by calculating the weighted intersection over minimum (IoM) of segmentation masks of each pair of annotations in an image (the definition of IoM is provided in Appendix A.1). For annotation pairs which have an IoM above 0.4,[2] the annotation with the lower prediction score is removed. This has the effect of removing annotations which are too similar to each other, and are likely to relate to the same target. We also rejected candidate annotations with prediction scores below a given threshold.[2] For each of the remaining candidate annotations, we fit a polyline annotation using the method described below.

To fit a polyline to a candidate annotation predicted by the Mask R-CNN model, we first perform a second-order[2] polynomial regression on the coordinates of each pixel within the bounding box, with weights taken from the segmentation mask. If the bounding box is taller than it is wide, we take the row (y) coordinates of the pixels to be the independent variable for the regression, rather than the default column (x) coordinates. We then set the endpoints of the motion blur as the two points on the regression curve which lie within the bounding box, and which have an independent variable coor-

---

[2] Configurable parameter when running Camfi.

dinate ten pixels[3] away from the edges of the bounding box. The rationale for setting these points as the end points is that the model was trained to produce bounding boxes with a ten-pixel margin from the manual polyline annotations (see above). The curve is then approximated by a piecewise linear function (a polyline) by taking evenly spaced breakpoints along the curve such that change in angle between two adjoining line segments is no greater than approximately 15°.[3]

Finally, a check is performed on the polyline annotation to determine if the motion blur it represents is completely contained within the image. If it is not, it is converted to a circle annotation by calculating the smallest enclosing circle of all the points in the polyline annotation using Welzl's algorithm (Welzl, 1991). The check is performed by measuring how close the annotation is to the edge of the image. If the annotation goes within 20 pixels[3] of the edge of the image then the motion blur is considered to not be completely contained within the image, and therefore the polyline annotation is converted to a circle annotation.

The automatically produced annotations are saved to a VIA project file, and tagged with their prediction score, enabling further downstream filtering or annotation visualisation and diagnostics, as well as editing by a human if desired. We ran automatic annotation on the entire image set (8640 images) on a laptop with a Nvidia Quadro T2000 GPU. Using the GPU for inference is preferred, since it is much faster than using the CPU. However, in some cases, images which had a lot of moths in them could not be processed on the GPU due to memory constraints. To solve this problem, `camfi annotate` provides an option to run inference in a hybrid

---

[3] Configurable parameter when running Camfi.

mode, which falls back to the CPU for images which fail on the GPU.

2.2.3.3 Validation:

This section introduces a number of terms which may be unfamiliar to the reader. Definitions of the following terms are provided in Appendix A: intersection over union (Appendix A.2), Hausdorff distance (Appendix A.3), signed length difference (Appendix A.4), precision-recall curve (Appendix A.5), and average precision (Appendix A.6). As mentioned above, we kept 50 randomly-selected annotated images as a test set during model training. We ran inference and validation on the full set of images, and on the test set in isolation. For both sets, we matched automatic annotations to the ground-truth manual annotations using a bounding-box intersection over union (IoU) threshold of 0.5.[4] For each pair (automatic and ground-truth) of matched annotations we calculated IoU, and if both annotations were polyline annotations, we also calculated the Hausdorff distance d_H and the signed length difference ΔL between the two annotations. Gaussian kernel density estimates of prediction score versus each of these metrics were plotted for diagnostic purposes. We also plotted the precision-recall curve and calculated the average precision AP_{50} for both image sets. To compare future automatic annotation methods to ours, we recommend using mean IoU $\overline{IoU}$, mean Hausdorff distance $\overline{d_H}$, mean length difference $\overline{\Delta L}$, the standard deviation of length difference $\sigma_{\Delta L}$, and AP_{50} as the set of comparison metrics.

---

[4] Configurable parameter when running Camfi.

2.2.4 Wingbeat frequency measurement

For observations of moths whose motion blur was entirely captured within the frame of a camera, we use a polyline annotation, which follows the path of the motion blur. This annotation can be obtained either manually or automatically, by the procedures described above. For the analyses presented in this paper, we used manual annotations. Since the moth is moving while beating its wings, we are able to observe the moth's wingbeat (see Fig. 2.1b). Incorporating information about the exposure time and rolling shutter rate of the camera, we are able to make a measurement of the moth's wingbeat frequency in hertz. We have implemented the procedure for making this measurement as part of camfi, in the sub-command called `camfi extract-wingbeats`. The procedure takes images from wildlife cameras (like those shown in Fig. 2.1b,c) and a VIA project file containing polyline annotations of flying insect motion blurs as input, and outputs estimates of wingbeat frequencies and other related measurements. A description of the procedure for a given motion blur annotation follows.

First, a region of interest image of the motion blur is extracted from the photograph, which contains a straightened copy of the motion blur only (see Fig. 2.1d). The precise method for generating this region of interest image is not important, provided it does not scale the motion blur, particularly in the direction of the motion blur. Our implementation simply concatenates the rotated and cropped image rectangles, which are centred on each segment of the polylines, with length equal to the respective segment, and with an arbitrary fixed width.[5] We used the default value of 100 pixels.

---

[5] Configurable parameter when running Camfi.

The pixel-period of the wingbeat, which we denote P, is determined from the region of interest image by finding peaks in the autocorrelation of the image along the axis of the motion blur (see Fig. 2.1e). The signal-to-noise ratio (SNR) of each peak is estimated by taking the Z-score of the correlation at the peak, if drawn from a normal distribution with mean and variance equal to those of the correlation values within the regions defined by the intervals $\left(\frac{1}{4}P^*, \frac{3}{4}P^*\right) \cup \left(\frac{5}{4}P^*, \frac{7}{4}P^*\right)$, where P^* is the pixel period corresponding to the given peak. The peak with the highest SNR is selected as corresponding to the wingbeat of the moth and is assigned as P. The total length of the motion blur (in pixels) may then be divided by P to obtain a non-integer wingbeat count for the motion blur. The SNR of the best peak is included in the output of the program, to allow for filtering of wingbeat data points with low SNR. It should be noted that the definition of SNR used here may differ somewhat from other formal definitions. For example, this definition admits negative values for SNR (albeit rarely), in which case the corresponding measurement will surely be filtered out after a SNR threshold is applied.

When running `camfi extract-wingbeats`, supplementary figures containing the region of interest images and corresponding autocorrelation plots, similar to those presented in Fig. 2.1d,e can be optionally generated for every polyline annotation.

To calculate wingbeat frequency F_w, in hertz, we need to know the length of time that the moth was exposed to the camera, which we call Δt. Unfortunately, this is not as simple as taking the exposure time as reported by the camera, which we call t_e, due to the interaction of the first-order motion of the moth with the rolling shutter of the camera. In particular,

$$\Delta t = t_e \pm \frac{|r_1 - r_0|}{R} \quad , \tag{2.1}$$

where r_0 and r_1 are the row indices (counting rows from the top of the image) of the two respective ends of the polyline annotation, and R is the rolling shutter line rate, which was measured to be 9.05×10^4 lines s^{-1} for the cameras we used (for a method of measuring R, see Appendix A.7). The "±" reflects the fact that it is impossible to tell in which direction the moth is flying from the images alone, leading us to two possible measurements of moth exposure time, corresponding to the moth flying down or up within the image plane of the camera, respectively. Under certain circumstances, this ambiguity can be resolved by observing that $\Delta t \geq 0$, i.e. insects cannot fly backwards through time. Intuitively, we may then attempt to calculate F_w by dividing the wingbeat count by Δt (these preliminary estimates of wingbeat frequency are included in the output of `camfi extract-wingbeats`). However, this would require the assumption that the moth has a body length of zero, since the length of the motion blur, which we denote as L, is the sum of the first order motion of the moth during the exposure and the moth's body length, projected onto the plane of the camera. Clearly, this assumption may be violated, as the insects have a non-zero body length in the images. We denote the body length of the moth projected onto the plane of the camera by the random variable L_b.

The statistical procedure for estimating the mean and standard deviation of observed moth wingbeat frequency, which accounts for both the time ambiguity and the non-zero body lengths of the moths, is as follows. We begin with the following model, which relates F_w to L_b and various measured variables.

$$L_i = F_{w_i} P_i \Delta t_i + L_{b_i} \quad , \tag{2.2}$$

where i is the index of the observation. We proceed by performing a linear regression of L on $P\Delta t$ (setting $P\Delta t$ as the independent variable) using the BCES method (Akritas and Bershady, 1996) to obtain unbiased estimators of $\bar{F_w}$ and $\bar{L_b}$, as well as their respective variances, $\sigma^2_{F_w}$ and $\sigma^2_{L_b}$. Values for Δt_i are taken as the midpoints of the pairs calculated in Eq. 2.1, with error terms equal to $\frac{|r_{1_i} - r_{0_i}|}{R}$. Values for L_i are assumed to have no measurement error. Where multiple species with different characteristic wingbeat frequencies are observed, an expectation-maximisation (EM) algorithm may be applied to classify measurements into groups which may then be analysed separately. We may then test the zero body length assumption, namely $\bar{l_b} = 0$, by calculating its t statistic.

2.2.5 Implementation

Our implementation of camfi and its associated tools is written in Python 3.9 (Python Software Foundation, https://www.python.org/). The latest version of camfi relies on (in alphabetical order): bces 1.0.3 (Nemmen et al., 2012), exif 1.3.1 (Thieding et al., 2021), imageio 2.9.0 (Silvester et al., 2020), Jupyter 1.0.0 (Kluyver et al., 2016), Matplotlib 3.4.2 (Hunter, 2007), NumPy 1.21.1 (Harris et al., 2020), Pandas 1.3.0 (McKinney et al., 2010), Pillow 8.3.1 (Kemenade et al., 2021), pydantic 1.8.2 (Colvin et al., 2021), Scikit-image 0.18.2 (Van der Walt et al., 2014), Scikit-learn 0.24.2 (Pedregosa et al., 2011), SciPy 1.7.0 (Virtanen et al., 2020), Shapely 1.7.1 (Gillies et al., 2007--), skyfield 1.39 (Rhodes, 2019), Statsmodels 0.12.2 (Seabold and Perktold, 2010), strictyaml 1.4.4, PyTorch 1.9.0 (Paszke et al., 2019), TorchVision 0.10.0 (Marcel and Rodriguez, 2010), and tqdm 4.61.2 (Costa-Luis et al., 2021).

Camfi is open source and available under the MIT license. The full source code for the latest version of camfi and all analyses presented in this paper are provided at https://github.com/J-Wall/camfi. The documentation for camfi is provided at https://camfi.readthedocs.io/. Camfi is under active development and we expect new features and new trained models to be added as new versions of camfi are released from time to time. All analyses presented in this paper were done using camfi 2.1.3, which is permanently available from the Zenodo repository https://doi.org/10.5281/zenodo.5194496 (Wallace, 2021a).

2.3 Results

2.3.1 Moth activity patterns

From the 8640 images analysed, a total of 1419 manual annotations were made. Of these, 259 were circle or point annotations, which we are able to use for quantifying general activity, but which cannot be used for wingbeat analysis. The remaining 1160 annotations were polyline annotations, which we used for both activity quantification and wingbeat analysis.

We observed a daily pattern of moth activity, with marked increase in the number of moths flying during evening twilight on most days (Fig. 2.2a).

This daily pattern is clearly pronounced in Fig. 2.2b. By considering just the period of evening twilight from each day of the study, we are able to quantify the relative evening moth activity levels over time, and compare them with abiotic factors such as the weather (Fig. 2.2c). We performed a Poisson regression of evening twilight moth activity levels (number of annotations, with exposure set to the number of images taken during evening twilight that day) against daily weather factors, (Bureau of Meteorology, 2019) and found that minimum morning temperature, minimum evening temperature, daylight hours, and temperature range had a significant joint effect on observed moth numbers (Fig. 2.2c, green trace: Wald test, $\chi^2 = 25.3, df = 4, p \ll 0.001$). It should be noted that since the study period was short (11 days), daylight hours are almost linearly confounded with study day, however for longer studies (e.g. over the entire summer) daylight hours would increase before the summer solstice, and later decrease, as does Bogong moth summer-range abundance.

2.3.2 Wingbeat frequency

Of the 1160 manual polyline annotations of flying moths, 580 yielded wingbeat measurements which had a SNR exceeding the threshold of 4.0 (Fig. 2.3a). The histogram of preliminary wingbeat calculations, which do not account for non-zero body-length (Fig. 2.3a), indicated that there were likely to be two classes of insect wingbeats observed, possibly corresponding to two separate species. It was noted that the preliminary wingbeat frequencies of the less common of the two classes were centred at a value approximately half that of the more common class, indicating the possibility

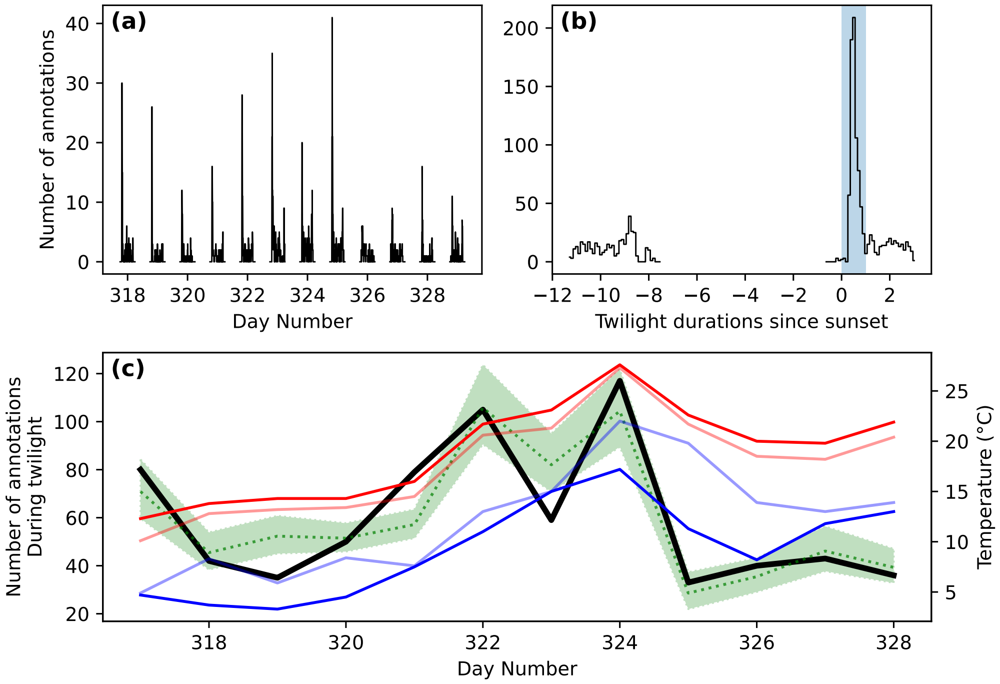


Figure 2.2: Moth activity levels during the November study period. **(a)** Number of moths observed across the study period from 10 cameras, with images taken at 10 min intervals. **(b)** Total number of moth observations by time after sunset (scaled by the duration of twilight) shows peak in activity during evening twilight (shaded blue). **(c)** Number of moths observed during twilight for each day of the study (black), shown with daily temperatures recorded by the BOM at nearby Cabramurra (Bureau of Meteorology, 2019): maximum (red), minimum (blue), 9 am (light blue), and 3 pm (light red). Predicted values for moth activity (and SE confidence interval) from a Poisson regression of number of annotations vs. daily weather factors (minimum morning temperature, minimum evening temperature, daylight hours [a proxy for study day], and temperature range) are shown in green.

that the lesser class represented an artefact of the wingbeat measurement process, where a given signal's period could conceivably be inadvertently doubled. To rule out this possibility, a subset of the wingbeat region-of-interest images were viewed, and no obvious evidence of erroneous measurements was observed. It was therefore concluded that the two observed classes of wingbeats represented a true biological signal, so for subsequent analysis we assumed there are two types of wingbeat represented in the data.

To produce unbiased estimates of mean wingbeat frequency (which accounts for non-zero body-length), we performed a linear regression of L vs. $P\Delta t$ using the BCES method (Akritas and Bershady, 1996) (Fig. 2.3b). Two target classes were identified using an EM algorithm, and regressed separately. The EM algorithm assigned 75 observations to the first class, and 505 observations to the second class, which we infer as representing Bogong moths. The slopes of the linear regressions give estimates of mean wingbeat frequency, and these are 23.7 Hz (SE = 1.8) and 48.6 Hz (SE = 1.4) for the two classes, respectively. The intercepts give estimates of mean pixel-body lengths at 21.2 pixels (SE = 30.4) and 30.4 pixels (SE = 18.9). The t statistics of the body length estimates are 0.6972 and 1.6067 respectively, for the null hypothesis that body length is zero. This leads to one-sided p-values of 0.244 and 0.054, respectively. Since both of these are above the canonical p-value threshold of 0.05, we conclude that the zero body length assumption is reasonable, and that a log-gaussian mixture model is sufficient to describe the observed wingbeat frequencies. After correcting for the (known) measurement error produced by the interaction between flight-direction ambiguity and the rolling shutter of the cameras (Eq. 2.1), the log-gaussian mixture model gives us estimates of wingbeat frequencies, which are 25.10 Hz (std = 2.88) and 49.40 Hz (std = 5.25), for the two classes of moth observed, respectively.

2.3.3 Automatic annotation

Automatic annotation performance was evaluated using a test set of 50 images, as well as the full set of 8640 images. Evaluation metrics for both sets are presented in Table 2.1. Each metric was similar across both image sets, indicating that the annotation model has not suffered from overfitting. This is also supported by the contour plots of prediction score vs. IoU, polyline Hausdorff distance, and polyline length difference (Fig. 2.4b,c,d, respectively). These plots show similar performance on both the full image set (8640 images) and the test set (50 images). Furthermore, they show that prediction scores for matched annotations (automatic annotations which were successfully matched to annotations in the manual ground-truth dataset) tended to be quite high, as did the IoU of those annotations, while both polyline Hausdorff distance d_H and polyline length difference ΔL clustered relatively close to zero. The precision-recall curves of the automatic annotator (Fig. 2.4e) show similar performance between the image sets, and show a drop in precision for recall values above 0.6. Training took 2640 iterations and completed in less than 2 h (Fig. 2.4a) on a machine with two 8-core Intel Xeon E5-2660 CPUs running at 2.2GHz and a Nvidia T4 GPU, and inference took on average 1.15 s per image on a laptop with a 6-core Intel Xeon E-2276M CPU running at 2.8GHz and a Nvidia Quadro T2000 GPU.

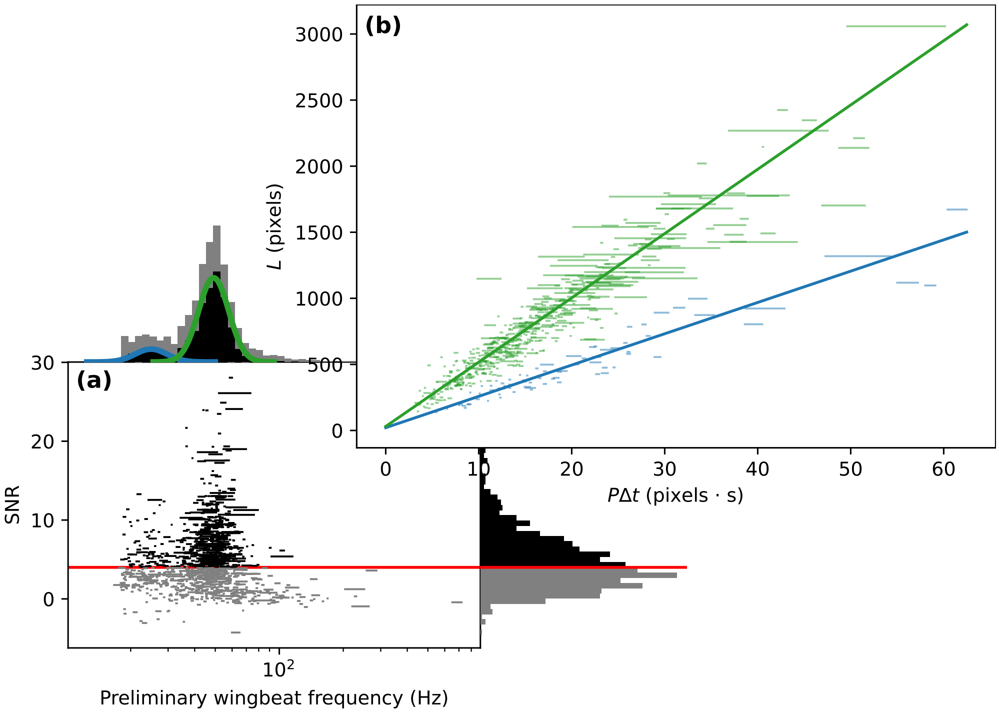


Figure 2.3: Moth wingbeat frequency measurements from wildlife camera images. Error bars indicate the two possible measurements arising from each observation, due to the interaction between flight-direction ambiguity and the rolling shutter of the cameras (Eq. 2.1). **(a)** Signal-to-noise ratio (SNR) vs. preliminary wingbeat frequency measurements on $\log_{10}$ scale, with SNR threshold (4.0) indicated by red line. Preliminary wingbeat measurements do not account for non-zero body length of observed moths. Marginal distribution histograms for both axes are shown. Data which exceeded the SNR threshold are in dark grey, and data which did not meet the SNR threshold are in light grey. The probability density functions associated with a $\log_{10}$-gaussian mixture model (GMM) of above-threshold preliminary wingbeat frequencies for two target classes are overlaid on the horizontal marginal distribution histogram (blue and green curves). **(b)** Length of motion-blur L vs. pixel-wingbeat period × exposure time of motion-blur $P\Delta t$ for observations which exceeded SNR threshold. Linear regressions are shown, which were obtained by the BCES method (Akritas and Bershady, 1996), extended to classify the data into two target classes, and regress each class separately using an expectation-maximisation (EM) algorithm. This regression eliminates the assumption of zero-body length. For convenience, the two classes are coloured in the same way as in panel a, however it should be noted that the classifications presented in each sub-figure are distinct. The slopes of the regressions estimate mean wingbeat frequency, and the intercepts estimate mean non-zero pixel-body length.

Table 2.1: Automatic annotation performance metrics when tested against the full image set (8640 images), and the test set (50 images). Performance metrics calculated are average precision AP_{50}, mean bounding-box intersection over union $\overline{IoU}$, mean Hausdorff distance of polyline annotations $\overline{d_H}$, mean signed length difference of polyline annotations $\overline{\Delta L}$, and the standard deviation of signed length difference of polyline annotations $\sigma_{\Delta L}$. Definitions of these metrics are provided in Appendix A.

Image set	AP_{50}	$\overline{IoU}$	$\overline{d_H}$	$\overline{\Delta L}$	$\sigma_{\Delta L}$
Full set	0.588	0.814	29.2	-3.31	46.4
Test set	0.687	0.805	28.8	5.16	51.0

2.4 Discussion

This paper demonstrates the utility of inexpensive wildlife cameras for the long-term monitoring of activity in flying insects, and describes how they may be used to measure the wingbeat frequency of those insects. We do not expect this method to completely replace other approaches for monitoring insects, such as trapping, which enables precise measurement of biodiversity and positive identification of species. Likewise, it will not completely replace other remote sensing approaches, such as radar and lidar, which facilitate detecting targets at long distances. However, it is clear that this method has significant potential to complement these other approaches, and in certain circumstances, replace them. For instance, in comparison to these other approaches, this method is particularly suited to monitoring assemblages of known species in remote areas, especially when it is known that the target insects are low-flying. An advantage of the presented method over trapping is that much greater temporal resolution is gained. In the present study one measurement was taken by each camera every ten minutes, and depending

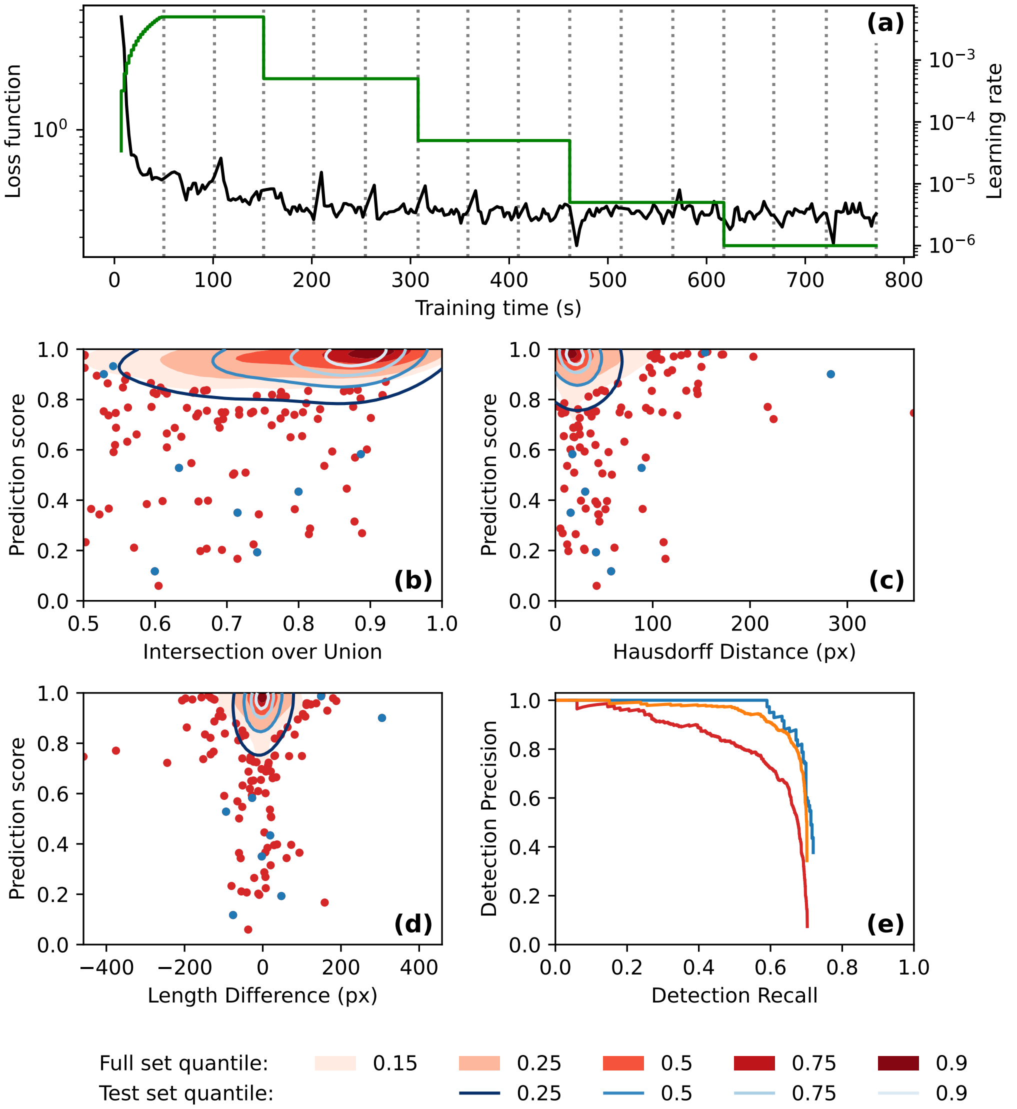


Figure 2.4: Automatic annotation evaluation plots. **(a)** Automatic annotation model training learning rate schedule (green) and loss function (black) over the course of training. Epochs (complete training data traversal) are shown with dotted vertical lines. **(b)**-**(e)** Similar performance was seen for both the full 8640-image set (red) and the test 50-image set (blue). **(b)**-**(d)** Gaussian kernel density estimate contour plots of prediction score vs. **(b)** bounding box intersection over union, **(c)** polyline Hausdorff distance, and **(d)** polyline length difference, for both image sets. Contours are coloured according to density quantile (key at bottom of figure). In each plot, data which lie outside of the lowest density quantile contour are displayed as points. **(e)** Motion blur detection precision-recall curve, generated by varying prediction score threshold. The precision-recall curve for the set of 928 images which had at least one manual annotation is shown in orange.

on the research question or absolute abundance of the insects being studied, this can easily be varied. This is in contrast to trapping studies, where only one measurement of abundance can be recorded per visit to the trap by the researcher. This provides an opportunity to use the present method to answer a variety of ethological research questions which may not be approachable with previous methods.

The measurement of wingbeat frequency has utility in distinguishing between multiple target species, especially when the observed flying insects are dominated by one known species, as is the case for the dataset we analysed, or where the wingbeat frequencies of the observed species are very different from each other. We observed two classes of wingbeat frequencies, centred at 25.10 Hz (std = 2.88) and 49.40 Hz (std = 5.25), respectively. It is likely that the former represents a larger insect, probably hawk moths (family: Sphingidae), which were observed in the light traps (Linda Broome, pers. comm.) and are known to have wingbeat frequencies of 25.2 Hz (std = 2.3) (Gau et al., 2021). The latter, more abundant class almost certainly represents Bogong moths given the light-trap confirmed abundance of Bogong moths at the study site during the study period. The author is not aware of any previous measurements of Bogong moth wingbeat frequency, however other noctuids have been recorded with similar wingbeat frequencies to our recordings (Hu et al., 2018 recorded *Agrotis* spp. of a range of body sizes with mean wingbeat frequencies ranging from 42 Hz to 58 Hz).

The method of generating summary statistics for observed wingbeat frequency is complicated somewhat by the measurement error introduced by the interaction between the ambiguity in insect flight direction and the rolling shutter of the cameras. This measurement error could be eliminated in one of two ways: 1. By taking two immediately successive exposures, which would enable inference of flight direction, or 2. By using cameras with a global shutter, which would prevent the flight direction of the insect from having any influence over the duration that the insect is exposed to the camera. Implementing either of these options is desirable, however they are not possible without significantly more expensive cameras than the type used in this study. This would limit the utility of the method for use in either large-scale or low-budget studies. Until the cost of wildlife cameras equipped with a global-shutter comes down, the most practical approach remains to handle this measurement error statistically.

This paper has presented a method for monitoring nocturnal flying insects, however there is no reason it couldn't be used for diurnal species as well, provided care is taken with regard to the placement of cameras. Namely, it would be important to have a relatively uniform background (such as the sky) in order to be able to see insects in the images during the day. In this case, the infra-red flash of the cameras would not be used and the insects would appear as dark objects on a light background. During the day, the exposure time of the cameras is much shorter than at night, so it would be impossible to use this method to measure wingbeat frequencies of day-flying insects. However, in some cases it may be possible to identify day-flying insects in the images directly. It may also be possible to recreate the type of images seen during the night in any lighting conditions by retrofitting the cameras with long-pass infra-red filters, neutral density filters, or a combination of both.

A key advantage of the present method over other approaches is that it can be readily scaled to large monitoring studies or programs, thanks to the low cost of implementation and the inclusion of the tool for automatic annotation of flying insect motion blurs. It is expected that studies implementing this method for target species which substantially differ in appearance from Bogong moths when in flight (and where the use of automatic annotation is desired) may have to re-train the Mask R-CNN instance segmentation model. We believe that the tools we have implemented make that process highly accessible.

2.5 Acknowledgements

EJW and JRAW are grateful for funding from the European Research Council (Advanced Grant No. 41298 to EJW), and the Royal Physiographic Society of Lund (to JRAW). JRAW is thankful for the support of an Australian Government Research Training Program (RTP) Scholarship. We thank Drs. Jochen Zeil, Ryszard Maleszka, Linda Broome, Ken Green, Samuel Jansson, Alistair Drake, Benjamin Amenuvegbe, Mr. Benjamin Mathews-Hunter, and Ms. Dzifa Amenuvegbe Wallace for invaluable collaboration and assistance, and Drs. John Clarke and Francis Hui for useful discussions relating to the wingbeat frequency analyses.

2.6 Author's contributions

JRAW, DD and EJW conceived the ideas; JRAW, TR, BB and DD designed the methodology; JRAW and TR collected the data; JRAW and BB analysed the data; JRAW led the writing of the manuscript with significant input from EJW. All authors contributed critically to the drafts and gave final approval for publication.

2.7 Data Availability

The images and manual annotations are available from the Zenodo repository https://doi.org/10.5281/zenodo.4950570 (Wallace, 2021b). All other data and code are available from the Zenodo repository https://doi.org/10.5281/zenodo.5194496 (Wallace, 2021a).

Chapter 3

Camera-based monitoring of Bogong moths in Alpine Australia reveals drivers of migratory behaviour

Jesse R A Wallace[1,2], Therese Reber[2], Lana Khaldy[2], Benjamin Mathews-Hunter[3], Ken Green[4], David Dreyer[2] and Eric J Warrant[1,2]

[1]Research School of Biology, Australian National University, 134 Linnaeus Way, Acton ACT 2601. [2]Lund Vision Group, Department of Biology, Lund University, Sweden. [3]School of Life and Environmental Sciences, The University of Sydney, Australia. [4]College of Asia and the Pacific, Australian National University.

Abstract

The Bogong moth *Agrotis infusa* is well known for its remarkable annual round-trip migration from its breeding grounds across eastern Australia to its aestivation sites in the Australian Alps, to which it provides an important annual influx of nutrients. Over recent years, we have benefited from a growing understanding of the navigational abilities of the Bogong moth. Meanwhile, the population of Bogong moths has been shrinking. Recently, the ecologically and culturally important Bogong moth was listed as endangered by the IUCN Red List, and the establishment of a program for long-term monitoring of its population has been identified as critical for its conservation. Here, we present the results of two years of monitoring of the Bogong moth population in the Australian Alps using a recently developed method for automated monitoring of flying insects, named Camfi. We found that the evening flights of Bogong moths occur throughout summer, and are modulated by daily weather factors. We present a simple heuristic model of the arrival to and departure from aestivation sites by Bogong moths, and confirm results obtained from fox-scat surveys which found that aestivating Bogong moths occupy higher elevations as the summer progresses. We also present the first recorded observations of the impact of bushfire smoke on aestivating Bogong moths. We observed a dramatic reduction in the size of a cluster of aestivating Bogong moths during the fire, and evidence of a large departure from the fire-affected area the day after the fire. Our results highlight the challenges of monitoring Bogong moths in the wild, and support the continued use of automated camera-based methods for that purpose.

3.1 Introduction

The Bogong moth *Agrotis infusa* is well known for its remarkable annual round-trip migration from its breeding grounds across eastern Australia to its aestivation sites throughout the high mountain areas of New South Wales, Victoria, and the Australian Capital Territory, where it forms aggregations numbering in the millions (reviewed by Warrant et al., 2016). Bogong moth aestivation was first reported during the 19th century (Bennett, 1834; Scott, 1873), but the moths have been known by Aboriginal people in the areas surrounding the Australian Alps for millennia (Keaney et al., 2016; Stephenson et al., 2020). Aboriginal people once converged on these mountainous regions during the spring-summer months to hunt and feast upon the abundant Bogong moth assemblages (Flood, 1996, 1980). In spite of this, Bogong moth migration was not understood until the 1950s, following the thorough studies of Common (1954, 1952).

In recent years, increasing efforts have been made to understand the migration of the Bogong moth from a neuroethological perspective (e.g. Adden et al., 2020b; Dreyer et al., 2018; Vries et al., 2017; Warrant et al., 2016), particularly with respect to how Bogong moths navigate. However, an open question remains as to what the proximate triggers for Bogong moth migration are (Warrant et al., 2016). As well as being interesting in its own right, the answer to this question is rapidly becoming critical to the conservation of the unique Australian Alpine ecosystem, which accommodates many species that rely on the annual influx of nutrients brought by the Bogong moth migration (Gibson et al., 2018; Green, 2011, 2003). Concerningly, an estimated 200-fold reduction in the Bogong moth population was observed between the

2016–2017 and 2017–2018 summers, following a slow, but consistent decline since the early 1980s (Green et al., 2021; Mansergh et al., 2019). This has led to the recent listing of the Bogong moth as endangered on the IUCN Red List (Warrant et al., 2021).

The question of what proximate cues trigger Bogong moth migration is complex, and is unlikely to be solved by a single study. Behavioural experiments are laborious, and indeed, to our knowledge, a behavioural paradigm to measure the timing of a Bogong moth's migration in response to controlled stimuli has yet to be developed. In the meantime it therefore seems prudent to make quantitative measurements of Bogong moth migratory timing in the wild. This will at least enable us to determine what proximal factors are correlated with the behaviour, which will greatly assist in narrowing the search-space for future experimentation.

Useful progress to this end has been made through long-term monitoring of migrating insects using vertical radar deployed on the Bogong moth migratory route (e.g. Hao et al., 2020). However, reliable monitoring of Bogong moths in their breeding grounds remains an unsolved challenge (Wintle et al., 2021). At the end of their spring migration, a number of methods have been used to monitor Bogong moths close to their aestivation sites, including light trapping (Gibson et al., 2018; Wintle et al., 2021), light beam surveys, (Monk, 2021), aestivation site surveys (Caley and Welvaert, 2018; Green et al., 2021), ski surveys of Bogong moth carcasses on the snow (Green et al., 2021), and fox scat surveys (Green, 2010b; Green and Osborne, 1981). Each of these methods have their idiosyncrasies, and are to varying degrees laborious, limiting their utility for large-scale long-term monitoring programs, such as the 100-site Bogong moth monitoring program recommended by Wintle et al. (2021).

In this paper, we present the results of two years of monitoring of Bogong moth flight activity near aestivation sites in the Australian Alps using wildlife cameras, and a newly developed method described in Chapter 2. We show that by monitoring the sites for the full span of the Bogong moth aestivation season, we are able to infer the arrival and departure dates of the moths from those sites. Moreover, we are able to quantitatively analyse the evening twilight flight activity of the aestivating Bogong moths described by Common (1952) over the entire duration of the summer, providing strong preliminary evidence for weather being an important driver of flight behaviour—and by extension, migratory behaviour—in the moth. Our results, and indeed the method we have developed to obtain them, may be of interest to land managers and conservationists who seek to measure the ongoing effects of management practices on the Bogong moth population, and to monitor it more generally.

3.2 Methods

The methods employed in this study closely follow those described in Chapter 2, and are described briefly below. Weather data were obtained from weather stations close to the camera sites (Bureau of Meteorology, 2021).

3.2.1 Camera placement and settings

Study sites were selected for their proximity to known Bogong moth aestivation sites. In the first study season (2019–2020), cameras were placed outside a Bogong moth aestivation site in a boulder field near the summit of Mt Kosciuszko, NSW, and outside two aestivation sites near the formerly unnamed peak now known as Ken Green Bogong (referred to in this paper as K.G. Bogong), near South Rams Head, NSW. In the second study season (2020–2021), a site near the summit of Mt. Gingera, ACT/NSW, which has been subject to a number of previous studies (Caley and Welvaert, 2018; Common, 1954; Keaney et al., 2016) was added. Data obtained in November 2019 from a boulder field near Cabramurra, NSW, in Chapter 2 were also included for certain analyses.

During the 2019–2020 season, a single camera was also placed inside the aestivation cave on K.G. Bogong, facing towards a cluster of aestivating Bogong moths (referred to as the "observation cluster"). This camera was not used for automated annotation, although occasionally flying moths were seen inside the cave. By the end of the season this camera had been flooded and was no longer usable.

3.2.2 Image annotation

A total of 109912 images of the sky was obtained during this study. Of these, we manually annotated 33780 images for Bogong moths. Of the 33780 manually annotated images, 4223 contained at least one annotation. We kept 200 of the 4223 images as a test set, combining them with the test set used in Chapter 2. The remaining 4023 of these images were combined with the training set used in Chapter 2, for a total training set of 4901 images. The Camfi annotation model was retrained on this image set and the newly trained model was evaluated on the combined test set. The newly trained model has been included with a recent release of Camfi, which is available at https://github.com/J-Wall/camfi.

All images obtained in this study were then automatically annotated using Camfi with the newly trained model. Wingbeat frequencies of each annotation were then measured using Camfi. For further analyses, the automatically obtained annotations were filtered by prediction score, wingbeat SNR, and wingbeat frequency. In particular, annotations with prediction scores less than 0.8, wingbeat SNR outside of the range $[1, 50]$, or wingbeat frequency outside of the range $[27, 78]$ Hz were excluded.

3.2.3 Data analysis

Of primary interest is the relative daily abundance of flying Bogong moths at the study sites and across the study period. We measured this by counting moth detections occurring during evening twilight, noting the number of images collected at a given site during the evening twilight of a given day as the exposure variable.

Various daily abiotic factors were regressed against counts of Bogong moths detected by Camfi during evening twilight. Before performing the regression, the Pearson correlations between each pair of factors were calculated, and highly correlated factors were removed using a greedy recursive algorithm. The algorithm proceeded by selecting the most highly correlated pair of factors, then removing the factor in the pair which was less well correlated with the evening detection count. The algorithm terminates when no pair of factors had a Pearson R^2 greater than a specified threshold, which we set to 0.3. The remaining factors were then jointly regressed against the evening detection count using a Poisson regression with image count as the exposure variable.

3.3 Results

An evaluation of the performance of the newly trained annotation model is presented in Appendix B.1. Overall, every evaluation metric marginally improved with respect to the previous Camfi annotation model (Chapter 2), with the exception of average precision, which slightly worsened. This is presumably a consequence of the present dataset containing images taken in a wider variety of lighting conditions.

Strong peaks in activity were observed during evening twilight across all study sites (Fig. 3.1, *right panel*), although pronounced peaks were not seen during morning twilight (Fig. 3.1, *left panel*). Evening twilight detection counts were highly variable, but clearly show that Bogong moths departed from the lower elevation sites (Mt. Gingera and K.G. Bogong) earlier in the season than from higher elevation sites, i.e. Mt. Kosciuszko (Fig. 3.2 for the 2019–2020 summer and Fig. 3.3 for the 2020–2021 summer). The camera placed at the K.G. Bogong site fell from its mount towards the end of the season (Fig. 3.3, *lower panel, shaded region*), reducing the camera's view of

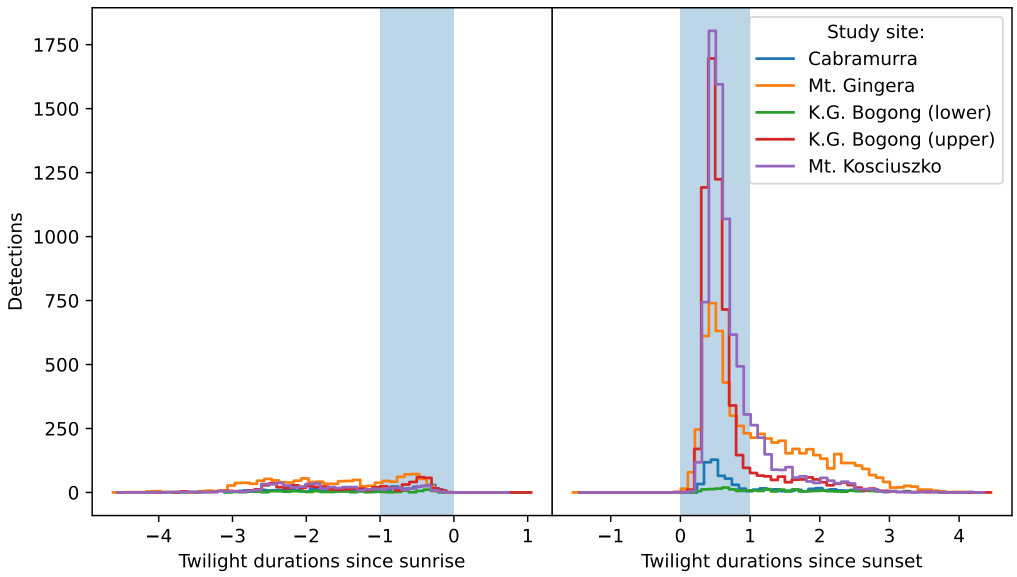


Figure 3.1: Total number of moth observations by time relative to sunrise (*left*, scaled by the duration of morning twilight) does not show peak in activity during morning twilight (*blue shaded region*), with the slight exception of K.G. Bogong (upper) and Mt Gingera sites, which show small peaks. Total number of moth observations by time relative to sunset (*right*, scaled by the duration of evening twilight) shows peak in activity during evening twilight (*blue shaded region*) across all study sites. Data from Cabramurra boulder field site is from Chapter 2.

the sky by about half, however it appears that the majority of moths had already left the area by the time this happened.

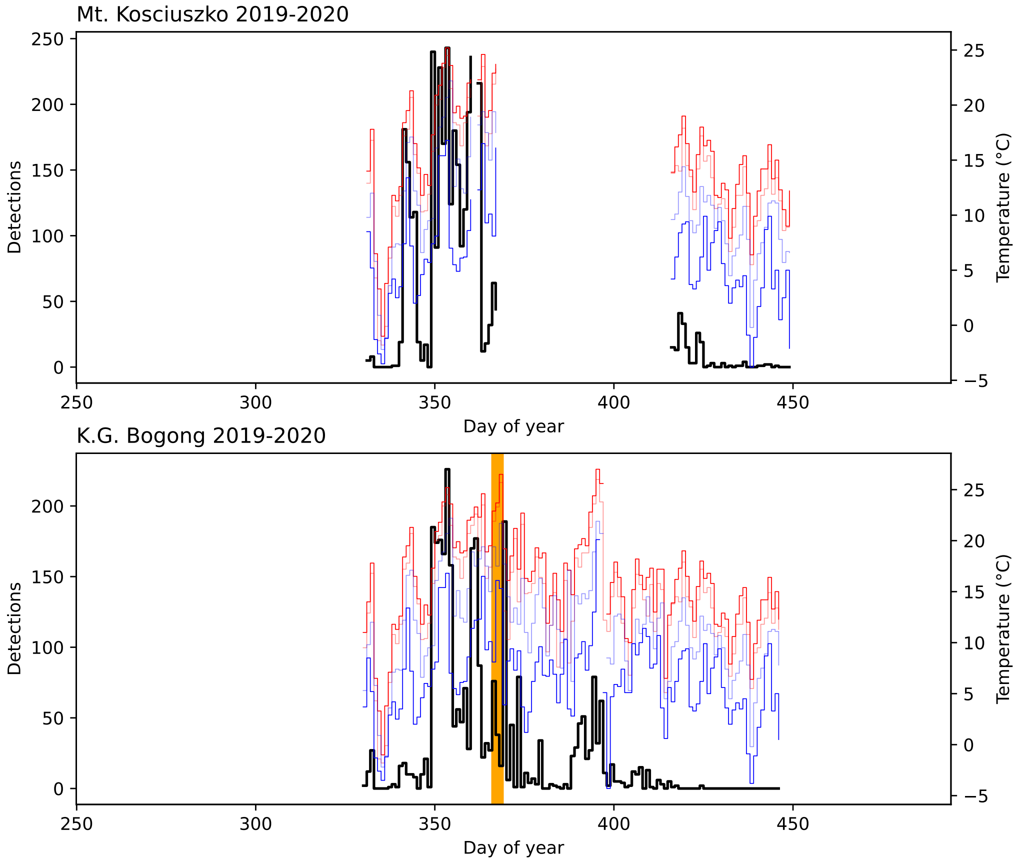


Figure 3.2: Number of Bogong moth detections (*black*) for each study day in the 2019–2020 summer season outside Bogong moth aestivation sites on Mt. Kosciuszko and K.G. Bogong, NSW, shown with daily temperatures recorded at Thredbo Top Station (Bureau of Meteorology, 2020): maximum (*red*), minimum (*blue*), 9 am (*light blue*), and 3 pm (*light red*). Data are missing for a portion of the season at the Mt. Kosciuszko site due to a camera malfunction. *Orange span* indicates a bushfire event which occurred 1 km SW of K.G. Bogong (the fire did not reach the site, although there was high levels of smoke in the air which would have entered the site).

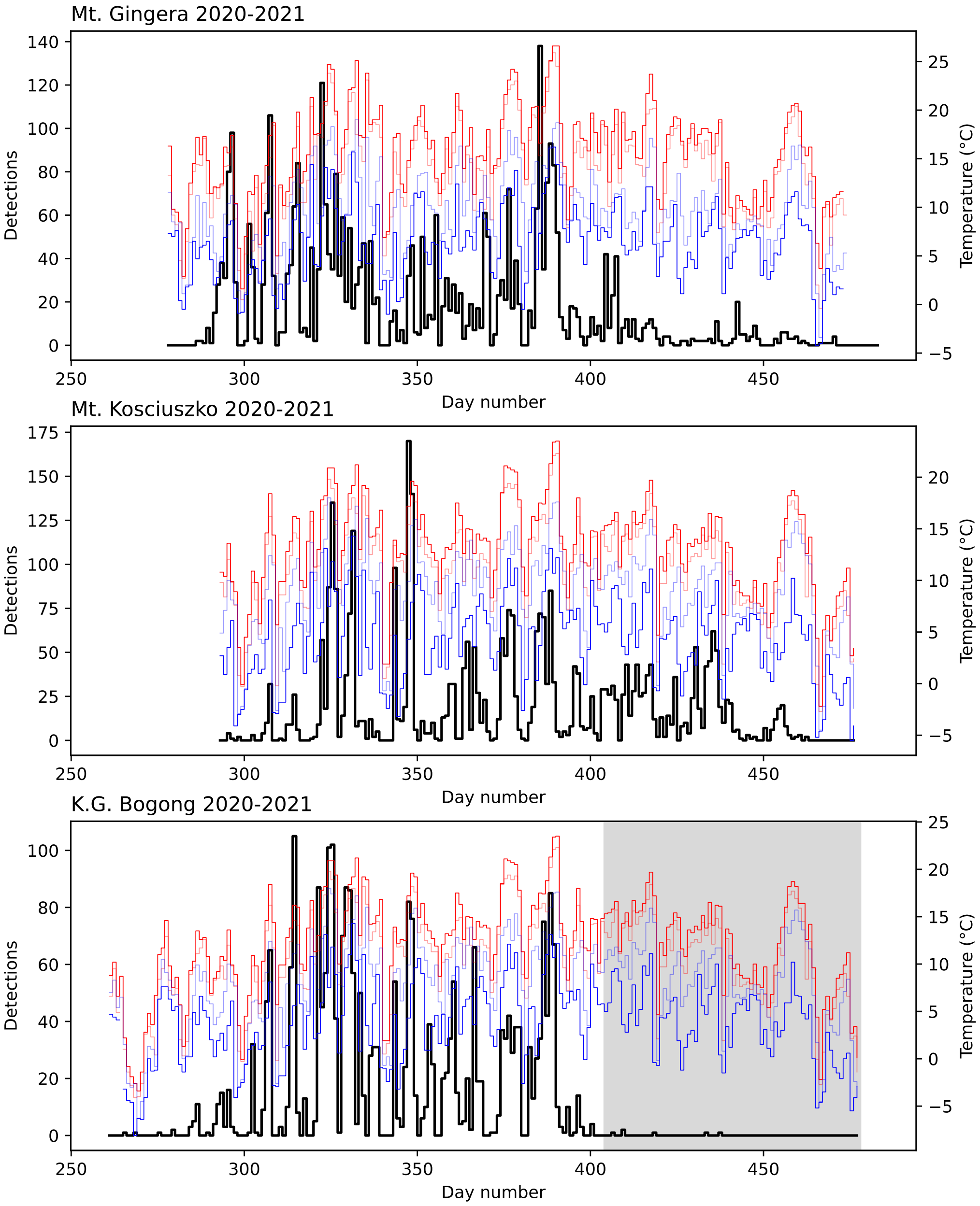


Figure 3.3: Number of Bogong moth detections (*black*) for each study day in the 2020–2021 summer season outside Bogong moth aestivation sites on Mt. Gingera, Mt. Kosciuszko, and K.G. Bogong, NSW, shown with daily temperatures recorded at Mt Ginini and Thredbo Top Station (Bureau of Meteorology, 2021): maximum (*red*), minimum (*blue*), 9 am (*light blue*), and 3 pm (*light red*). *Shaded region* on lower plot indicates period where camera had fallen from its mount, reducing its view of the sky by about half.

3.3.1 Predictors of activity

A total of ten abiotic factors were found to be significantly correlated with evening twilight counts of flying Bogong moths (Fig. 3.4a). A greater number of moths were observed at higher elevation sites (Fig. 3.4b) and when twilight duration was longer (i.e. in the middle of summer, Fig. 3.4d). Study year was also positively correlated with moth counts, suggesting that the Bogong moths were more abundant in the 2020–2021 summer than in the 2019–2020 summer. The most important weather factors were daily maximum temperature (which had a positive effect on moth counts, Fig. 3.4c) and maximum wind speed (which had a negative effect on moth counts, Fig. 3.4e). Note that very few Bogong moths were observed flying on days which had maximum temperatures lower than 10°C (Fig. 3.4c). Daily temperature range, relative humidity (measured at 9 am), and daily minimum temperature were negatively correlated with moth counts, while latitude and rainfall were positively correlated with moth counts. Scatter plots of all covariates in our model are shown in Fig. B.2, and residuals of the fitted model are shown in Fig. B.3 (both in Appendix B.2).

3.3.2 Arrival and departure of Bogong moths

During the 2020–2021 summer season, the cameras were placed at the aestivation sites before the Bogong moths had arrived, and removed after they had left. This means the detection data obtained from those cameras contain information regarding the arrival and departure dates of the Bogong moths. For example, at Mt. Gingera, the first Bogong moth was detected on the 13th October (Fig. 3.3, *top panel, day 286*). At K.G. Bogong, this

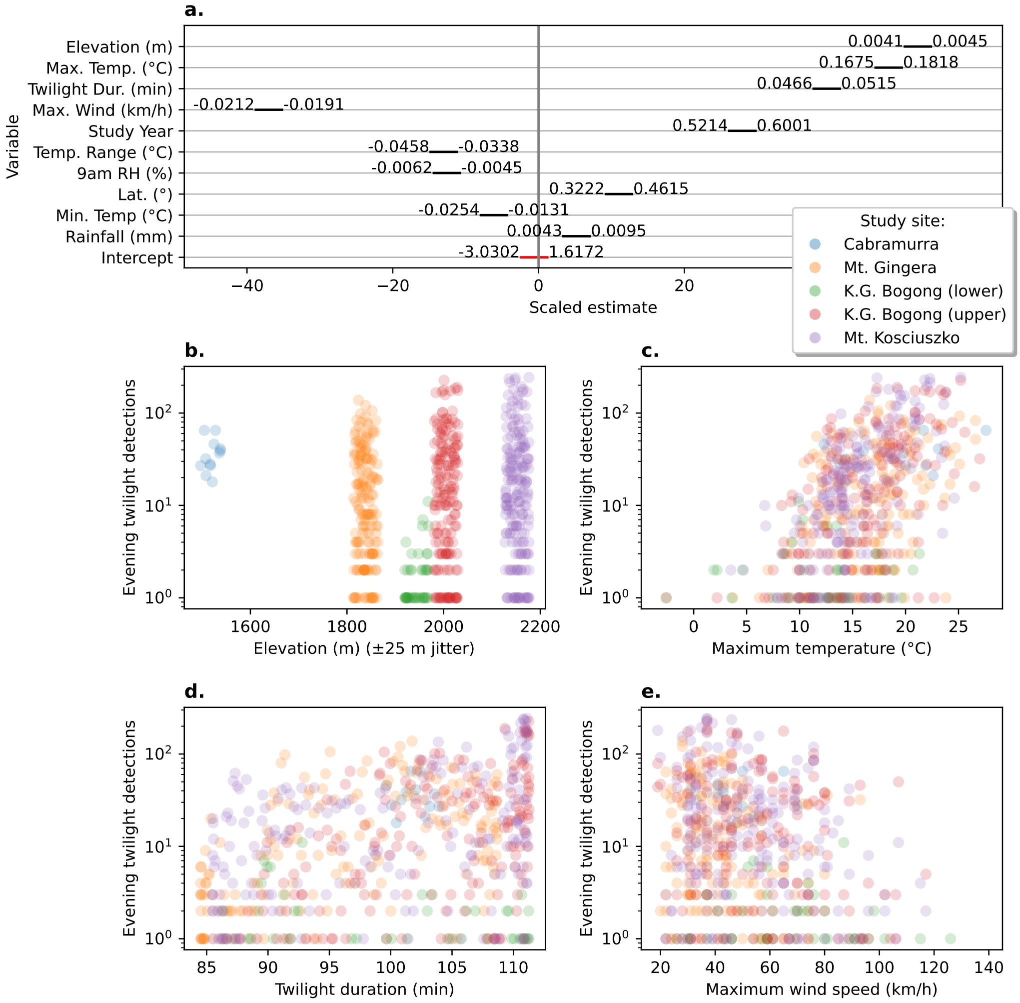


Figure 3.4: Effect-sizes and plots of number of detections during evening twilight against significantly associated abiotic factors. **a.** Scaled estimates of effect size of abiotic factors on Bogong moth evening flight intensity (as measured by number of Camfi detections) from a mixed-effect Poisson generalised linear model of detections against these factors. *Black bars* show 95% confidence interval of estimates (scaled by effect size). Values to either side of *black bars* represent bounds of 95% confidence interval in the units of the respective factor (corresponding to the gradient of the regression, in that dimension). Negative values indicated that increases in the value of the factor lead to a decrease in moth counts (and positive values, the opposite). **b.** Scatter plot of detections per evening twilight by elevation. Random fluctuation (jitter) is applied to elevation to increase readability. **c.** Scatter plot of detections per evening twilight by daily maximum temperature. **d.** Scatter plot of detections per evening twilight by duration of evening twilight. **e.** Scatter plot of detections per evening twilight by daily maximum wind gust speed. Points in *b*–*e* are coloured by study site, as per study-site key (*right, towards top*).

date was 11th October (Fig. 3.3, *bottom panel, day 284*). A few detections were made prior to this date, however upon inspection these were found to be false-positives caused by rain. The first detection of a Bogong moth at Mt. Kosciuszko was on 22nd October (Fig. 3.3, *middle panel, day 295*) although cameras were only placed there on 20th October, so it is possible that some moths had arrived earlier. However, snow was still present on the ground in front of the aestivation site on Mt. Kosciuszko until it melted on 23rd October (JRAW, personal observation), so if there was an earlier arrival, it was probably only by a few days.

While records of earliest arrival are interesting, they are also subject to substantial noise, owing to the fact that the marginal probability of a particular moth being detected at all is very small. Earliest arrivals are also not necessarily representative of the predominant behaviour in the population. Therefore, we would like to use the detection data across the entire season to model the arrival and departure of the majority of the population. To do this, we propose a simple heuristic model of the evolving relative abundance of evening-flying Bogong moths in an area, as the moths arrive at—and later depart from—the area (Fig. 3.5). This model is based purely on detection data, and is independent of weather factors, etc.

The model separates the aestivation of Bogong moths at a particular site into two phases; arrival and departure. In the arrival phase (Fig. 3.5, *solid lines*), the relative abundance of aestivating moths is modelled by the cumulative maximum of the mean number of detections per image over all preceding evenings. The arrival phase ends on the day where this value reaches its maximum across the entire summer. The departure phase (Fig. 3.5,

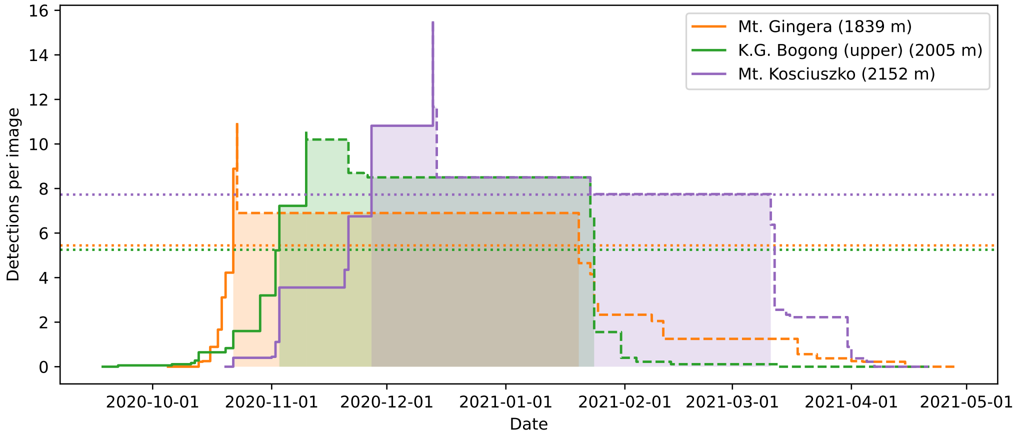


Figure 3.5: A simple heuristic model of the arrival and departure of Bogong moths to summer aestivation sites applied to data obtained from automated camera monitoring in the summer of 2020–2021. *Solid lines:* Cumulative maximum detections per image, plotted until the date that absolute maximum is reached, for the respective location. This roughly models the arrival of moths to the location. *Dashed lines:* Reverse-cumulative maximum of detections per image, plotted from date of absolute maximum, for the respective location. This roughly models the sum of departure and mortality of moths from the location. *Dotted lines:* Show half of the maximum detections per image for respective location (for calculating median date of arrival and departure). Dates after median date of arrival, and before median date of departure for each location are *shaded*. Elevations shown are of the camera placement, rather than the summit elevations of the mountains.

dashed lines) is modelled similarly, this time using the reverse-cumulative maximum.

An obvious set of descriptive statistics arise; namely, median date of arrival and median date of departure. These are shown for each study site in the 2020–2021 season in Fig. 3.5 (*shaded areas*) and in Table 3.1. A clear signal of Bogong moths arriving at higher elevation aestivation sites later than lower elevation sites is present (Fig. 3.5, *solid lines*; Table 3.1). This trend appears to also apply to departures, albeit slightly less clearly (Fig. 3.5, *dashed lines*; Table 3.1).

To assess whether temperature could explain the later arrival and departure of Bogong moths at higher elevations, we calculated the 3-day average maximum temperature at each location in the lead-up to the median arrival and departure (Table 3.1). Notably, temperatures were relatively high at lower elevation sites (Mt. Gingera: 20.1°C, K.G. Bogong: 17.3°C) in the days leading up to the median *arrival* date at the higher-elevation Mt. Kosciuszko site (2020-11-27; Table 3.1). Also, temperatures in the lead-up to median *departure* dates at the lower elevation sites were relatively high, respectively (Mt. Gingera: 18.4°C, K.G. Bogong: 20.7°C), while pre-departure temperatures at Mt. Kosciuszko were comparatively cool (13.0°C; Table 3.1).

Table 3.1: Median date of arrival ($A_{1/2}$) and departure ($D_{1/2}$) of Bogong moths from aestivation sites during 2020–2021 summer. Elevations shown are of the camera placement, rather than the summit elevations of the mountains. 3-day average maximum is calculated across the 3 days preceding the date listed (inclusive), from the nearest weather station (Bureau of Meteorology, 2020) assuming an adiabatic lapse rate of 9.1°C/1000 m elevation (Green, 2014).

Date	Mt. Gingera (1839 m)	K.G. Bogong (2005 m)	Mt. Kosciuszko (2152 m)
Mt. Gingera $A_{1/2}$ (2020-10-22)	14.0°C (std = 2.0°C)	11.2°C (std = 1.7°C)	9.8°C (std = 1.7°C)
K.G. Bogong $A_{1/2}$ (2020-11-03)	14.0°C (std = 2.5°C)	14.0°C (std = 4.1°C)	12.7°C (std = 4.1°C)
Mt. Kosciuszko $A_{1/2}$ (2020-11-27)	20.1°C (std = 2.4°C)	17.3°C (std = 1.9°C)	16.0°C (std = 1.9°C)
Mt. Gingera $D_{1/2}$ (2021-01-20)	18.4°C (std = 2.2°C)	16.1°C (std = 0.9°C)	14.7°C (std = 0.9°C)
K.G. Bogong $D_{1/2}$ (2021-01-24)	24.3°C (std = 1.6°C)	20.7°C (std = 2.3°C)	19.4°C (std = 2.3°C)
Mt. Kosciuszko $D_{1/2}$ (2021-03-11)	16.4°C (std = 1.4°C)	14.4°C (std = 1.5°C)	13.0°C (std = 1.5°C)

3.3.3 Impact of January 2020 bushfire

On 4th January 2020, a major bushfire which had been burning in the area in the preceding few days (Fig. 3.2, *orange span*) came within 1 km of the K.G. Bogong site. Despite the thick smoke, Bogong moths were seen flying

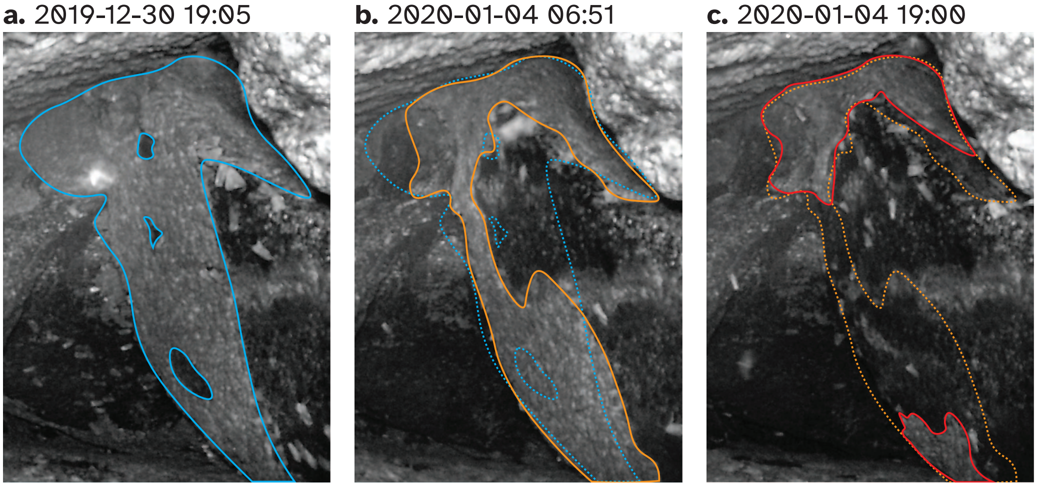


Figure 3.6: Progression of cluster of aestivating Bogong moths in cave on K.G. Bogong over the worst few days for Kosciuszko National Park during the 2019–2020 bushfire season. **a.** Prior to bushfire. Cluster is outlined with *solid blue trace*. **b.** Morning of 4th January 2020, the day which saw the bushfire come within 1 km of the site. Cluster of Bogong moths is outlined with *solid orange trace*. Trace from *a* is overlaid for comparison (*dotted blue trace*). **c.** Evening of 4th January 2020. Cluster of Bogong moths is outlined with *solid red trace*. Trace from *b* is overlaid for comparison (*dotted orange trace*). Times shown are Australian Eastern Daylight Time (AEDT; UTC+11:00).

outside their aestivation cave (Appendix B.3 Fig. B.4). The following day, a large number of flying Bogong moths were detected, presumably indicating a departure of a portion of the moths from the site (Fig. 3.2, peak to the right of *orange span*).

A reduction in the number of aestivating Bogong moths on K.G. Bogong during the bushfire was reflected by our observation cluster, which dramatically reduced in size over the course of the fire (Fig. 3.6). Notably, a significant portion of this reduction happened during the day (Fig. 3.6b–c), despite Bogong moths typically being night-active. The remaining cluster (Fig. 3.6c) did not change much during the following few weeks before the camera was flooded on 20th January.

3.4 Discussion

Our results clearly and quantitatively demonstrate that the summer flights of Bogong moths described by Common (1954) occur predominantly during evening twilight, and occur throughout the Bogong moth's entire summer aestivation. Furthermore, the intensity of these flights is modulated by daily weather factors, with Bogong moths favouring warmer evenings with lower wind speeds for flying, confirming that the patterns seen in Chapter 2 hold over the entire summer.

The influence of weather on animal migration *in general* has been well studied (see review by Shamoun-Baranes et al., 2010). Broadly, our results agree with those of previous studies comparing weather with flight activity of migratory insects. Namely, wind speed has a negative effect on flight activity (e.g. Gregg et al., 1994), and temperature has a positive effect (e.g. Chapman et al., 2002; Krauel et al., 2015).

For simplicity's sake, we chose to use a relatively simple linear model relating moth counts to various abiotic factors. If long-term monitoring of Bogong moths continues using our camera-based method, and an increasing number of years of data become available, more complex models will become more appropriate. For example, we have modelled our study year as having a linear effect on moth counts. For a two-year study, this is valid, however when additional years are added, this should be changed to a random effect (or perhaps an effect depending on annual climactic factors, depending on the research question).

Our survey of three known Bogong moth aestivation sites over the summer of 2020–2021 shows that occupation of higher elevation aestivation sites by Bogong moths occurs later in the season than lower elevation sites. There are three possible explanations for the later arrival dates at the higher elevations: 1) The higher sites are blocked by snow (or are otherwise unsuitable) earlier in the season, but are open later to allow occupation by later arrivals of Bogong moths coming directly from the breeding grounds, 2) Bogong moths from lower sites move higher as summer progresses, or 3) a combination of 1 and 2.

Blockage of high-elevation aestivation sites by snow (possibility 1) would certainly prevent Bogong moths from occupying those sites, but this does not appear to be a satisfactory explanation for the delay we observed, as most of the remaining snow near the highest site (Mt. Kosciuszko) melted on 23rd October 2020, more than an entire month prior to the median arrival of Bogong moths at this location. Additionally, sub-zero temperatures have been recorded in occupied Bogong moth aestivation sites (Green et al., 2021), so it seems unlikely that low temperatures alone would have *prevented* Bogong moths from migrating directly to high-altitude sites early in the season.

On the other hand, high temperatures do appear to be a reasonable explanation for Bogong moths *avoiding* lower elevation sites, which could motivate movement to higher elevations (possibility 2 or possibility 3). The median arrival date at the highest elevation site (Mt. Kosciuszko) coincided with lower elevation sites experiencing 3-day average maximum temperatures above 16°C (Table 3.1). Incidentally, Green et al. (2021) estimated that 16°C is the maximum temperature (inside a cave) that permits aestivation. Interestingly, 3-day average maximum temperatures leading up to the median *departure* dates at Mt. Gingera and K.G. Bogong were 18.4°C and 20.7°C at those locations, respectively (i.e. well above 16°C; Table 3.1). However, the 3-day average maximum temperature at Mt. Kosciuszko leading up to its median *departure* date was lower, at just 13.0°C. Therefore, it could be that departures from the Mt. Gingera and K.G. Bogong sites were motivated by high temperatures (and resulted in movements to higher elevations), while departures from Mt. Kosciuszko were motivated by temperatures falling (to 13°C), indicating the approaching autumn, thus triggering the return migration to the breeding grounds.

Notably, possibility 2 is also supported by previous results from fox scat surveys (Green, 2010b), which showed a departure of Bogong moths from sub-alpine areas into alpine areas as the summer progressed. However, properly disentangling each of these possibilities requires observations of the movements (or lack thereof) of Bogong moths between elevations during the summer, after their arrival in the mountains. Such observations could be made using a similar method to that used in this study, with cameras deployed in elevation transects on a single mountain.

Our simple model of Bogong moth arrival and departure is robust to periods of evening-flight inactivity (e.g. due to unfavourable weather conditions for flight), and follows naturally from the following two heuristics: 1) the maximum relative density of flying moths in the vicinity of an aestivation site is representative of the relative abundance of aestivating moths in the area, and 2) most Bogong moths arrive at and leave from the vicinity of an aestivation site on relatively few nights (so evening flights with lower relative density are generally station-keeping movements, rather than migrations to other aestivation sites or returning to the breeding grounds). As with the model for relating evening flights to abiotic factors, the arrival and departure model could perhaps be extended to include—depending on the research question—the effects of other factors, such as daily weather and annual climate.

Fire appears to have a pronounced effect on assemblages of aestivating Bogong moths. We observed a marked reduction in the size of our observation cluster of aestivating Bogong moths during the day that a bushfire came close to the site (but did not affect it directly). Presumably this was mediated by smoke entering the cave and disturbing the moths. Interestingly, from flight data, we observed (what we assume to be) a large departure of Bogong moths from a bushfire-affected area the day *after* the fire, and thus the day after the marked reduction in the size of the observation cluster. It could be that the reduction of the observation cluster on the day of the fire was caused by Bogong moths falling from their perch on the cave wall, but remaining inside the cave, rather than perishing or departing that day. These moths could then have departed the following day when conditions outside were less dangerous.

In recent years, especially since their dramatic population crash in 2017 (Green et al., 2021; Mansergh et al., 2019), Bogong moths have enjoyed increased attention from those interested in their conservation, and in the conservation of the Australian Alpine ecosystem more generally. In particular, the need for the implementation of a long-term monitoring program for Bogong moths for the ongoing conservation of the Australian Alpine ecosystem has been identified (Wintle et al., 2021).

The high level of variability in counts we observed across each night of our study highlights the importance of regular measurements throughout the summer for such a monitoring program. Not only does the proportion of Bogong moths flying on a given night depend on the weather, but the Bogong moth population also moves between aestivation sites over the course of the summer, complicating the interpretation of sparse data collected from infrequent light-trapping surveys, particularly when these surveys do not simultaneously collect counts from multiple locations.

Ideally, a long-term monitoring program for Bogong moths would collect counts of moths every day at every study site from mid-September until mid-May, completely covering the Bogong moth aestivation season. For a large-scale program with many study sites, light trapping or similarly labour-intensive approaches would be a costly undertaking. Conversely, a comparatively “hands-off” and non-invasive approach (i.e. not involving trapping), such as the automated camera-based monitoring method used in this study, could be relatively easily and inexpensively scaled, without the need for a large team of dedicated light-trappers. For instance, by replacing the wildlife cameras used in this study with permanent solar-powered and Internet-connected camera stations, the data acquisition portion of a large-scale Bogong moth monitoring program could be completely automated, with visits to study sites only needed for placement, maintenance, and eventual retrieval of the equipment. Such a program would produce an incredibly rich and informative dataset for ongoing efforts to model the dynamics of the Bogong moth population, and their migration to and from the Australian

Alps.

Finally, one could also imagine a wide variety of possible study designs using the method, as it provides the opportunity to trivially increase the number of cameras at each site, or change the frequency of image captures to address specific research questions surrounding the behaviour of the moths.

3.5 Data accessibility

The software used for the analyses is archived and available from https://doi.org/10.5281/zenodo.5242596. The raw data (images) are available on Zenodo: Cabramurra 2019 dataset: https://doi.org/10.5281/zenodo.4950570, Ken Green Bogong 2019–2020 dataset: https://doi.org/10.5281/zenodo.4971714, Mt Kosciuszko 2019–2020 dataset: https://doi.org/10.5281/zenodo.5039891, Ken Green Bogong 2020–2021 dataset: https://doi.org/10.5281/zenodo.4972022, Mt Kosciuszko 2020–2021 dataset: https://doi.org/10.5281/zenodo.5040011, and Mt Gingera 2020–2021 dataset: https://doi.org/10.5281/zenodo.5040018. Additional supplementary data (e.g. configuration files and annotation files) are available from https://doi.org/10.5281/zenodo.6583127.

3.6 Author contributions

EJW, DD, and JRAW conceived the project. JRAW, TR, BMH, DD, EJW, and KG performed the fieldwork. KG provided essential knowledge of the locations of Bogong moth aestivation sites. TR, LK, BMH, and JRAW performed the manual annotations of images of flying moths. JRAW analysed the data. JRAW wrote the first draft of the manuscript, with input from EJW.

3.7 Acknowledgements

EJW and JRAW are grateful for funding from the European Research Council (Advanced Grant No. 741298 to EJW), and the Royal Physiographic Society of Lund (to JRAW). JRAW is thankful for the support of an Australian Government Research Training Program (RTP) Scholarship. EJW holds Scientific Permits for collection and experimental manipulations of Bogong moths in several alpine national parks and nature reserves (NSW Permit SL100806). We are extremely grateful to Dr. Peter Caley for useful discussions, and to Dr. Stanley Heinze for essential collaboration and providing computational resources.

Chapter 4

Oriented evening flight behaviour in the Bogong moth revealed through automated video tracking

Jesse R A Wallace[1,2], David Dreyer[2], Jochen Zeil[1] and Eric J Warrant[1,2]

[1]Research School of Biology, Australian National University. [2]Lund Vision Group, Department of Biology, Lund University, Sweden.

Abstract

During their period of summer dormancy, Australian Bogong moths *Agrotis infusa* undertake seemingly random evening flights, filling the air with densities in the dozens per cubic metre. The purpose of these flights is unknown, but they may serve an important role in Bogong moth navigation, which remarkably enables them to return to the same exact summer sites—generation after generation—after migrating around 1000 km, and with no opportunity to learn their route or destination from prior generations. The recent development of the camera-based insect monitoring method, Camfi, enables quantitative observations of Bogong moth behaviour at an unprecedented scale. To gain a better understanding of the summer evening flights of Bogong moths, we have extended Camfi to facilitate automated video tracking of flying insects, taking the already-high throughput of the method to a new level. We used this new method to record the evening flight behaviour of Bogong moths in two elevational transects below the summit of Mt. Kosciuszko, NSW, on a single night in February 2021, and found that these flights were not random, but were systematically oriented in directions relative to the azimuth of the summit of the mountain. These results stimulate interesting and plausible hypotheses relating to previously unexplained summer evening flight behaviour of Bogong moths, and the mechanisms of their long-distance navigation.

4.1 Introduction

During their period of summer dormancy, known as aestivation, Australian Bogong moths *Agrotis infusa* remain huddled within cool, dark crevices of granite outcrops that dot the peaks of the Australian Alps, tiling the walls with an estimated density of up to 17,000 moths per square metre (Common, 1954). However, during evening twilight, the moths are known to emerge from their hiding spots and undertake seemingly random flights (Common, 1954; Wallace et al., 2021; Warrant et al., 2016). Although these flights are only undertaken by a portion of the moths at a particular site, they are enough to fill the air with densities probably reaching dozens per cubic metre (Wallace, personal observations).[1] The purpose of these flights is unknown, although observations have been made of Bogong moths using them to visit water to drink (Common, 1954; Warrant et al., 2016).

It could be that the evening flights of aestivating Bogong moths are used as a sort of learning flight to calibrate their navigational machinery, akin to how homing insects familiarise themselves with an area of interest (Collett and Zeil, 2018), or how night-migratory birds calibrate their star compasses and other compass systems prior to migration (reviewed by Foster et al., 2018; Pakhomov and Chernetsov, 2020). Alternatively, one might hypothesise that these flights are undertaken by moths who are dissatisfied with their resting place—perhaps being too warm or not dark enough—and are seeking a more favourable site in which to continue their aestivation. The former possibility may be necessary, particularly as the departure date

---

[1] To illustrate the point, a note from JRAW's field book recounting one of these evening flight events reads, "I decided to see if it was enough to simply reach out an open hand and close it into a fist in order to catch a moth. It worked—first try."

for the return migration draws near. We might expect the latter possibility to play a more important role during the first few months of aestivation, as Bogong moths are known to occupy higher and higher elevation sites as the summer progresses (Green, 2003). Disentangling these possibilities is important for our understanding of the mechanisms of Bogong moth navigation, which remarkably enable them to return to the exact same summer sites—generation after generation—following a migration of around 1000 km, having had no opportunity to learn their route or destination from prior generations.

The recent development of Camfi, a camera-based system for monitoring evening flight behaviour in wild Bogong moths (Chapter 2), presents an opportunity to make quantitative observations of the dynamics of the behaviour with an unprecedented scale and spatiotemporal resolution. In this paper, we demonstrate how observations made using Camfi can begin to disentangle the causes of Bogong moth evening flights, and we provide evidence of directed flights undertaken by the moths near aestivation sites, giving clues as to the purpose of the flights.

In order to measure the direction of flight, we have extended Camfi to facilitate automated video tracking of insects flying above the camera. Full details of this new method are presented in this paper, and the method has been implemented as part of the Camfi package, freely available from https://github.com/J-Wall/camfi.

4.2 Methods

4.2.1 Detection of flying Bogong moths using Camfi

Chapter 2 introduced Camfi, a method for monitoring the activity of flying insects using still images obtained from off-the-shelf wildlife cameras. Camfi has been used to monitor the activity of migratory Bogong moths arriving at, aestivating in, and departing from their summer range in the Australian Alps (Chapter 3). Camfi performs automatic detection of flying insects using the Mask R-CNN framework (He et al., 2017), and at the time of writing, has been trained on a set of 4901 manually annotated images of flying Bogong moths (Chapter 3). In addition, Camfi automatically measures wingbeat frequency of detected insects in still images, which is useful for assigning species identity to observations of flying insects.

An advantage of Camfi is its flexibility with regard to the temporal resolution of data collection. Depending on the research question, cameras can be set to capture an image at relatively long intervals, on the order of minutes, or they can be set to capture images at a very high rate, which in the case of video clips is on the order of hundredths of a second (typically 25-30 frames per second). However, when analysing Camfi data which have been obtained from high-rate captures (namely, videos), individuals will be detected multiple times, since each moth will be seen in each of many consecutive video frames as they pass by the camera. This results in detection counts being inflated by insects which have lower angular velocities relative to others from the perspective of the camera, and therefore spend more time in-frame. Therefore, to facilitate the use of videos by Camfi, we need to be able to track observations of individuals in a sequence of video frames, so we can count each individual only once.

In the following sections, we introduce an extension to Camfi which enables analysis of video data. This includes proper handling of video files, as well as tracking of individuals through consecutive frames. In addition to ensuring individual insects are only counted once per traversal of the camera's field of view, the new method allows for measurement of the direction of displacement of insects as they travel through the air.

4.2.2 MULTIPLE OBJECT TRACKING

Multiple object tracking is a challenging problem which arises in many computer vision applications, and which has been approached in a variety of different ways (reviewed by Luo et al., 2021).

A common approach to multiple object tracking is "detection-based tracking" (also known as "tracking-by-detection"), in which objects are detected in each frame independently, and then linked together using one of a number of possible algorithms. Typically, this requires the use of a model of the motion of the objects to be tracked, along with a method which uses the model of motion to optimise the assignment of detections to new or existing trajectories. In many approaches, the modelled motion of tracked objects is inferred by combining information about the position of the objects in multiple frames. An obvious challenge arises here because the model of motion requires reliable identity information of objects detected in multiple frames, whereas the identity of the objects usually must be inferred from their motion (including their position). This circular dependency—between the inference of object identity and the model of motion—can be dealt with in a number of ways, including probabilistic inference via a Kalman filter (Reid, 1979) or a particle filter (e.g. Breitenstein et al., 2009), or through deterministic optimisation using a variety of graph-based methods.

Our approach to multiple object detection removes the requirement of an explicit model of motion entirely, by utilising two peculiar properties of the Camfi object detector. The first of these properties is that the Camfi detector obtains information about the motion of the flying insects it detects from the motion blurs the insects generate, which it stores in the form of a polyline annotation.[2] Since this information is obtained from a single image, and therefore a single detection, it does not depend on the identity of the insect, solving the previously mentioned circular dependency problem. The second property is that the Camfi detector is robust to varying exposure times, owing to the fact it has been trained on images with a variety of exposure times. This in turn means that the detector is robust to the length of the insects' motion blurs. Ultimately, these two properties, along with the fact that the insects appear as light objects on a dark background, mean that it is possible to use the Camfi detector to make a single detection of an individual insect traversing multiple consecutive frames. Thus, the trajectories can simply be formed using bipartite graph matching of overlapping polyline annotations, using only information provided by the detections themselves, using the method described in Section 4.2.3.

---

[2] This information does not extend to the direction the insect is flying with respect to the camera, but it does include the orientation of flight (with 180° ambiguity).

4.2.3 Automated flying insect tracking

The algorithm described in this section has been implemented in Python, and is packaged together with the Camfi software, available under the MIT licence from https://github.com/J-Wall/camfi.

We will now describe our algorithm for tracking flying insects in short video clips. The algorithm uses a detection-based tracking paradigm, relying heavily on the Camfi flying insect detector described in Chapter 2. Accordingly, we will not present a detailed explanation of the detector here, but will instead focus on the process of linking detections into trajectories. Readers interested in the details of the detector should consult Chapter 2.

An example of the sequence of steps taken by the tracking algorithm described in this section is illustrated in Fig. 4.1. For brevity, the example shows the algorithm operating on three frames only, however the algorithm can operate on any number of frames, up to the memory constraints of the computer it is running on.

First, the video frames are prepared for flying insect detection. A batch of frames is loaded into memory (e.g. Fig. 4.1a–c, although typically this would be a video clip). The maximum image of each sequential pair of frames is then calculated by taking the maximum (brightest) value for each pixel between the two frames (Fig. 4.1d–e). This produces images with lengthened motion blurs of the in-frame flying insects, approximating the images which would be obtained if the exposure time of the camera were doubled. Importantly, the motion blurs of an individual insect in consecutive time-steps overlap each other in these maximum images.

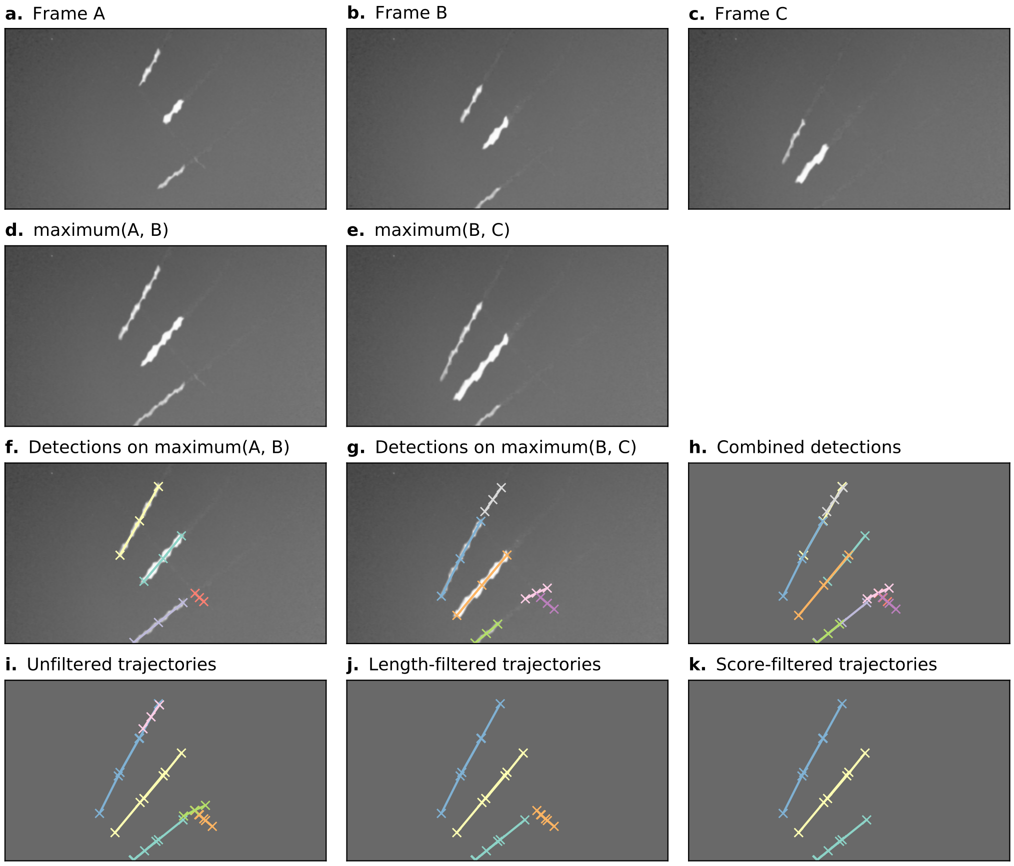


Figure 4.1: Automatic annotation is performed by Camfi on the maximum image of each pair of consecutive frames, allowing trajectories to be built from overlapping detections. Here, an example of this process is shown for three consecutive video frames. **a–c.** Three consecutive video frames containing multiple flying insects. **d–e.** The maximum image of each sequential pair of frames. **f–g.** Flying insects are detected in the two-frame maximum images using Camfi. **h.** Detections from **f** and **g** together on a plain background. **i.** Detections from sequential time-steps are combined into trajectories using bipartite graph matching on the degree of overlap between the detections. **j.** Trajectories containing fewer than three detections are removed. **k.** Finally, trajectories are filtered by mean detection score (trajectories with mean detection score lower than 0.8 are removed).

Detection of flying insects is performed on the maximum images using the Camfi detector (Chapter 2), producing candidate annotations of insect motion blurs to be included in trajectories (Fig. 4.1f–g). The Camfi detector produces polyline annotations which follow the respective paths of the motion blurs of flying insects captured by the camera. Because the motion blurs of individual insects overlap in consecutive frames, so too do the annotations of those blurs (e.g. Fig. 4.1h). This enables the construction of trajectories by linking overlapping sequential detections.

Detections in successive time-steps are linked by solving the linear sum assignment problem using the modified Jonker-Volgenant algorithm with no initialisation, as described by Crouse (2016). In order to do this, a formal definition of the cost of linking detections is required. We call this cost the "matching distance", which we denote by d_M. Consider two polyline annotations P_a and P_b, which are sequences of line segments defined by the sequences of vertices $(a_i)_{i=0}^{n-1}$ and $(b_j)_{j=0}^{m-1}$, respectively, where $a_i, b_j \in \mathbb{R}^2$. We define $d_M(P_a, P_b)$ as the second smallest element in $\{d(a_0, P_b),\ d(a_{n-1}, P_b),\ d(b_0, P_a),\ d(b_{m-1}, P_a)\}$, where $d(x, P)$ is the Euclidean distance from a point $x \in \mathbb{R}^2$ to the closest point in a polyline $P \subset \mathbb{R}^2$. This definition of d_M is efficient to compute, and allows us to discriminate between pairs of detections which come close to each other by chance (perhaps at very different angles) and pairs of detections which closely follow the same trajectory (i.e. roughly overlap each other).

After solving the assignment problem, a heuristic is applied to reduce spurious linking of detections into trajectories, where links with d_M values above a specified threshold are removed. Trajectories are built across the entire batch of frames by iteratively applying the detection linking procedure for each consecutive pair of time-steps (Fig. 4.1i). Trajectories containing fewer than three detections are removed (Fig. 4.1j), as are trajectories with low mean detection scores (Fig. 4.1k). We used a mean score threshold of 0.8 to produce the final set of trajectories for further analyses. When analyses relating to flight track directions are required, we apply an additional filtering step to constrain analysis to detections inside a circular region of interest within the frame. This eliminates directional bias arising from the non-rotationally-symmetrical rectangular shape of the video frames.

Diagnostic plots of tracking performance over an entire short video clip can be made by taking the maximum image of the entire video clip, and plotting the detected trajectories as a single image using a different colour for each trajectory (e.g. Fig. 4.2). For example, we can see good performance of the tracking procedure in Fig. 4.2a, where all trajectories except one appear to have been correctly built. The one exception is an insect close to the centre of that figure which appears to have had its trajectory split in three parts (seen as three different coloured segments), most likely due to occlusion by another insect. Fig. 4.2b shows the result of constraining these trajectories to a circular region of interest to remove directional bias (in this case, this happened to solve the aforementioned split trajectory, but only by coincidence—the orange and purple tracks were removed for overlapping the edge of the circle).

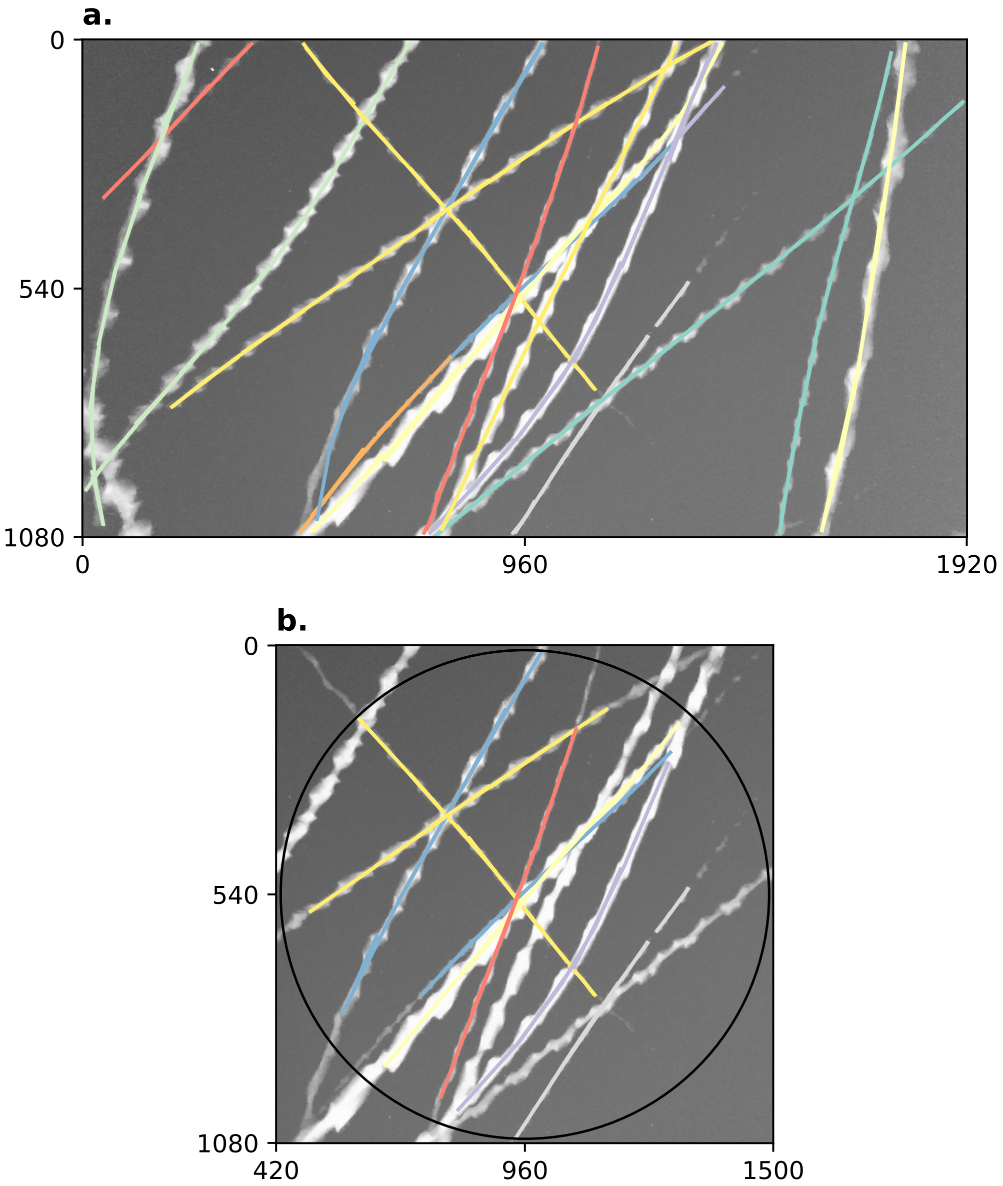


Figure 4.2: Example summary of trajectories followed by insects flying past a camera during a 5 s video clip. Axes on both plots show pixel row and column numbers. **a.** Maximum (brightest) value of each pixel across every frame in the clip with annotations overlaid. Visible bright streaks are made by the motion blurs of Bogong moths flying past the camera. The colour of an annotation indicates its membership in a unique trajectory, as predicted by our method. **b.** Annotations constrained to circular region of interest. Using only these trajectories eliminates directional bias resulting from the non-rotationally symmetrical rectangular shape of the frame. Black circle shows region of interest.

4.2.4 Camera placement and settings

A total of ten cameras (BlazeVideo, model SL112) were placed in two transects below the summit of Mt Kosciuszko, NSW on the afternoon of 18th February 2021, and collected the following morning. The first transect, which we call `kosci_south`, was placed on the south-eastern slope, running from the shore of Lake Cootapatamba up to the Kosciuszko South Ridge, and ranging in elevation from 2046 m to 2151 m. The second transect, which we call `kosci_north`, was placed on the north-western slope below the summit, with five cameras ranging in elevation from 2050 m to 2220 m. The positions of each camera are shown in Fig. 4.3a. The `kosci_north1` and `kosci_north2` locations were both within 10 m of known Bogong moth aestivation sites.

The cameras were placed such that their lenses pointed up into the sky, and the compass orientation of each camera was noted so that analysis of flight direction could be performed. Fig. 4.3b shows an example of the placement of one of the cameras. The cameras were set to take an image, along with a 5 s video clip every 30 s for the duration of the evening. Illuminance of the clear sky was recorded at multiple time points during evening twilight, near the `kosci_north2` location (Fig. 4.3a) using a digital luxmeter (Hagner, model E4-X). Luminance was also recorded from the rock face and a white standard in a few locations inside and outside a Bogong moth aestivation cave, also near the `kosci_north2` location, using a digital photometer (Hagner, model ERP-105).

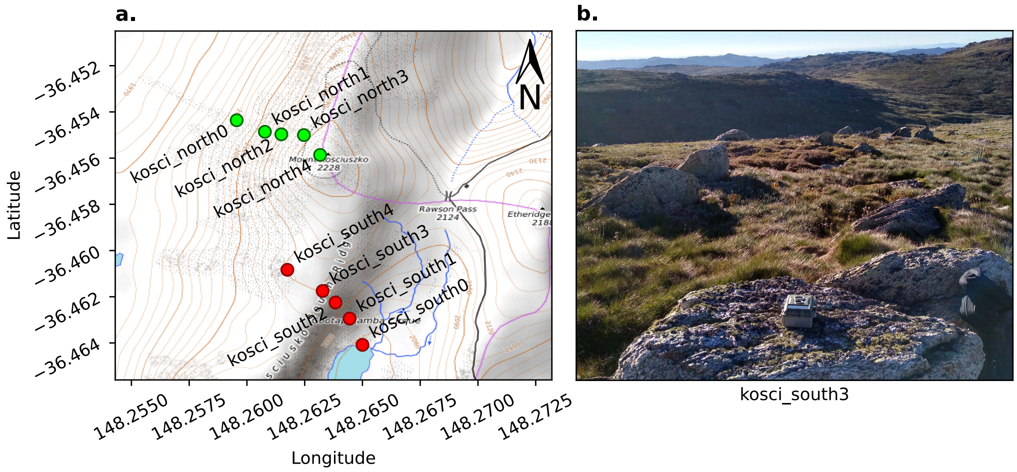


Figure 4.3: Cameras were placed on two transects on the slopes of Mt Kosciuszko, NSW. **a.** Map of camera locations. Contour lines show 10 m changes in elevation. The transects were `kosci_south` (*red*), on the south-eastern slope towards Lake Cootapatamba and `kosci_north` (*green*), on the north-western slope, below the summit. **b.** Example placement of camera at `kosci_south3` location. Map data available under the Open Database Licence at openstreetmap.org. © OpenStreetMap contributors, SRTM. Map tiles credit: © OpenTopoMap (CC-BY-SA).

4.2.5 Computational analyses

Flying insects were detected in the 5 s video clips using Camfi (Chapter 2) and tracked using the method described above (Section 4.2.3). Track directions were modelled with the orientation models described by Schnute and Groot (1992) using the CircMLE R package (Fitak and Johnsen, 2017). Maximum likelihood models were selected using Akaike's information criterion (AIC, Akaike, 1973).

4.3 Results and Discussion

At approximately 20:30 Australian Eastern Daylight Time (AEDT; UTC+11:00) the cameras switched to night mode and started using their infra-red flash. Video clips taken before this time were omitted from analysis as it was found that detection was unreliable for video clips taken in day mode. A total of 6,515 night-mode video clips were recorded, and from these 11,147 flying insects were detected. The vast majority of these are likely to be Bogong moths, as we observed a large number of them (and no other species) flying close to our vantage point near `kosci_north2` throughout the evening. Sky illuminance varied from 106.5 lx to 0.0132 lx over the course of evening twilight (Fig. 4.4b, *red trace*).

4.3.1 Activity levels

A strong peak in activity was observed during evening twilight at all sites on both transects. Activity plummeted just before 21:00 AEDT, coinciding with the end of nautical twilight (Fig. 4.4a–b). Outdoor illuminance dropped from about 1 lx during the activity peak to below 0.1 lx after activity had plummeted (Fig. 4.4e). Of the 11,147 total flying insect detections, 8,589 occurred before 21:00 AEDT (from 576 video clips), and 10,163 occurred before 21:30 AEDT (from 1,176 video clips; coinciding with the end of astronomical twilight). This agrees with previous observations of Bogong moth flight activity exhibiting large peaks during evening twilight (Common, 1954; Wallace et al., 2021; Warrant et al., 2016) (see also Chapter 3). The activity peak was most pronounced at the `kosci_north1` and `kosci_north2` locations (Fig. 4.4d), presumably owing to the proximity of these locations to known Bogong moth aestivation sites.

Previous work has demonstrated that occupied Bogong moth aestivation sites vary in elevation over the course of the summer, with moths occupying higher elevation sites as summer progresses and temperatures in-

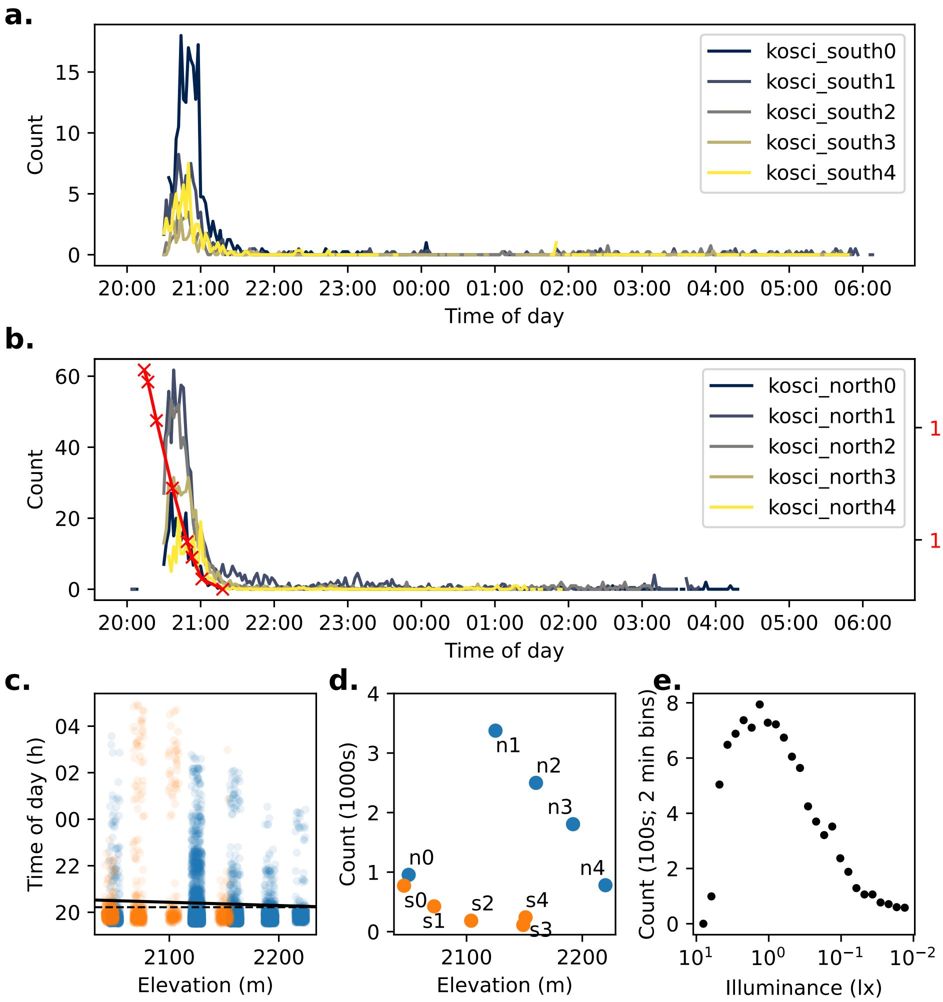


Figure 4.4: Summary of detections of flying insects on two transects of Mt. Kosciuszko on 18th–19th February 2021. **a–b.** Time series of counts of flying insect trajectories detected at each location in `kosci_south` (*a*) and `kosci_north` (*b*) transects. Counts have been smoothed by taking the average over 2 min bins. Illuminance readings recorded from the sky close to the `kosci_north2` site are shown in *b* (*red trace*). **c.** Time of each detection by elevation. Jitter (random fluctuation) is applied to elevation values to assist readability. A linear regression of time against elevation is shown by *solid black line* ($R^2 = 0.0065$, slope $= -5.59$ s/m). Despite the small effect size, the slope is statistically significant (Wald test, $p = 1.8 \times 10^{-17}$; null "zero slope" hypothesis is indicated with *dashed black line*). **d.** Total number of detections at each location by elevation. Location names are labelled "n0" = "kosci_north0", "s3" = "kosci_south3", etc. Points in *c* and *d* are marked according to transect; *blue* = `kosci_north`, and *orange* = `kosci_south`. **e.** Total detection count in 2 minute bins (pooled across all locations) plotted against illuminance (log-linearly interpolated from recorded measurements; *b, red trace*). Measurements taken between 20:28 and 21:18 are included.

crease (Green, 2003) (see also Chapter 3). There are also some indications that towards the end of summer—as temperatures start to drop—Bogong moths may re-occupy lower altitude sites, possibly as they start their return migration to the north, north-east, and east (Common, 1954). Clearly, these patterns of site occupation require movement of individuals between elevations. Since the mode of Bogong moth locomotion is predominantly flight, and they are known to be particularly strong flyers (Dreyer et al., 2018; Warrant et al., 2016), we would expect these movements between elevations to occur over short time periods—on the order of minutes or hours, rather than the days or weeks that previous monitoring methods have measured (e.g. direct observation (Caley and Welvaert, 2018; Common, 1954); fox scats (Green, 2003); daily-pooled still-image camera monitoring (Chapter 3)).

If there was a trend for Bogong moths to move between elevations on the night of our recordings, then we would expect the timing of the peak in detections to vary with elevation. However, this type of temporal shift may be difficult to detect, especially if the duration of travel for an individual moth between the elevations is short with respect to the total duration of the detection peak (since, in that case, most of the variation in detection time would be explained by variation in the times that moths take flight, rather than movement of moths across an altitudinal gradient). Fortunately, the present method provides tremendous statistical power to detect such weak interactions, owing to the sheer volume of detections it generates.

Indeed, a statistically significant—albeit extremely weak—correlation between time of detection and elevation was observed (Wald test, $p = 1.8 \times$

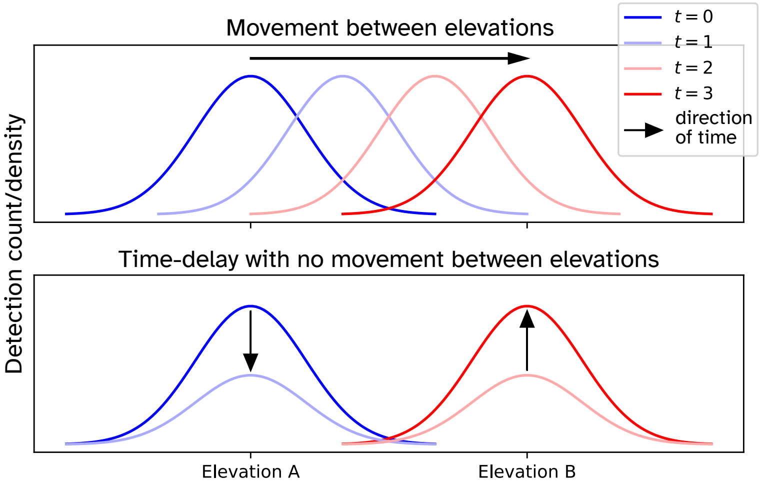


Figure 4.5: Scenarios which could lead to the observation that detection time depends on elevation. **Top panel:** In the first scenario, a lag in detection times is observed at elevation B, relative to elevation A, due to the movement of flying moths from elevation A to B. **Bottom panel:** In the second scenario, the lag is due to moths at elevation B emerging from—and returning to—their aestivation sites later than those at elevation A.

10^{-17}, linear regression $R^2 = 0.0065$, slope $= -5.59$ s/m; Fig. 4.4c, *black line*). This corresponds to a delay in detections of roughly 16 minutes from the highest site (`kosci_north4`; 2220 m) to the lowest site (`kosci_south0`; 2046 m). We could tentatively take this as an indication that the bulk of the moths are moving downhill, although from this analysis alone, we are unable to disentangle that hypothesis from the hypothesis that moths at lower altitudes merely emerge from (and/or return to) their aestivation crevices later than higher-altitude moths (see Fig. 4.5 for illustration).

So far, we have attempted to detect the movement of Bogong moths along an altitudinal gradient by recording their location (i.e. displacement) over time. However, this analysis does not incorporate any information regarding the identity of the detected moths. This, along with the fact that moths may be present at a given location without being observable (for instance, a moth might not be airborne at a particular time), prevents us from concluding—with absolute certainty—that moths are indeed moving between elevations. If we knew that a particular moth had been detected at a given elevation, and detected again at another elevation a few minutes later, we could say with certainty that the moth moved between those elevations. Alas, there is barely enough information in the images taken by the wildlife cameras to positively identify species, let alone to identify individual moths that are members of a local population numbering in the millions.

4.3.2 Evidence of orientation behaviour

Displacement is, of course, not the only way to measure movement. We can also measure its derivative with respect to time; namely, velocity (the combination of direction of displacement, which we call "track direction", and speed). In our case, we are only interested in the track direction of flights, which conveniently our method measures.

Flight track directions at each respective site in both transects showed significant departures from uniform circular distributions (Fig. 4.6), as determined by Moore's modified Rayleigh tests ($p < 0.05$ for all locations, Moore, 1980). Furthermore, track directions from each pair of locations were significantly different from each other, as determined by pairwise Mardia-Watson-Wheeler tests ($p < 0.05$ for each pair, Mardia, 1969). Thus, the flights of the Bogong moths were directed, and the direction of flight depended on location. This in itself is not surprising, however it is ethologically relevant, since directed movement requires behavioural control in response to external stimuli (Cheung et al., 2007).

Generally speaking, the distributions of flight track directions at each location were bimodal (Fig. 4.6; Table C.1) and the two modes were not separated by 180° (i.e. the bimodality of the directions was not a result of axially-directed flight). There was one notable exception to this trend, with moths detected at `kosci_south0` showing a unimodal south-easterly flight track direction tendency (Fig. 4.6, southernmost site; Table C.1). Trends in flight track direction which were seen during nautical twilight (i.e. before 21:00 AEDT) were continued throughout the night, albeit with a much lower density of moths (Fig. C.2).

It is clear from our analyses that the Bogong moths exhibited orientation behaviour, although it is not immediately clear *how* they were orienting themselves. We know from laboratory assays of orientation behaviour that Bogong moths are able to orient themselves relative to visual landmarks in conjunction with the Earth's magnetic field (Dreyer et al., 2018), and celestial cues (Dreyer and Adden et al., in prep.). Another possible source of directional information is wind, which is known to be used to control flight direction in another species of migratory noctuid moth, *Autographa gamma* (Chapman et al., 2008), and it is likely that Bogong moths also possess this ability. We wish to evaluate these possible sources of directional information with regard to our measurements of the orientation behaviour of Bogong moths in the wild.

Cues from celestial objects and the Earth's magnetic field would be roughly the same across all study locations, given the small geographical area covered by the study (all locations were within 1.2 km of each other). Therefore, these cues alone could not explain the differences in the distri-

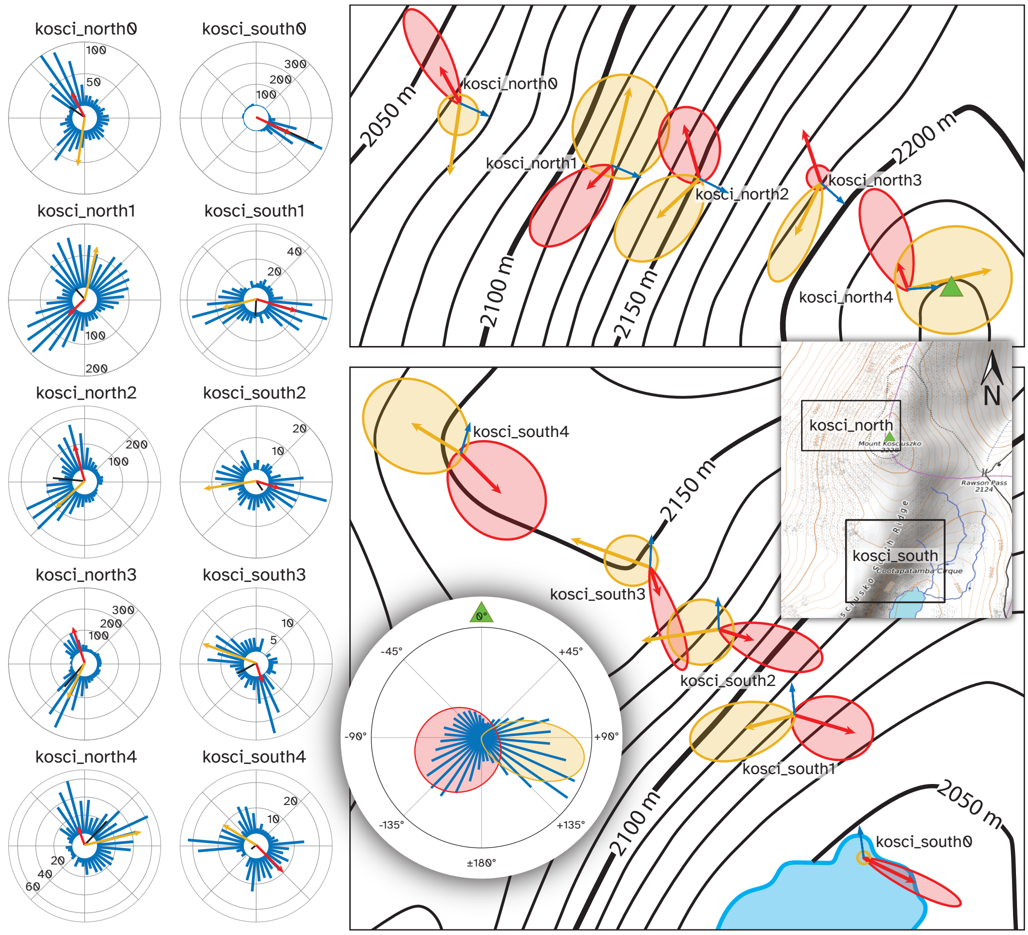


Figure 4.6: Distribution of flying insect track directions over the course of the evening of 18th February 2021, by location. **Left panel:** *Blue bars* show histograms of track directions for the given location, with scale (counts) indicated on circular axes. *Black bars* show mean vector of all detections at the respective location. *Red arrows* show direction and weight of the first component of a bimodal von Mises distribution for the respective location, computed using the CircMLE R package (Fitak and Johnsen, 2017) and *yellow arrows* show the second component. Where only a *red arrow* is shown, the data were better explained by a unimodal distribution. **Top-right panel:** Arrows from the *left panel*, with von Mises probability density functions of each component also plotted. Plots are placed at their respective camera locations on a map of the `kosci_north` transect. *Blue arrows* show the azimuth of the summit of Mt. Kosciuszko from the respective locations. *Green triangle* shows summit of Mt. Kosciuszko. **Bottom-right panel:** Follows the conventions of the *top-right panel*, showing `kosci_south` transect. **Inset, right:** Map of Mt. Kosciuszko showing locations of top-right and bottom-right panels. *Green triangle* shows summit of Mt. Kosciuszko. **Circular inset, bottom:** Histogram (*Blue bars*) of flight track directions relative to the azimuth of the summit of Mt. Kosciuszko (*green triangle*) for all detections across all locations, shown with probability density functions of components of a bimodal von Mises model of the data (*red* and *yellow regions*).

butions of flight track directions observed across the locations (Fig. 4.6, *left panel*).

Wind speeds at 3 pm, 18th February, and 9 am, 19th February 2021 were moderate to fresh—4 ms^{-1} easterly and 5 ms^{-1} north-northwesterly, respectively (recorded at Thredbo Top Station, circa 4.6 km from the summit of Mt. Kosciuszko; Bureau of Meteorology, 2021). These wind speeds are similar to the likely airspeed of a motivated Bogong moth, and could have an important impact on their resultant track direction, especially for high-flying moths. The speed and direction of wind can be modulated by topography, so it is possible that wind varied between the study locations, although this was not measured, so we cannot rule out the possibility that wind could explain the observed differences in flight track directions across the locations. However, bimodal distributions of flight track directions would be hard to explain with wind alone.

We can say with certainty that the terrestrial visual panorama varies greatly across the study locations, especially between the two transects, which are on opposing sides of the highest point in the mountain range. Dreyer et al. (2018) showed Bogong moths orienting themselves relative to the azimuth of an abstraction of the silhouette of a mountain peak (namely, a black triangle on a white background, above a black horizon). Mountains are striking visual landmarks in the Australian Alps (Paterson, 1890)[3] where Bogong moths spend their summer. It is therefore reasonable to predict that wild Bogong moths exhibiting directed flights in their summer range would fly in directions relative to mountain peaks. As it happens, the summit of

---

[3] From *The man from Snowy River* (Paterson, 1890),
“And down by Kosciusko, where the pine-clad ridges raise
Their torn and rugged battlements on high, ...”

Mt. Kosciuszko (2228 m) is the highest peak in Australia, and it is also the closest peak to all of the study locations in both transects. We therefore proceed by comparing the azimuth of the summit of Mt. Kosciuszko with flight track directions of the Bogong moths we detected.

Indeed, flight track directions relative to the azimuth of the summit of Mt. Kosciuszko clustered bimodally (pooled across all locations, confirmed by AIC-based maximum-likelihood model selection, Table C.3; distributions shown in Fig. 4.6, *right panels* and *circular inset*). The respective means of the two components of a bimodal von Mises model, fit to flight track directions relative to the azimuth of the summit were $-118°$ (SD: 80.2°) and $+103°$ (SD: 28.8°) (Fig. 4.6, *circular inset*). As both of these are greater (in absolute terms) than 90°, the Bogong moths were, in aggregate, moving away from the summit. And since there is no higher point in Australia than the summit of Mt. Kosciuszko, the moths were also moving downhill, on paths which would take them around the mountain.

4.3.3 WHY DO AESTIVATING BOGONG MOTHS TAKE FLIGHT?

Our analyses of the rich dataset produced by our new method have so far told us *what* the Bogong moths were doing (moving downhill), and have provided us with a robust hypothesis for *how* they were doing it (flying relative to the azimuth of the nearest—and highest—summit). What remains to be answered is, *why* do Bogong moths behave this way? Indeed, why do they take flight almost every evening throughout summer—a decidedly non-dormant activity—when they are supposedly aestivating?

We have presented evidence from both displacement and velocity data indicating that on the evening of 18th February 2021, Bogong moths on Mt. Kosciuszko were, in aggregate, moving downhill. This movement was characterised not by a straight-line departure from the peak of the mountain, nor a departure in a particular direction. Instead, it was characterised by motion *relative* to the azimuth of the peak, with moths presumably fixing the direction of their flight by holding the azimuth of the summit at a constant obtuse angle, with respect to their direction of travel, leading to trajectories that would resemble portions of outward logistic spirals centred on the summit, when viewed from above.

Interestingly, this *almost* matches qualitative observations made by JRAW and EJW from the same vantage point near `kosci_north2`, about 14 months earlier. An excerpt from JRAW's field notes from 20:45 on the 28th December 2019 reads,

> "When I look up the hill, I can see fast-moving moths moving right to left [and left to right]. And when I look to the side, along the mountain, the overarching movement is a slow movement uphill. There seems to be two different modes of moth flight—there's a slow upward movement, and then a fast lateral movement in both directions."

Perhaps we were seeing the equivalent pattern of flight directions to those on 18th February, 2021, with an uphill rather than downhill trend (in this case, flight directions would form an acute angle with the azimuth of the summit, resulting in inward-logistic-spiral trajectories).

If this is true, an appealing explanation for the up- and downhill movements is that Bogong moths were seeking new aestivation sites of higher elevation on 28th December 2019, while on 18th February 2021, the moths on Mt. Kosciuszko were getting ready to leave. This would make sense, as temperatures typically don't peak until January in Australia, so it is likely that Bogong moths are still in the forward half of their round-trip migration in late December (and higher elevations have lower temperatures, thanks to adiabatic expansion). Notably, daytime temperatures in the Australian Alps were high in the last few days of 2019, reaching 23.3°C on 28th December at Thredbo Top Station (Bureau of Meteorology, 2019),[4] while on 18th February 2021, the temperature only reached a more moderate 15.6°C at the same location (Bureau of Meteorology, 2021). We know from long-term monitoring data that the bulk of the Bogong moths had already left the lower elevation aestivation sites of Mt. Gingera and Ken Green Bogong by mid-February 2021, and that numbers on Mt. Kosciuszko were declining in that month (Chapter 3, Fig. 3.3), so it is reasonable to conclude that the return migration had begun.

A possible explanation for why Bogong moths move laterally (i.e. in a logistic spiral) around the mountain, rather than in a straight line, is that in addition to altering elevation, these flights are used to calibrate the moths' internal compasses. The flights only occur just after sunset, so the flying moths would be able to see the azimuth of the sunset, which is an extremely stable compass cue. Similarly, they could be using their magnetic sense to perceive the Earth's magnetic field (Dreyer et al., 2018), an even more stable

---

[4] and 24.6°C on the 30th, and then 26.5°C just four days later (4th January 2020) which was the worst day for Kosciuszko National Park (KNP) of the 2019-2020 bush fire season, which saw over 200,000 ha of KNP burn.

compass cue. Meanwhile, the moths could be taking snapshots of the terrestrial, and possibly celestial panorama (as suggested for dung beetles, el Jundi et al., 2016), which they could later use as a terrestrial compass cue (Zeil, 2012) while they remain in the area, helping them to navigate at the start of their return migration to their breeding grounds. There is a distinct possibility that each of these cues are taken together, and these evening flights are used by Bogong moths to calibrate multi-sensory internal compasses which they eventually use for their return migration. Such multi-sensory compass calibrations are thought to be performed by migratory songbirds (reviewed by Foster et al., 2018; Pakhomov and Chernetsov, 2020).

In order to see the entire distant terrestrial panorama in the direction of their return migration, a moth would either have to fly up above the summit of the mountain, or it would have to traverse around the summit while taking snapshots, since approximately half of the panorama would be occluded by the mountain for a moth flying below the summit. The latter would likely be a safer strategy, as a higher altitude flight could present a risk of strong winds blowing the moth away from the mountain entirely, forcing the moth to expend more energy to return if it is not yet ready to leave. This process would also be useful for the forward phase of Bogong moth migration, as it would enable the moths to check the horizon for taller, and therefore more favourable (particularly in late summer) mountains, which they could then orient relative to in the subsequent leg of their journey, akin to beacon-aiming performed by wood ants (Graham et al., 2003).

4.4 Additional remarks

Some of the most interesting behavioural phenomena involve highly complex and fragile mechanisms which are easily disturbed, rendering them challenging to study. Animal migration is a conspicuous example of such fragility. For instance, merely eclosing otherwise wild Monarch butterflies in captivity is enough to disrupt their migratory orientation behaviour (Tenger-Trolander et al., 2019). Therefore, it is important that we are able to support laboratory results of animal behaviour with data obtained in the wild, ideally without any potentially disruptive manipulations (i.e. exposing the animal only to natural stimuli). In this paper, we have shown that for certain questions, our new method—which enables us to inexpensively make ethological observations of wild insects in a high-throughput, quantitative, and completely non-invasive manner—allows us to do just that. Conversely, laboratory-based experimentation is extremely useful for testing specific hypotheses, as it allows us to present animals with controlled stimuli of our own choosing (we could, for instance, change the azimuth of a prominent landmark).

Our results have generated a number of interesting and plausible hypotheses relating to the previously unexplained summer evening flight behaviour of Bogong moths. The first hypothesis is that the evening flights serve a specific navigational purpose. In particular, these flights might be used by Bogong moths to calibrate their internal compasses by integrating directional information provided by the azimuth of the setting sun, the geomagnetic field, and the visual panorama. Second, the visual panorama might be used by the moths on both the forward and return legs of their migration. On the forward leg, it would enable them to identify other mountains which may be more suitable for continuing their aestivation on (e.g. higher mountains). On the reverse leg, it could be used as a reliable compass, which provides valid directional information for moths remaining in the local area. Such a compass would be especially useful for helping the moths select favourably-directed winds for their return migration (Chapman et al., 2008). Third, to navigate effectively, the Bogong moths may need to see the visual panorama in the direction of their migration, and to access this, they fly around the nearest prominent mountain peak, rather than flying to a high altitude, where there is a risk of being blown off course.

Mouritsen (2018) listed twenty of the most important open questions in long-distance navigation research for the next twenty years. Included in this list is the question,

> “How does the pinpointing-the-goal phase work in a Monarch butterfly or Bogong moth, which can pinpoint their very specific wintering [or, for the Bogong moth, summering] locations even though they have never been there before?”

If our hypothesis that Bogong moths use their summer evening flights during their forward migration to identify taller and taller mountains turns out to be true, then this could go some way to answering this question. The Bogong moths may have never had an opportunity to learn where the tallest mountains in the Australian Alps are, but by employing a beacon-aiming navigational strategy they could simply find out when they get there. Of course, there must be other factors at play as well, since Bogong moths don’t all end up on the highest peak (i.e. Mt. Kosciuszko). Lower peaks with otherwise favourable conditions (e.g. with availability of crevices with suitable temperature, humidity, and darkness) could also end the forward migration.

4.5 Author contributions

EJW, JZ, and JRAW conceived the project. EJW and JRAW designed the experiment. JRAW collected the data, devised the algorithms, and wrote the software. DD and JRAW analysed the data. JRAW wrote the first draft of the manuscript. All authors critically interpreted the results, and contributed to the writing and editing of the manuscript.

4.6 Acknowledgements

EJW and JRAW are grateful for funding from the European Research Council (Advanced Grant No. 741298 to EJW), and the Royal Physiographic Society of Lund (to JRAW). JRAW is thankful for the support of an Australian Government Research Training Program (RTP) Scholarship. EJW holds Scientific Permits for collection and experimental manipulations of Bogong moths in several alpine national parks and nature reserves (NSW Permit SL100806). We are extremely grateful to Mark Rullo for assisting with the fieldwork, and to Dr. Stanley Heinze for useful discussions and providing computational resources.

Part II

The molecular basis of directed migration

This part is made up for two chapters, and like in Part I, these chapters are structured as research papers. The purpose of these chapters is to begin to develop an understanding of the genetic basis of unique life history of the Bogong moth, including its long-distance navigation. Of particular interest are the following questions: What is the molecular basis of the geomagnetic sense in the Bogong moth? Are there specific genetic drivers of the moth's heritable migratory direction? And what is the impact of the moth's close association with the parasitic nematode on both genomic and epigenomic systems? To satisfactorily answer these questions would require a research campaign spanning years, and possibly decades. However, considerable progress can be—and has been—made in the course of a PhD.

First, we must get a handle on the entity which contains the heritable information for the exceptional navigator that is the Bogong moth: its genome. Thus, in Chapter 5, we present, for the first time, the sequence of the genome of the Bogong moth. This is an essential piece of research infrastructure for many types of molecular experiments, and its production helps elevate the status of the Bogong moth to a *bona fide* model system. Already in this chapter, we present the results of a few of these types of experiments, and report several interesting—and even unexpected—findings.

Armed with the Bogong moth's genome, we can begin to tackle the second question I presented above. Namely, are there specific genetic drivers of the moth's heritable migratory direction? There are two steps to answering such a question. The first is to find possible drivers through a correlative genome-wide association study (GWAS), which compares genotype with phenotype (in this case, migratory direction). The second step is to apply the so-called “reverse genetics” paradigm, to allow us to (causatively) infer the function of the phenotype-associated genetic variants identified through the GWAS. Genome-wide association studies are relatively expensive, and are never guaranteed to work, especially in the case of highly complex, multigenic traits (which a behavioural trait like migratory direction is likely to be). Undeterred, and buoyed by an undying fascination with the so far inexplicable navigational prowess of the Bogong moth, we tried it anyway. In Chapter 6, we present the results of our (successful) GWAS, identify promising candidates for future reverse-genetic experimentation, and discover that the Bogong moth population is more-or-less panmictic, a result which has far reaching implications for understanding its navigational strategies, and for its conservation.

Chapter 5

The genome of a remarkable nocturnal navigator, the migratory Australian Bogong moth *Agrotis infusa*

Jesse R A Wallace[1,2], Kristina Brauburger[2], Robert Kucharski[1], Jean-Marc Lassance[3,4], Allison M Churcher[5], Estelle Proux-Wéra[6], Lucile Soler[7], Mosè Manni[8], Brendan Beaton[1], Paul Clémençon[9], Stanley Heinze[2], Evgeny Zdobnov[8], Ryszard Maleszka[1] and Eric J Warrant[1,2]

[1]Research School of Biology, Australian National University. [2]Lund Vision Group, Department of Biology, Lund University, Sweden. [3]Department of Biology, Lund University, Sweden. [4]Department of Organismic and Evolutionary Biology, Harvard University, USA. [5]NBIS, Department of Molecular Biology, Umeå University, Sweden. [6]Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Sweden. [7]Department of Medical Biochemistry and Microbiology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Sweden. [8]Department of Genetic Medicine and Development, Faculty of Medicine, University of Geneva, Switzerland. [9]ENS de Lyon, Université Calude Bernard, Biosciences.

Abstract

The endangered Australian Bogong moth is a nocturnal navigator that migrates over long distances in an extraordinarily directed and precise manner. During its aestivation period in mountain caves some adult Bogong moths become infected by parasitic nematodes. Remarkably, the following generations are capable of accurately repeating the migratory cycle to the same small set of destinations without prior knowledge of the migratory route. To facilitate studies on the genetic mechanisms controlling this unique biology, we assembled a high-quality draft of the Bogong moth's genome, performed tissue and migratory phase-specific transcriptional profiling, and mapped the genomic distribution of methylated cytosines. Here we report several interesting and even unexpected findings emerging from our multilevel analyses, including some noteworthy features of the genomic landscape, contrasting migratory phase-dependent transcriptomes including differentially spliced isoforms, and an unexpectedly high level of non-CpG methylation, not seen previously in invertebrates. To complement this study, we also sequenced and analysed the genome of the Bogong moth-associated parasitic nematode, *Amphimermis bogongae.* Our data open a new line of research that will help to unravel a set of currently unknown genetic instructions behind the Bogong moth's marvellous biology.

5.1 Introduction

The Bogong moth *Agrotis infusa* (Lepidoptera: Noctuidae) (Fig. 5.1, *inset, upper left*) is an iconic Australian insect that is well known to Australians, with deep cultural importance for many of the country's indigenous peoples (Flood, 1996; Warrant et al., 2016). This moth is also famous for its spectacular long-distance nocturnal migration (Common, 1954; Warrant et al., 2016). Travelling in spring from a broad arc of winter breeding areas in south-eastern Australia (Fig. 5.1), an estimated 4.4 billion newly-emerged adult moths (Green, 2010b) travel up to 1000 km in different directions, arriving at a geographically tightly-restricted assemblage of high-elevation caves in the Australian Alps (Fig. 5.1). After entering the caves, moths densely tile the cave walls (ca. 17,000 moths/m^2: Common, 1954) and begin 3-4 months of summer dormancy, known as aestivation (Fig. 5.1, *inset, lower right*). At the beginning of the following autumn, the same individuals that arrived months earlier leave the caves and make a return migration to the breeding areas. Once there, they mate, lay their eggs and die, with their offspring developing during the coming winter and emerging as adults during the following spring (Fig. 5.1, *inset, upper right*). After emergence, the next generation of moths begins (and completes) the entire migratory cycle afresh.

Insect migrants typically undertake seasonal movements in search of better conditions—an improved climate, more abundant food or a reduced risk from predators and infectious diseases (Chapman et al., 2015). For the vast majority of these insects, migrants travel from one broad latitudinal zone to another, rather than to a specific destination, and do so over several

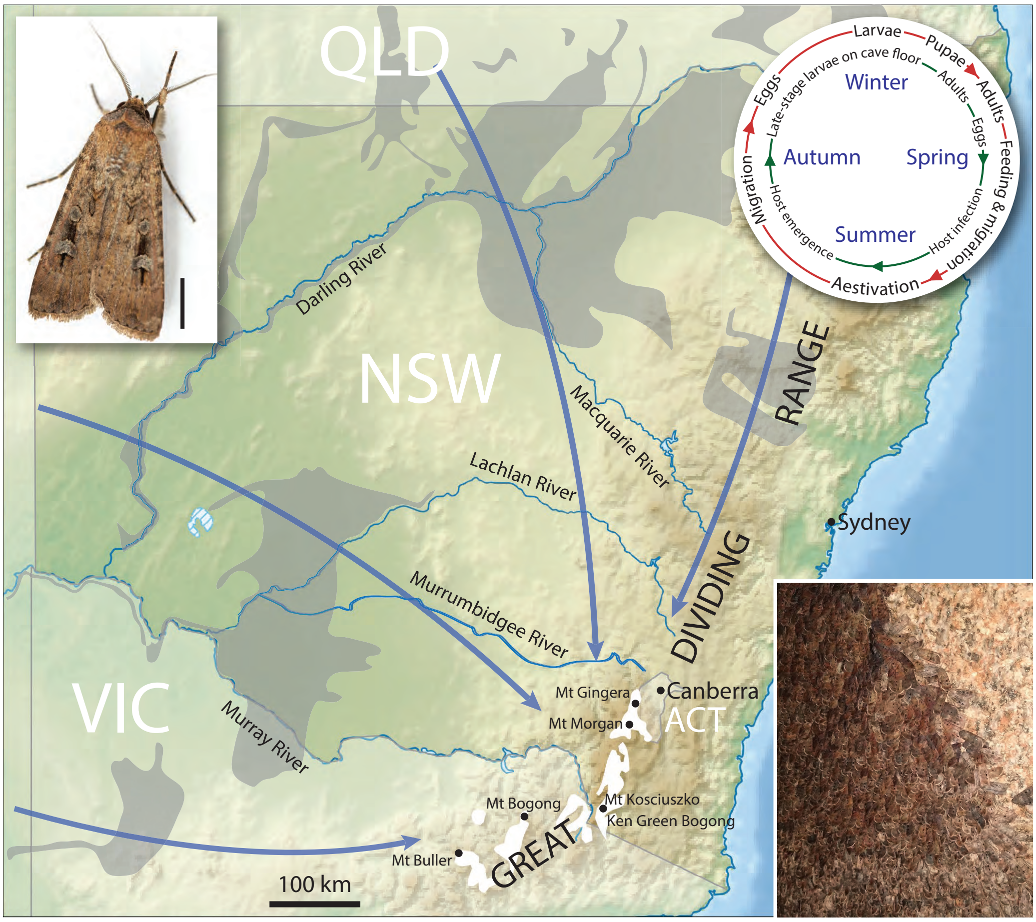


Figure 5.1: The Bogong moth. **Inset, upper left:** A male Bogong moth (*Agrotis infusa*). Scale bar = 5 mm. Photo courtesy of Dr. Ajay Narendra, Macquarie University, Australia. **Main:** Likely migratory routes (*arrows*) of moths during spring to alpine regions in south-eastern Australia. Autumn migration occurs in the reverse directions. Areas of grey cracking clays—favoured soils for Bogong moth winter development—are shown in *grey*. The *white areas* represent elevations above 1500 m, where all known summer aestivation sites are located. **Inset, upper right:** The life cycles of the Bogong moth (*outer red circle*) and the parasitic mermithid nematodes *Amphimermis bogongae* and *Hexamermis cavicola* (*inner green circle*). The nematode life cycle occurs entirely within the Bogong moth aestivation cave. Bogong moths undergo a spring migration to escape the increasingly warm conditions of the breeding grounds. Derived from information given in Common (1954) and Welch (1963). **Inset, lower right:** Around 17,000 moths/m^2 undergo a summer aestivation of up to four months on the walls of specific caves in the Australian Alps before making the return migration in autumn.

generations (Gao et al., 2020). The highly directed and univoltine (i.e. single generational) migration of the Bogong moth (and possibly a couple of other noctuid species: Gao et al., 2020; Oku, 1983; Pepper, 1932), stands in stark contrast to this. So too does the migration of the Monarch butterfly *Danaus plexippus*, that makes a highly directed autumn migration from broad areas of the northern USA and southern Canada to a specific mountainous overwintering destination in central Mexico (Urquhart, 1987). However, unlike the Bogong moth, the Monarch butterfly's yearly migratory cycle is completed over several generations (i.e. it is multivoltine).

Thus, the Bogong moth is a univoltine and highly-directed nocturnal long-distance migrant that, upon arrival in the mountains, enters a long period of summer dormancy. It is also an accomplished navigator, relying on the Earth's magnetic field (Dreyer et al., 2018) and the stars (Dreyer et al., in preparation) as navigational compasses for travel along its inherited migratory route towards its alpine destination (a place it has never previously visited). Apart from several species of night-migratory birds (Alerstam, 1993; Mouritsen et al., 2016; Wiltschko, 1983), the Bogong moth is the only animal known to use these two compass cues for long-distance navigation at night, making it a unique and remarkable insect.

This impressive migratory strategy has led to the evolution of an exclusive association with two species of rare parasitic mermithid nematodes—*Amphimermis bogongae* and *Hexamermis cavicola*—that are only found within the Bogong moth aestivation caves (Common, 1954; Welch, 1963). After surviving as final-stage larvae throughout the alpine winter, buried deep within moist layers of cave floor detritus, these free-living nematodes moult to become sexually mature egg-laying adults at the beginning of spring (Fig. 5.1, *inset, upper right*). Eggs hatch to produce infective larvae just as the Bogong moths begin to arrive in the caves. Via unknown mechanisms, possibly involving the larvae ascending the walls of the cave within downward-flowing streamlets of water, many moths become infected, thus becoming a rich food supply for the growing nematode. The larvae eventually reach a length of up to 20 cm (Welch, 1963), filling the moth's body cavity and killing the moth by the time the nematode exits in its final larval stage in late summer (after which it returns to the cave floor in preparation for the coming winter). Mermithid nematode infection of adult insects is incredibly rare (larval infection is the norm), which underscores the remarkable evolutionary coupling of the life cycles of these two nematodes with that of their adult Bogong moth host.

Despite this fascinating range of features, there have been no attempts to understand the path from genotype to organismal complexity and behaviour of the Bogong moth.

What is the molecular basis of the geomagnetic sense in the Bogong moth? Are there specific genetic drivers of the moth's heritable migratory direction? What is the impact of the moth's close association with the parasitic nematode on both genomic and epigenomic systems? These are some of the essential questions that can only be answered by using modern molecular tools to explore the moth's genetic blueprint. To advance our understanding of this intriguing—and now endangered (Warrant et al., 2021)—species, we sequenced and assembled its genome, annotated the predicted proteome, and examined gene expression in various tissues in both migrating and aestivating moths. As part of this project, we also generated an initial genome-wide map of methylated cytosines, the so called “methylome”. Together with other epigenomic mechanisms, this important modification of DNA provides the level of regulatory flexibility that is required for generating developmental and behavioural diversity from a limited number of genes (Law and Jacobsen, 2010; Maleszka, 2016). Additionally, we sequenced the genome of the Bogong moth-associated parasitic nematode, *A. bogongae*, to test for signatures of selection associated with the evolution of these two interlocked host-parasite life cycles.

Here we report several interesting and even unexpected findings emerging from our multilevel analyses, including some noteworthy features of the Bogong moth’s genomic landscape, contrasting migratory phase-dependent transcriptomes including differentially spliced isoforms, and an unexpectedly high level of non-CpG methylation.[1] Such a prominent level of non-CpG methylation contradicts the recent proposal that this type of DNA modification is a vertebrate invention (Mendoza et al., 2021), and suggests a novel role in some insects.

Although it is not straightforward to move from the level of molecular networks to that of brain circuitry operating in real time to generate behaviours (Hyduke and Palsson, 2010; Miklos and Maleszka, 2011), this new powerful set of data opens virtually unlimited research possibilities into the role of molecular mechanisms linking genotype to phenotype. The Bo-

---

[1] Methylation of cytosine residues is a common post-replication modification of DNA seen across the domains of life, and which is known to play a role in the regulation of gene transcription. The most common context in which cytosine methylation is observed are CpG sites, where a cytosine is followed by a guanine in the genome. Context-dependent DNA methylation is a consequence of the fact DNA-methyltransferases—the enzymes which mediate DNA methylation—act in a sequence-specific manner.

gong moth is now poised to become an accessible model enabling high-tech studies into how emergent phenomena become permanent features of insects' sophisticated behaviours.

5.2 Methods

5.2.1 DNA EXTRACTION AND GENOME SEQUENCING

The specimens used for genome sequencing all originated from a mating of one male and one female Bogong moth, caught in January 2019 at Mt Selwyn, Australia. Larvae were raised on an artificial white-bean-based diet at the Biology department of Lund University, Sweden.

5.2.1.1 Oxford Nanopore sequencing

One female pupa was collected right after pupation (≤ 1 day) and flash-frozen in liquid nitrogen. The cocoon was removed and high molecular weight DNA was extracted using 1.5 ml pestle tube homogenisation and CTAB lysis (2% CTAB, 1.4 M NaCl, 20 mM EDTA, 0.2% BME, 100 mM tris-HCl (pH 8), 1.2 mg/ml Proteinase K, 60°C for 60 min), followed by phenol-chloroform-isoamyl (25:24:1) extraction. Subsequently, impurities were removed using a high salt, low ethanol, chloroform purification. Two aliquots containing approximately 11 μg DNA each were taken from the isolated DNA sample and sheared to 75 kb and 20 kb, respectively. The first aliquot was then size-selected to ≥ 15 kb using a *BluePippin* (Sage Science). Libraries for each of the sheared aliquots were prepared using the *LSK-109* *sequencing kit* (Oxford Nanopore Technologies), and loaded on two separate *R9.4 PromethION flow cells.* Sequencing was performed on a *PromethION beta* machine (Oxford Nanopore Technologies), using MinKNOW for PromethION (version 19.05.1, Oxford Nanopore Technologies). Base calling was performed with guppy (v.3.0.3, Oxford Nanopore Technologies). DNA isolation and sequencing were performed by the Uppsala Genome Center, Uppsala, Sweden.

5.2.1.2 10x Chromium sequencing

The remainder of the DNA isolated for the Oxford Nanopore sequencing (approximately 0.6 μg) was used for 10x Chromium sequencing. Sequencing libraries were prepared using the *Chromium™ Genome Chip Kit* (cat # 120257/58/61/62; 10x Genomics) according to the manufacturer's protocol. Libraries were sequenced on the *HiSeq X* platform (paired-end 150 bp read length, v2.5 sequencing chemistry; Illuminia). Library preparation and sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden.

5.2.1.3 Hi-C sequencing

For Hi-C sequencing, one female originating from the same mating as the individual used above was raised to adulthood, then anaesthetised by cooling to 4°C. Wings and legs were removed, and the body was descaled, and then flash-frozen in liquid nitrogen. The thorax was separated from the head and abdomen, and ground into a flour-like powder on dry ice. The sequencing library was prepared using the *Dovetail™ Hi-C Preparation Kit* (v.1.03 4-13-18, with custom variation for sample filtering; Dovetail Genomics). Clustering was done by *cBot* (Illumina) and samples were sequenced on a single NovaSeq 6000 SP lane (2×150 bp reads, XP kit; Illumina). Bcl to FastQ conversion was performed using bcl2fastq_v2.20.0.422 from the *CASAVA* software suite (Illumina). DNA extraction, library preparation and sequencing were performed by the National Genomics Infrastructure Genomics Applications, Stockholm, Sweden.

5.2.1.4 Bisulfite sequencing

A single whole Bogong moth pupa from the same mating pair (to keep heterogeneity to a minimum) was used for bisulfite sequencing. DNA isolation was performed as described by Evans et al. (2013). Bisulfite conversion was performed using the *Zymo EZ DNA Methylation Kit (D5001)* (Zymo Research), as per the manufacturer's protocol. Libraries were prepared using the *True Methyl kit* (Illumina), as per the manufacturer's protocol. Input DNA amount was 100 ng as per NanoDrop reading. Two technical replicates were prepared and sequenced but only one was used for analysis.

5.2.1.5 Sequencing the genome of a Bogong moth-associated nematode

Four samples of the Bogong moth-associated parasitic nematode *A. bogongae* were collected from a Bogong moth aestivation cave on Mt. Kosciuszko, NSW. Total DNA was extracted from these samples using a *Quick-DNA™ MagBead Plus kit* (Zymo Research, 2017) and quantified using a *Qubit™ dsDNA HS Assay Kit* on a *Qubit™ 3 Fluorometer* (Thermo Fisher Scientific, 2017) using the manufacturers' respective protocols. Sequencing libraries were prepared using the *Illumina® DNA PCR-Free Prep, Tagmentation* with the *IDT® for Illumina® DNA/RNA UD Indexes Set A, Tagmentation* (Illumina, 2020) using the manufacturer's protocol, and a unique index was used for each sample. The DNA content of the libraries were quantified using a *Qubit™ ssDNA Assay Kit* on a *Qubit™ 3 Fluorometer* (Thermo Fisher Scientific, 2017), then pooled in equal DNA proportions. To ensure library quality, the pooled library was first sequenced on the Illumina MiSeq platform with a target of 1 million reads, and quality control checks were performed. Based on the read counts from the MiSeq run, the libraries were re-pooled and sequenced on the Illumina NovaSeq platform.

5.2.2 QUALITY CONTROL

Standard sequencing quality control was performed using FastQC. We generated a k-mer profile using the Chromium 10x data and *Jellyfish* (version 2.2.6, Marçais and Kingsford, 2011). k-mer sizes of 17, 21, and 25 were used to generate a k-mer histogram (Fig. D.1). To assess contamination levels, in all of the genomic short-read libraries (except the bisulfite-converted libraries), reads were classified using Kraken 2 (Wood et al., 2019) with the National Center for Biotechnology Information (NCBI) non-redundant nucleotide (nt) database. Most reads from Bogong moth short-read libraries were unclassified, or identified as Lepidoptera (11%), as expected, with very low levels of bacterial contamination. Similar results were obtained from running Kraken 2 classification on short reads from *Operophtera brumata* (NCBI accession: SRR1618582) and *Agrotis ipsilon* (NCBI accession: SRR8103939) (data not shown). We recommend that read classification should be part of quality control for *all* sequencing experiments, as our first attempt at sequencing the Bogong moth genome was hindered by lab-based contamination, and sequencing ultimately had to be repeated.

5.2.3 GENOME ASSEMBLY

The longest 32 Gb of the PromethION reads (estimated 50x coverage) were assembled using the *redbean/wtdbg2* assembler (version 2.5, Ruan and Li, 2020). The Chromium 10x reads were then aligned to the assembly using the *Long Ranger align* pipeline (version 2.1.4, 10x Genomics). Polishing was done with *Pilon* (version 1.22, Walker et al., 2014) using the 'jumps' and 'diploid' flags. The Hi-C data was aligned and analysed using *BWA* (version 0.7.17, Li and Durbin, 2009) and the *Juicer* pipeline (version 1.9.9, Durand et al., 2016b). The *3D-DNA* pipeline was used to correct, order, orient, and anchor scaffolds (Dudchenko et al., 2017). The assembly was reviewed using *Juicebox Assembly Tools* (Dudchenko et al., 2018; Durand et al., 2016a) followed by three more rounds of *Pilon* polishing with the Chromium 10x data. Manual inspection of the resultant assembly revealed that some indels remained uncorrected. These indels were corrected using a novel INDEL-correction tool we developed, named *Windel* (described in Appendix D.1; available from https://github.com/J-Wall/windel). A final round of polishing with *Pilon* was then performed.

5.2.4 Genome annotation

5.2.4.1 Evidence used for annotation

Protein sequences were collected for the annotation from the *Uniprot Swiss-Prot* database (561356 proteins downloaded November 2019, Magrane et al., 2011). Transcriptome evidence from an RNA-Seq experiment of Bogong moths (Section 5.2.6) was also used. In particular, a *de novo* transcriptome assembly was made using *Trinity* (Grabherr et al., 2011), and a reference-guided assembly was made using *StringTie* (Pertea et al., 2015).

5.2.4.2 Repeat masking

We created a species-specific repeat library modelled using the *RepeatModeler* package (version 1.0.11, Smit and Hubley, 2008-2015). As repeats can be part of actual protein-coding genes, the candidate repeats modelled by *RepeatModeler* were vetted against our protein set (minus transposons) to exclude any nucleotide motifs stemming from low-complexity coding sequences. From the repeat library, identification of repeat sequences present in the genome was performed using *RepeatMasker* (version 4.0.9_p2, Smit et al., 1996-2010), and *RepeatRunner* (Smith et al., 2007).

5.2.4.3 ***Ab-initio*** training

We used multiple gene finders to improve the final genome annotation. In particular, we trained *Augustus* (version 3.3.3, Stanke et al., 2006), and *Snap* (version 2013_11_29, Korf, 2004), using a custom *Nextflow* (Di Tommaso et al., 2017) pipeline, available at https://github.com/NBISweden/pipelines-nextflow/tree/master/AbinitioTraining.

5.2.4.4 Gene build

Gene builds were computed using the *MAKER* pipeline (Holt and Yandell, 2011), which includes the following software: *exonerate* (version 2.4.0), *BLAST* (version 2.9.0), *RepeatMasker* (version 4.0.9_p2, Smit et al., 1996-2010), *BioPerl* (version 1.7.2), *Augustus* (version 3.3.3, Stanke et al., 2006), *tRNAscan-se* (version 1.3.1), and *Snap* (version 2013_11_29, Korf, 2004).

An evidence-guided build was computed using the *MAKER* software to construct gene models directly from both aligned transcript sequences and reference proteins. This annotation was then passed again to *MAKER*, together with a curated *ab-initio* profile. Additionally, we used *EVidenceModeler* (version 1.1.1, Haas et al., 2008), which allows the construction of gene models based on the set of exons produced by the *ab-initio* tools which are most consistent with the evidence.

5.2.4.5 Functional annotation

Functional inference for genes and transcripts was performed using the translated CDS features of each coding transcript. We ran *BLAST* (version 2.9.0) against the *Uniprot Swiss-Prot* reference dataset, and *InterProScan* (version 5.30-69.0, Jones et al., 2014), and then parsed output from both analyses to extract and reconcile relevant metadata into predictions of canonical protein names and functions. This was performed using a custom *Nextflow* pipeline, available from https://github.com/NBISweden/pipelines-nextflow/tree/master/FunctionalAnnotation.

5.2.5 DNA methylation analysis

Analysis of the whole-genome bisulfite sequencing data followed that done by Welsh et al. (2017). Mapping, de-duplication, and methylation data extraction were performed using *Bismark* (version 0.23.0, Krueger and Andrews, 2011). A report splitting by cytosine context (i.e. CG, CHG, CHH) was generated using a custom script (written by Sylvain Forêt). Strand merging (for CG-context) and methylation calling using a binomial test (assuming bisulfite conversion rate of 0.99) with Benjamini-Hochberg multiple-testing correction were performed using custom scripts (also written by Sylvain Forêt). Summary statistics from the above analyses were obtained using a custom R script which depends on the *GenomicFeatures* library (Lawrence et al., 2013).

5.2.6 Transcriptome and differential expression experiment

To compare the expression profiles of Bogong moths in different migratory phases, and to provide transcriptomic evidence for genome annotation, we performed RNA-Seq on eye, brain, and antennae tissue dissected from twenty migrating and twenty aestivating Bogong moths (combined into pools of five individuals each; Fig. 5.2).

The migrating moths were collected between the hours of 20:10–01:04 over multiple nights in November 2017 at Mt. Kaputar, NSW. The moths were captured into sample jars by attracting them to a white sheet suspended between two trees using a search light (model gt175, Ammon Luminaire Company) and a LepiLED lamp (Brehm, 2017). Whole moths were immediately dropped into individual tubes containing approximately 10 ml absolute ethanol. The aestivating moths were collected from a cave on Ken Green Bogong, NSW, in January 2018. The moths were captured into a plastic container and transported to the lab, where they were kept in the fridge (to simulate the cave temperature and keep them in the aestivating state). The next night, between 22:53–01:17, the moths were dropped into individual tubes containing approximately 10 ml absolute ethanol. This way, the timing of ethanol-fixing for migrant and aestivating samples was matched, controlling for circadian cycling of expression profiles.

Antennae, brains, and retinas of the ethanol-fixed samples were dissected, and total RNA from four pools of five individual moths from each condition (for a total of 24 tissue-condition-replicate pools; Fig. 5.2) was extracted using the *ISOLATE II RNA Mini kit* (Bioline) according to the manufacturer's instructions. Each pool contained both male and female moths. mRNA enrichment with oligo(dT) beads, library preparation, and sequencing (HiSeq-PE150, Illumina) was performed by Novogene (Hong Kong). These data were used for genome annotation (Section 5.2.4), differential expression analysis (Section 5.3.4), and alternative splicing analysis (Section 5.3.5).

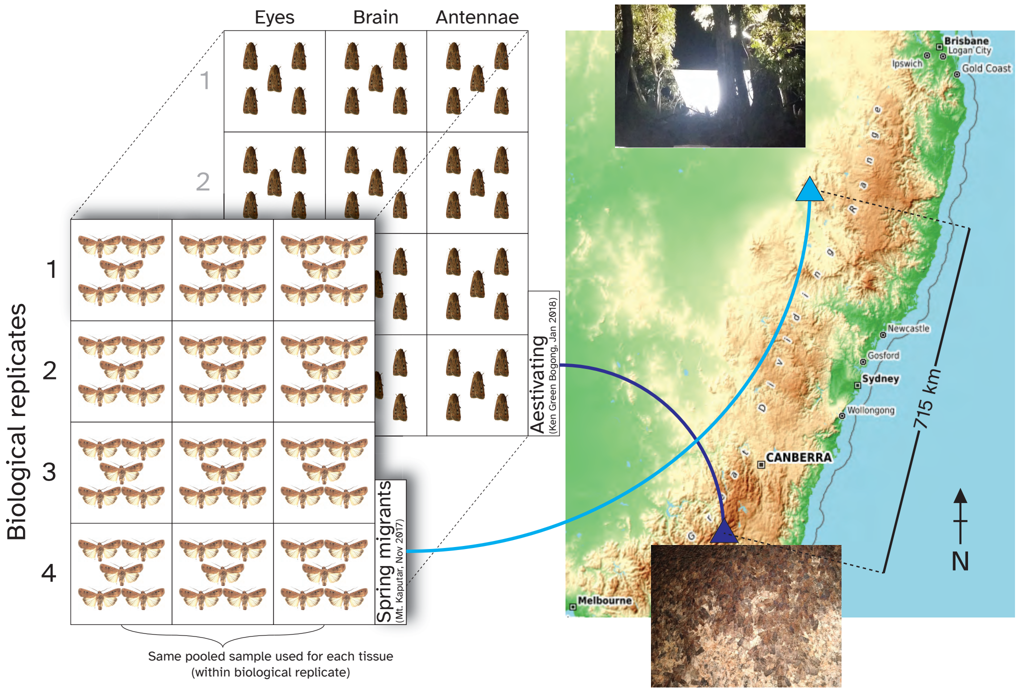


Figure 5.2: Study design of migratory-phase differential expression experiment. Spring-migrating Bogong moths (**front-left panel**) were caught via light-trapping (**right panel, upper inset**) in November 2017 at Mt. Kaputar, NSW (*light blue triangle*), which is along the Bogong moths' migratory route. Aestivating Bogong moths (**back-left panel**) were collected in January 2018 from a cave wall (**right panel, lower inset**) on Ken Green Bogong, NSW (*dark blue triangle*). Individuals were pooled into mixed-sex groups of five, and RNA was extracted from eye, brain, and antennae tissue from each pool separately. Four pools for each condition and tissue were sequenced, for a total of 24 sequenced samples (**left panel**).

5.2.6.1 Alternative splicing analysis

Data from the RNA-Seq experiment were aligned to the genome using *STAR* (version 2.7, Dobin et al., 2013). Differential alternative splicing events were detected from the alignment using the *rMATs* package (Shen et al., 2014). Mapping splice sites to domains was a multi-step process which involved first mapping domains to the Bogong moth proteome, and then mapping splice sites to proteins (and thereby mapping splice sites to domains). Mapping of domains to the proteome (obtained from the genome annotation; Section 5.3.2) was done using *pfam_scan* (available from https://anaconda.org/bioconda/pfam_scan) against *Pfam* (release 34.0, Mistry et al., 2021). Mapping of differential alternative splicing events to domains was performed using a custom R script which depends on the *GenomicFeatures* library (Lawrence et al., 2013).

5.2.7 ASSEMBLY OF *A. BOGONGAE* GENOME

To estimate the expected size of the genome of *A. bogongae*, and confirm the purity of the four sequenced samples, we performed *k*-mer analysis on each of them (Appendix D.2). One of the samples was found to be contaminated, and was excluded from further analysis.

The genomes of each of the uncontaminated samples were assembled separately using *SPAdes* (version 3.15.3, Prjibelski et al., 2020), after performing read trimming using *Trimmomatic* (version 0.39, Bolger et al., 2014). Additionally, an assembly of the samples pooled together was made using *MaSuRCA* (version 4.0.5, Zimin et al., 2013), which was scaffolded using the

SPAdes assemblies using *LINKS* (version 1.8.7, Warren et al., 2015).

5.3 Results and Discussion

5.3.1 GENOME ASSEMBLY

We generated a 595 Mb assembly containing 4,229 contigs. Mean contig size was 38,315 bp and the contig/scaffold N50 was 15.7 Mb. 82% of the assembly is contained in the 31 mega scaffolds/chromosomes (Fig. D.2). The genome is AT-rich (62.1%). Genome completeness was assessed using BUSCO (version 2.0.1, Simão et al., 2015) and the insect odb9 dataset. The results indicated that 96.4% of the reference genes were complete in the assembly. This includes 91.6% single-copy and 4.8% duplicated genes.

5.3.2 ANNOTATION

The annotation includes 19,259 genes, which is towards the upper end for insect genomes. For comparison, the 431 Mb genome of the silk worm *Bombyx mori* has over 20,000 genes, and the postman butterfly *Heliconius melpomene* has 12,829 genes. Putative functions were inferred for 16,132 genes, and for 25,211 mRNAs. 11,116 genes were assigned names according to the *Uniprot Swiss-Prot* reference dataset. Because of the apparent quality of automated annotations, only a few dozen genes of special interest were manually curated. These include the cryptochromes, the yellow protein family, and the DNA methylation toolkit. The coding component occupies 42% of the entire genome. While the average exon length in *A. infusa* is similar to that in other Lepidoptera, the average intron length of 2,947 bp is significantly longer, in line with the idea that intron length correlates with increased genome size (Suetsugu et al., 2013). In a smaller 273 Mb *Danaus plexippus* genome, the mean intron size is 809 bp, and in *B. mori* (genome size 431 Mb), it is 1,904 bp.

5.3.3 GENE ORTHOLOGY AND PHYLOGENETICS

A number of insect species with available annotated genomes (focusing on lepidoptera) were selected for phylogenetic analysis. Gene orthologies were obtained using *OrthoDB* (version 10.1, Kriventseva et al., 2019). A phylogeny based on these orthologies, placing the Bogong moth amongst the selected species, is presented in Fig. 5.3. The Bogong moth was placed near the other noctuid moths in the phylogeny (*Spodoptera* spp., *Helicoverpa armigera*, and *Heliothis virescens*), as expected.

5.3.4 THE TRANSCRIPTOMES OF MIGRATING AND AESTIVATING BOGONG MOTHS

There were 4742 significantly differentially expressed genes (adjusted $p < 0.05$) between spring-migrant and aestivating Bogong moths. Of these, 2451 genes were up-regulated in the spring migrant group, and 2291 genes were up-regulated in the aestivating group.

A principal-component analysis of expression levels across all of the pooled samples showed clear evidence of gene-expression profiles being dominated by tissue type (Fig. 5.4a, *marker shape*). This result reassures us

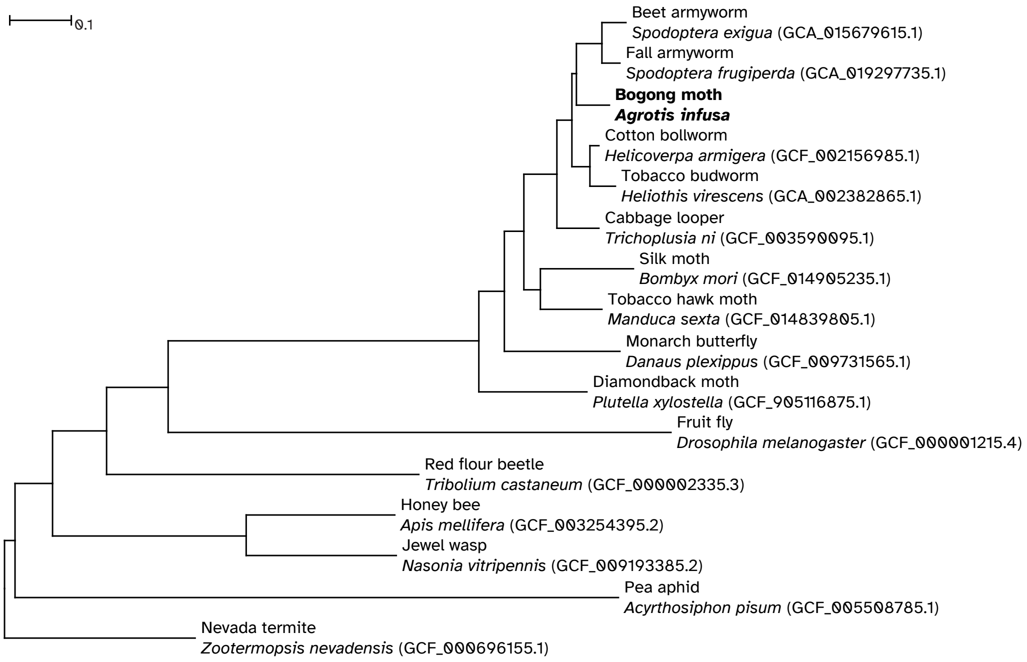


Figure 5.3: Phylogeny of various insect species, including the Bogong moth, with available annotated genomes based on gene orthologies obtained using *OrthoDB* (version 10.1, Kriventseva et al., 2019). Where available, common names are displayed along with binomial names. NCBI accession numbers are shown in parentheses.

that the experiment worked as expected, since antennae, brain, and retinal tissues are made up of very different cell types, and therefore must have very different gene expression profiles. Furthermore, within each tissue type, a clear distinction between spring-migrant and aestivating behavioural phases is evident (Fig. 5.4a, *marker colour*), confirming that gene regulation plays an extremely important role in coordinating these behavioural regimes.

Since the spring-migrant and aestivating samples were (necessarily) collected 715 km—and two months—apart (Fig. 5.2), there is a risk that the observed differential expression between the two groups is a result of genetic variation, rather than regulatory plasticity. To rule out this possibility, we performed population genetics analyses on the RNA-Seq data. We detected ca. 1 million biallelic SNPs from the aligned RNA-Seq reads, which

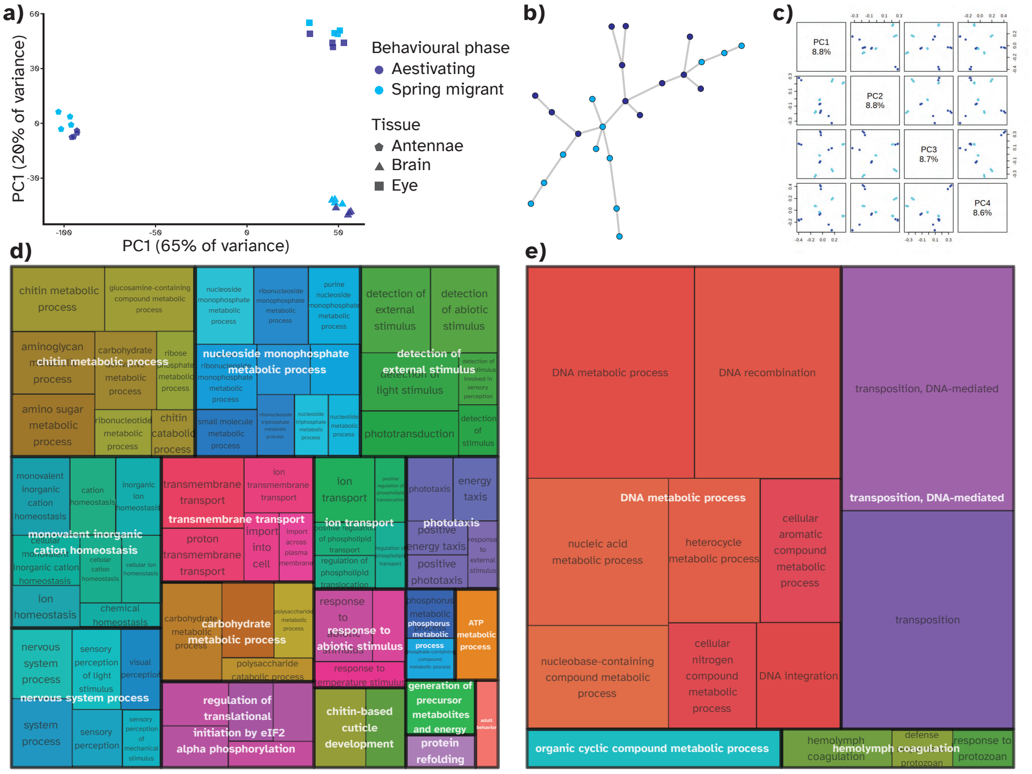


Figure 5.4: Differential gene expression analysis of RNA-Seq reads from 24 pooled samples of antennae, brain, and retinal tissues exposes profound differences in the expression profiles of migrating and aestivating Bogong moths. **a)** Principal component analysis of gene expression shows clear clustering by tissue type (*marker shape*). Within each tissue type, a clear distinction between the expression profiles of migrating and aestivating moths (*marker colour*). **b)** Neighbour-joining tree from 20,000 randomly-selected biallelic SNPs shows no population structure exists between migrating and aestivating moths **c)** Similarly, Principal component analysis of biallelic SNPs reveals no discernible population structure. Samples in *b* and *c* are coloured by behavioural phase, as per *a*. **d)** Tree-map of enriched biological processes terms from GO enrichment analysis of up-regulated genes in sensory and neurological tissues of *spring-migrant* Bogong moths. **e)** Tree-map of enriched gene ontology GO terms from GO enrichment analysis of up-regulated genes in sensory and neurological tissues of *aestivating* Bogong moths.

we reduced to ca. 11,000 SNPs by performing linkage disequilibrium (LD) pruning. We further thinned these data by randomly selecting 20,000 SNPs, and then performed neighbour-joining and principal-component analysis on this reduced set. Both analyses failed to distinguish between behavioural phase (Fig. 5.4b–c), and indeed, no discernible population structure was observed at all, suggesting that the sampled Bogong moths come from a single inter-breeding population. This result serves to increase our confidence that the results of the differential expression analysis represent the interaction of gene regulation with behavioural state, rather than confounding genetic variability.

Since there are over four thousand differentially expressed genes in the spring-migrant and aestivating Bogong moth regulatory regimes, drawing biologically-relevant conclusions about each one of them is intractable. Rather, to summarise the overarching trends in the two regulatory regimes, we performed gene ontology (GO) term enrichment analysis to identify biological processes which are up-regulated in either behavioural phase. In the migratory Bogong moths, a wide variety of processes were up-regulated (Fig. 5.4d). Notably, these included a number of terms which are undoubtedly important for navigation, such as “detection of external stimulus”, “nervous system process”, and “phototaxis”. Furthermore, a host of energy-intensive processes were up-regulated in the migrating moths. These include various metabolic processes, ion transport, and protein synthesis (“regulation of translational initiation by eIF2α phosphorylation”; Fig. 5.4d). Most likely, this result actually reflects selective *down*-regulation in *aestivating* Bogong moths, rather than *up*-regulation in *migrating* moths, *per se*. Indeed, all of these energy-intensive processes are thought to typically be down-regulated in aestivating animals across diverse phyla, to facilitate conservation of resources over long periods of dormancy (reviewed by Storey and Storey, 2010). However, it is unclear to what extent these processes may be up-regulated in migrating Bogong moths versus non-migrating—but also non-aestivating—Bogong moths (e.g. recently eclosed pre-migratory adults, which were not included in this study).

In contrast, relatively few terms were *up*-regulated in aestivating Bogong moths. A similar pattern of reduced expression is observed in hibernating mammals, which tend to exhibit an overall reduction in transcription activity, while selectively up-regulating a small set of genes which are presumably especially important for survival during prolonged torpor (reviewed by Morin and Storey, 2009). In the Bogong moth, these were mostly confined to DNA modification and repair-like processes, under “DNA metabolic process” and “transposition, DNA-mediated” (Fig. 5.4e). These processes act to maintain the integrity of DNA molecules, and the information contained within them. Thus, these are likely to be important for the moths’ longevity, and therefore survival through their long (in terms of a moth’s life cycle) period of summer dormancy, and subsequent return migration to their breeding grounds. In mammals, torpor (typically hibernation) has long been known to be protective against the damaging effects of ionising radiation (e.g. Musacchia and Barr, 1968), which causes damage to DNA (Hutchinson, 1985). This discovery has even led to the—*bona fide*, non-science-fiction—suggestion of using induced torpor to aid in human exploration of outer space (reviewed by Puspitasari et al., 2021). Investigations of gene expression in the brains of hibernating squirrels have shown up-regulation of DNA-repair pathways in the hypothalamus (Schwartz et al., 2013), which could partially explain this protective effect, and appears analogous to our results in the brain and sensory tissue of the Bogong moth.

Interestingly, the term, "hemolymph coagulation" is also up-regulated in the aestivating moths (Fig. 5.4e). This is an important process in insect immune systems, analogous to mammalian blood clotting (Schmid et al., 2019), and is probably up-regulated in response to aestivation-cave-specific pathogens, including the parasitic nematodes, *A. bogongae* and *H. cavicola*, and/or disease-causing bacteria, viruses, and fungi. Indeed, aestivating Bogong moths are estimated to cluster with densities up to 17,000 moths/m^2 (Common, 1954), so it is not hard to imagine that communicable diseases would spread easily in these populations.

5.3.5 Alternative splicing

Alternative splicing (AS) is one mechanism by which a fixed number of genes can generate a multitude of protein isoforms via shuffling exon inclusions in mRNAs. This process occurs with a very high frequency in the nervous system and contributes to development, axon guidance, synaptogenesis, etc. (reviewed by Furlanis and Scheiffele, 2018; Su et al., 2018). We examined our RNA-Seq data to determine the extent to which AS is utilised in the brains, antennae, and eyes of Bogong moths, and if there are differences in the AS variants between spring migrants and aestivating Bogong moths. In total, 102,212 spliced junctions were analysed. Of these, 14,537 (14.2%) were found to be alternatively spliced. As shown in Fig. 5.5a–c, there are easily detectable AS events not only between tissues, but also between migratory phases. One example of these AS events in the antennae is visu-

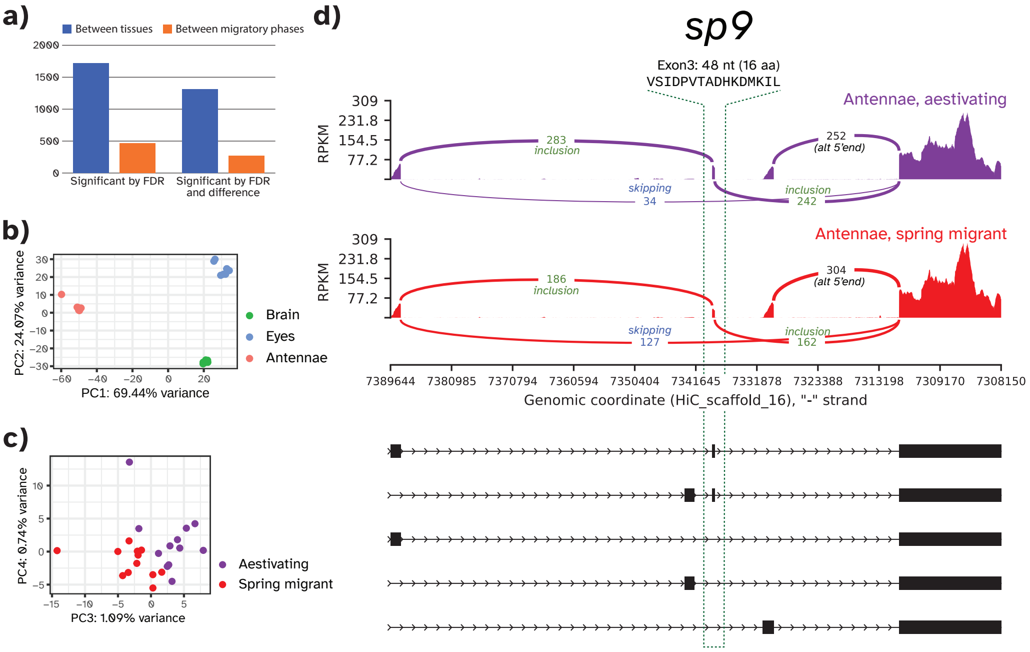


Figure 5.5: Alternative splicing (AS) in the Bogong moth transcriptome. **a)** Average number of differential AS events between tissues and migratory phases. **b)** First two principal components of AS variation show clear distinction between tissue types. **c)** Third and forth principal components of AS variation show clear distinction between migratory phase. **d)** Example of an AS event in the antennae. A larger portion of *sp9* mRNAs with exon 3 skipped is seen in spring migrants (*red*), when compared with aestivating moths (*purple*). The full set of *sp9* isoforms are shown in *black*.

alised in more detail in Fig. 5.5d. It shows the differential inclusion of exon 3 in the gene encoding *sp9*, a transcription factor that mediates expression of several brain genes. A larger portion of *sp9* mRNAs with exon 3 skipped is seen in spring migrants, when compared with aestivating moths.

5.3.6 DNA METHYLATION

Methylated cytosines in the CpG context (mCpGs) have been found in 66.53% of the currently annotated genes. Assuming that all annotated genes encode proteins, this is a relatively large proportion of methylated genes for an insect. Relative to other insects, the Bogong moth's global genomic methylation is similar or higher (Bewick et al., 2017). Because Lepidoptera have only one type of DNA methyltransferase—the so-called maintenance DNMT1—and lack DNMT3 that adds new methyl tags to DNA (Werren et al., 2010), the underlying mechanism of this modification in Lepidoptera is not fully understood (Lyko and Maleszka, 2011). One possibility is that DNMT1 has dual *de novo* and maintenance activity, or there are other novel enzymes providing the DNMT3-type activity (Wedd and Maleszka, 2016).

Fig. 5.6a shows the distribution of mCpGs within the genomic regions. There are more mCpGs in introns than in exons, and quite surprisingly, a lot in promoters. In other insects like the honey bee, virtually no promoter methylation has been found, with most mCpGs located in exonic regions and only a small amount in introns (Foret et al., 2012; Lyko et al., 2010). There are some indications that intragenic CpG methylation could alter AS by affecting elongation polymerase efficiency, giving more time for upstream splice sites to assemble a functional spliceosome before having to compete with downstream sites (Perales and Bentley, 2009), although the function of CpG methylation in intergenic and promoter regions of insect genomes is unknown (Glastad et al., 2014).

Another unexpected feature of DNA methylation in the Bogong moth is a high level of non-CpG methylation (Fig. 5.6b). In several previously analysed insect methylomes, no methylation in this context was detectable, leading to the conclusion that all insects—and indeed, all invertebrates—do not have this type of epigenomic modification (Mendoza et al., 2021). Our finding contradicts this notion and suggests that in some insects non-CpG

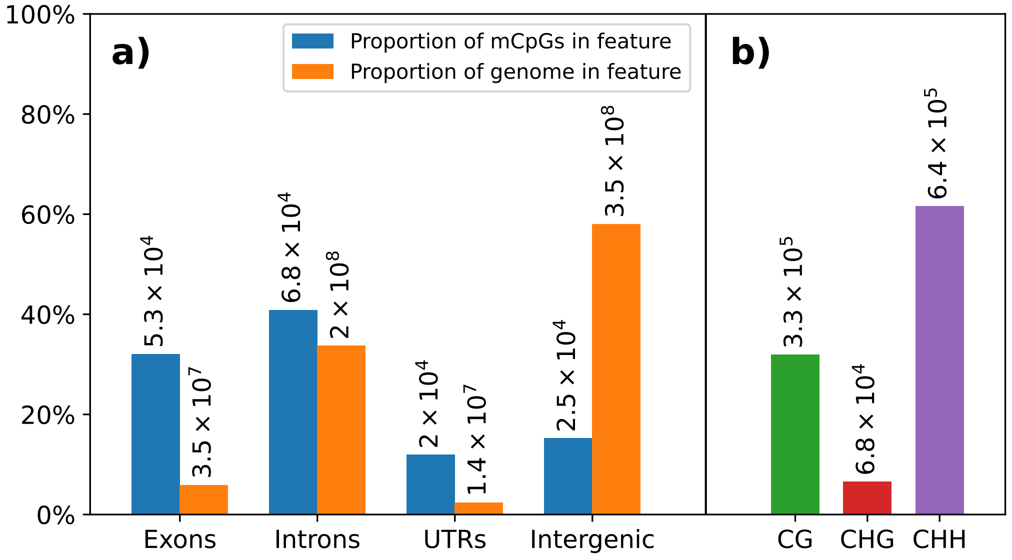


Figure 5.6: Genomic context of methylated DNA in the Bogong moth. **a)** Distribution of mCpGs located in genomic features (*blue*), shown with the proportion of the genome covered by those features (corrected for overlaps; *orange*). **b)** Distribution of sequence contexts of methylated cytosines (H = any of A, T, or G). In both plots, the value leading to each proportion is displayed above each bar (e.g. 3.5×10^7 bases are in exons, representing about 5.9% of the 5.9×10^8 bases in the entire genome).

methylation is prominent and might be functionally relevant, specifically in brain functions, as proposed for vertebrates (*ibid.*).

Our discovery of these unusual patterns of DNA methylation in the genome of the Bogong moth suggest that it may have novel and yet-to-be-determined functions. This provides grounds for renewed investigations into the functional role of DNA methylation in insects more generally (Lyko and Maleszka, 2011), and in particular in our newly established model system—the Bogong moth.

5.3.7 THE GENOME OF *A. BOGONGAE*

Assembly of the genome of *A. bogongae* from short reads yielded a 301 Mb assembly, with 56,615 contigs, a maximum contig length of 269 kb, and contig N50 of 26.5 kb. The assembly contained 36.6% of the nematode BUSCOs (34.5% single-copy, 2.1% duplicated). Given the low quality of the assembly—which is a result of only using short reads—we did not attempt to perform genome-wide annotation.

However, we were able to ascertain the presence of various genes by performing TBLASTN searches of coding sequences from related nematode species, *Romanomermis culcivorax* (NCBI accession: GCA_001039655.1), *Trichurus muris* (NCBI accession: GCA_000612645.2), and *Caenorhabditis elegans* (NCBI accession: GCF_000002985.6), and their corresponding annotations. To do this, nucleotide sequences of CDS (exons) were extracted from each genome using a custom R script, which depends on Biostrings (Pagès et al., 2013) and GenomicFeatures (Lawrence et al., 2013) libraries. The exonic sequences were then translated using ORF phase information from the genome annotation files, and these (amino acid) sequences were used to query the *A. bogongae* genome assembly using TBLASTN (Gertz et al., 2006) with E-value cutoff set to 0.1.

Surprisingly, we found that *A. bogongae* possesses the full invertebrate DNA methylation toolkit. This opens the possibility that the parasitic nematode uses epigenetic modification to hijack the epigenome of the Bogong moth, in order to bypass immune responses (Silmon de Monerri and Kim, 2014). Another possibility is that the nematode uses epigenetic modification to aid in regulating its genome during winter, when it presumably hibernates while it waits for the return of its Bogong moth host the following spring.

These results, while preliminary, indicate interesting avenues for further exploration of the interactions of *A. bogongae* and the Bogong moth.

An obvious next step would be to supplement the short read sequencing data with long reads, and reads from bisulfite-converted DNA, as we have done with the genome of the Bogong moth. This would allow a more complete overview of the nematode's genome, and insights into the function of its methylation toolkit.

5.4 Conclusion

We have performed deep, hybrid sequencing and *de novo* assembly of the Bogong moth genome. Additionally, we have sequenced the transcriptome of the Bogong moth, and used this with a protein database to generate high-quality functional annotations for over 16 thousand genes. We have discovered a profound difference in the transcriptional profiles of Bogong moths during different phases of their remarkable round-trip migration to the Australian Alps, confirming that gene regulation—at both the gene expression and alternative splicing levels—plays an extremely important role in coordinating their migration and aestivation. Furthermore, we have found that a high level of DNA methylation is present in the Bogong moth genome, including CpG methylation in all sequence types, and surprisingly, large amounts of non-CpG methylation, contradicting the recent proposal that this type of DNA modification is a vertebrate invention (Mendoza et al., 2021).

Our results provide the foundation for research into the molecular basis of the Bogong moth's impressive navigational abilities, including investigations into the source of their inherited migratory direction, geomagnetic sense, and their unique interaction with cave-dwelling parasitic nematodes.

5.5 Author contributions

EJW, KB, and SH conceived the project. RM and EZ provided advice on the analytical strategy. KB optimised the moth rearing protocol and designed the RNA-Seq experiment. JRAW and EJW performed the fieldwork for the RNA-Seq experiment, and made other sample collections. RK prepared the bisulfite sequencing libraries. JRAW prepared the nematode libraries. JRAW, KB, RK, JML, AMC, EPW, LS, MM, BB and PC analysed the data. JRAW, RK, JML, EJW and MM produced the figures in the main text. JRAW, RM, and EJW wrote and edited the first draft of the manuscript. KB, RK, AMC, EPW, LS, and PC provided additional text relating to methodological details.

5.6 Acknowledgements

EJW and JRAW are grateful for funding from the European Research Council (Advanced Grant No. 741298 to EJW), the Air Force Office of Scientific Research (Grant No. FA9550-14-1-0242 to EJW), the Swedish Research Council (Grant No. 2016-04014 to EJW) and the Royal Physiographic Society of Lund (to JRAW). JRAW is thankful for the support of an Australian Government Research Training Program (RTP) Scholarship. EJW holds Scientific Permits for collection and experimental manipulations of Bogong moths in several alpine national parks and nature reserves (NSW: Permit SL100806; Vic: Permit 10008966). The authors acknowledge support from the National Genomics Infrastructure (NGI) Genomics Applications, Stockholm, funded by Science for Life Laboratory, the Knut and Alice

Wallenberg Foundation and the Swedish Research Council. We also acknowledge SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science, NGI/Uppsala Genome Center, and UPPMAX for assistance with massively parallel sequencing and computational infrastructure. Work performed at NGI/Uppsala Genome Center has been funded by RFI/VR and Science for Life Laboratory, Sweden. Work performed at the SNP&SEQ Technology Platforom (part of NGI and Science for Life Laboratory) is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. We would like to thank Joanna Maleszka for identification of the nematode species, and for advice regarding insect dissection, and David Szakal for assistance with fieldwork.

Chapter 6

Large-scale whole-genome sequencing of migratory Bogong moths *Agrotis infusa* reveals genetic variants associated with migratory direction in a panmictic population.

Jesse RA Wallace[1,2], Ryszard Maleszka[1] and Eric J Warrant[1,2]

[1]Research School of Biology, Australian National University. [2]Lund Vision Group, Department of Biology, Lund University, Sweden.

Abstract

One of the most interesting macroscopic phenomena in the animal world is seasonal migration. A central goal of research into animal migration is to better understand the mechanisms that evolved to solve the complex challenges which a migratory life history presents. Each year, and with a high degree of species-level site fidelity, the Australian Bogong moth makes a return migration of up to and over 1000 km between widely distributed breeding grounds and a specific set of aestivation sites in the Australian Alps. It does this without any opportunity to learn the migratory route or the location of the aestivation sites from either older generations or repeated migrations, meaning that the information required by the moth to navigate during its migration must be inherited. The migratory direction, and therefore the inherited navigational information in Bogong moths, varies with breeding site, providing us with an opportunity to search for the source of that heritability by comparing the genomes of moths collected from different breeding areas. We successfully sequenced whole nuclear genomes of 77 Bogong moths collected from across their breeding grounds and summer range, and found that the Bogong moth population contains a large amount of (mostly rare) variation. We found no evidence of population structure, indicating that Bogong moths are panmictic. A genome-wide scan for signals of selection indicate that the Bogong population has recently recovered from a past bottleneck, however genomic regions which have likely undergone balancing selection were also detected. Despite panmixia, four genetic variants in breeding-ground-caught Bogong moths were found to be significantly associated with geographic location, and therefore migratory direction. While these results do not imply a causative link between genetic variation and the migratory direction phenotype, they do indicate promising future avenues of research into the molecular basis of long-distance navigation.

6.1 Introduction

Migratory behaviour is one of the most conspicuous and fascinating phenomena in the natural world (Dingle, 2014). It has evolved independently in many taxa across the animal kingdom, despite frequently requiring complex physiological and neurological adaptations for successful execution (Chapman et al., 2015). These adaptations are particularly pronounced in animals that navigate over great distances from a distinct origin to a specific destination, and that return to the origin after the season changes, especially if the origin and/or destination remain stable for an individual or population from one year to the next. For such fidelity to be possible, the migrant must either learn its migratory destination or inherit the information required to find it.

The Australian Bogong moth *Agrotis infusa* is a wonderful example of an animal that migrates in an extraordinarily directed and precise manner (Fig. 6.1) (reviewed by Warrant et al., 2016). The migratory journey of the adult Bogong moth starts in their breeding grounds, the dry plains of southern Queensland, western NSW, western Victoria, and eastern South Australia (Common, 1954; Warrant et al., 2016), a vast arc of country that spans at least 7° of latitude and 9° of longitude—1,400 km from end to end. In spring, the moth leaves its breeding area and flies up to 1,000 km to a specific set of sites located in the mountaintops of the Australian Alps, a narrow strip of alpine territory stretching between Mt. Gingera in the north and Mt. Buller in the south, a summer range which spans less than 2° of latitude and 2° of longitude, or some 274 km from end to end. Thus, the Bogong moth's breeding grounds are vast relative to their summer range.

This means that the required direction of migration must vary, depending on the starting point in the breeding grounds (*arrows* in Fig. 6.1) (Warrant et al., 2016). For instance, moths migrating to the Australian Alps from southern Queensland must fly south, whereas moths migrating to the Australian Alps from eastern South Australia must fly east.

Once they have reached the Australian Alps, Bogong moths seek out the cool cracks and crevices of particular granite outcrops dotted across the mountain ridges. Here they spend the summer huddled together in a dormant state known as aestivation (Common, 1952). In autumn, they leave these outcrops and return to their breeding grounds, where they breed, lay their eggs, and die, ending their migratory journey (Green, 2010b). The following generation of Bogong moths will return to the same set of summer sites as their ancestors, sites known to have been repeatedly and consistently occupied by Bogong moths for millennia (Keaney et al., 2016; Stephenson et al., 2020). This journey is achieved despite the moth having had no opportunity to learn the location of those sites from the previous generation. The information the moth requires to reliably navigate to its alpine destination during the spring migration must therefore be inherited.

The impressive migration of the Bogong moth, along with its abundance and accessibility, has seen it become an important emerging model for long-distance nocturnal navigation, subject to fruitful and ongoing research which has begun to unravel the sensory and neurobiological mechanisms that the moth uses to successfully migrate (Adden et al., 2020b; Dreyer et al., 2018; Warrant et al., 2016). However, there remains an important open question which this study aims to address: How are naïve moths capable of

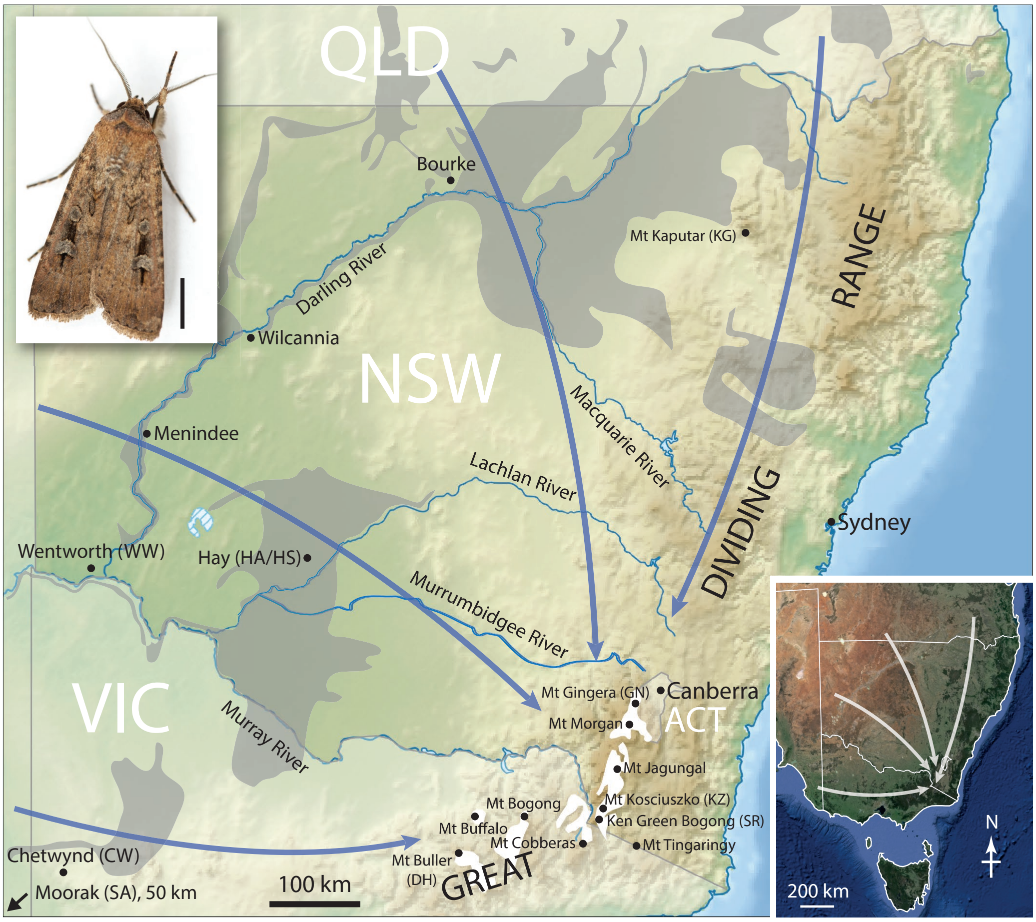


Figure 6.1: Map of Bogong moth migratory routes showing the names of the major rivers in New South Wales (*NSW*), and the locations of various towns and mountain peaks, mentioned in the text (shown in the context of continental south-eastern Australia: *inset, lower right*). Areas of grey cracking clays—favoured soils for Bogong moth winter development—are shown in *grey*. Bogong moths fly in different directions (*arrows*) towards the Australian Alps from various regions of south-eastern Australia, from as far distant as eastern South Australia, western Victoria (*VIC*), western and north-western New South Wales, and southern and south-eastern Queensland (*QLD*). The *white areas* represent elevations above 1500 m, where all known summer aestivation sites are located. *ACT* = Australian Capital Territory. Adapted from Green et al. (2021). **Inset, upper left:** The Australian Bogong moth *Agrotis infusa* (Boisduval, 1832). Scale bar = 5 mm. Reproduced with the kind permission of the photographer; Ajay Narendra, Macquarie University, Australia.

using their navigational toolkit to migrate from broadly distributed breeding grounds to a relatively restricted area with such remarkable species-level site fidelity?

Previously, the degree of linkage between specific regions in the Bogong moth breeding grounds and their migratory destination (i.e. the aestivation sites)—a concept known as 'migratory connectivity'—was not even known (Gao et al., 2020). Under a scenario where migratory connectivity in the Bogong moth is high, Bogong moths would have a propensity to return after aestivation to the specific region of their breeding grounds where they hatched months earlier. We would then expect genetic structure to develop across the Bogong moth population, as gene flow would necessarily be limited between subpopulations from different breeding regions. Alternatively, if migratory connectivity is low, there would be a high degree of mixing across the Bogong moth population, which would lead to an absence of population structure.

The degree of migratory connectivity present in Bogong moths has important implications for how Bogong migratory direction is inherited. A high-migratory-connectivity scenario would readily facilitate a regime favouring a genetically-determined migratory direction, akin to that seen in migratory songbirds (e.g. Lundberg et al., 2017)—Bogong moths would then simply inherit a fixed migratory direction from their parents. However, in a low-migratory-connectivity scenario, additional information would be required in order for the Bogong moth to successfully complete its migration. This information could come in the form of sensory cues, perhaps relating to the environment in which the moth hatches, or, as in the high-connectivity scenario, it could come in the form of a genetically-heritable fixed spring-migratory direction. Whether or not this inherited migratory direction is correct (for a given breeding ground region) would only be determined once the merciless process of natural selection has acted. This latter possibility may be reasonable, given the high fecundity of Bogong moth females, which can each lay up to 2,000 eggs (Warrant et al., 2016).

In this study, we aim to determine whether Bogong moth migratory direction is genetically heritable, and if so, to infer the putative sources of its heritability. In doing so, we also aim to shed light on the level of migratory connectivity in the Bogong moth. We will proceed by taking advantage of the research opportunities presented by the recently sequenced Bogong moth reference genome (Chapter 5). By re-sequencing the whole genomes of 77 Bogong moths collected from locations distributed across their entire breeding grounds and their summer aestivation range, we will search for genetic variation that could explain the variation in their migratory directions. As is the case for most genome-wide association studies, the design of this study does not permit the detection of causative links between genotype and migrtory direction. However, by finding associations it reduces the search-space for future work with this aim.

In addition to progressing our understanding of Bogong moth migration, these population genetics data will also prove useful for understanding Bogong moth ecology more generally. The Bogong moth has the peculiar status of being both endangered (Warrant et al., 2021), and simultaneously considered a pest (Common, 1954; Farrow and McDonald, 1987, see also https://moths.csiro.au/species_taxonomy/agrotis-infusa/). This duality is a direct result of its migratory life history—in summer adult Bogong moths provide essential nutrients and energy to the Australian Alpine ecosystem (Green, 2011), but in winter Bogong cutworms are known to cause damage to crops in their breeding grounds (Common, 1990). Regardless of whether management objectives are conservation or control, an understanding of the structure of the Bogong moth population is fundamentally important for informing strategies to achieve them.

6.2 Methods

6.2.1 Sample material

In the Southern Hemisphere spring months of 2017-2019, living Bogong moths were sampled from various locations across their breeding grounds and spring migratory routes. Moths were attracted to a white sheet suspended in a tree and illuminated by a 1000 W xenon searchlight and/or a LepiLED lamp (Brehm, 2017). Bogong moths were collected by hand using sample jars. Moths were also collected during the summers of 2017-2020 from Bogong aestivation sites in the alpine regions of New South Wales and Victoria. The whole moth samples were fixed in absolute EtOH shortly after collection, and were stored at either room temperature or 4°C. This eliminated logistical issues associated with attempts to keep samples at low temperatures in the field, and proved effective in maintaining the quality of the samples' DNA. The collection locations are listed in Table 6.1, and are also included on the map in Fig. 6.1.

Table 6.1: Sample collection locations. Samples are grouped by two-letter abbreviations (Abbr.). Migratory condition (Cond.) is noted as either spring migrant (SM) or aestivating (A). n denotes the number of samples sequenced for each group, and n' denotes the number of samples which passed sequencing quality control.

Abbr.	Location	Cond.	Lat/Lon	Date	n	n'
CW	Chetwynd	SM	37°14'S 141°30'E	10-2018	13	11
DH	Mt Buller	A	37°09'S 146°26'E	11-2018	12	5
GN	Mt Gingera	A	35°34'S 148°46'E	02-2019	5	4
HA/HS	Hay	SM	34°33'S 144°52'E	10-2017	9	9
KG	Mt Kaputar	SM	30°17'S 150°09'E	11-2017	22	19
KZ	Mt Kosciuszko	A	36°27'S 148°16'E	02-2021	8	8
SA	Moorak	SM	37°52'S 140°45'E	10-2018	12	11
SR	Ken Green Bogong	A	36°31'S 148°15'E	02-2018	10	9
WW	Wentworth	SM	34°05'S 141°54'E	10-2018	1	1

6.2.2 DNA EXTRACTION

To reduce the chance of contamination, brain and thoracic muscle tissue were dissected from the moth samples and stored at -80°C prior to DNA extraction. DNA was extracted from the brain and muscle tissue separately using a *Quick-DNA™ MagBead Plus kit* (Zymo Research, 2017) and quantified using a *Qubit™ dsDNA HS Assay Kit* on a *Qubit™ 3 Fluorometer* (Thermo Fisher Scientific, 2017) using the manufacturers' respective protocols. Results from the first few extractions indicated that DNA yield was significantly higher from the thoracic muscle tissue when compared with the brain tissue, so for the remainder of the samples DNA was extracted from the muscle tissue only.

6.2.3 Genomic DNA sequencing

A total of 92 Bogong moth samples were selected for sequencing, ensuring sufficient DNA yields for each selected sample, as well as adequate coverage of their geographic range in both the breeding grounds and their mountainous summer aestivation range. Sequencing libraries were prepared using the *Illumina® DNA PCR-Free Prep, Tagmentation* with the *IDT® for Illumina® DNA/RNA UD Indexes Set A, Tagmentation* (Illumina, 2020) using the manufacturer's protocol, and a unique index was used for each sample.

The DNA content of the libraries were quantified using a *Qubit™ ssDNA Assay Kit* on a *Qubit™ 3 Fluorometer* (Thermo Fisher Scientific, 2017), then pooled in equal DNA proportions. To ensure library quality, the pooled library was first sequenced on the Illumina MiSeq platform with a target of 1 million reads, and quality control checks were performed. Based on the read counts from the MiSeq run, the libraries were re-pooled and sequenced on the Illumina NovaSeq platform, for a target of 10-20x average coverage per sample.

6.2.4 Sequencing quality control

For both the MiSeq and NovaSeq runs, quality control was performed using a custom pipeline written in Snakemake (Mölder et al., 2021). Read trimming was performed using Cutadapt (Martin, 2011). To assess contamination levels, trimmed reads were classified using Kraken 2 (Wood et al., 2019) with the National Center for Biotechnology Information (NCBI) non-redundant nucleotide (nt) database, and were aligned to the Bogong reference genome (Chapter 5) using BWA-MEM 2 (Vasimuddin et al., 2019). General quality control statistics were obtained using FastQC, SAMtools (Li et al., 2009), and Picard tools "collectinsertsizemetrics". Summary reports were generated from the output of the above software using MultiQC (Ewels et al., 2016).

6.2.5 Variant calling

Sequence variant (SNP/INDEL) calling was performed using the data from both the MiSeq and NovaSeq runs, using a slightly modified version of the variant-calling pipeline, Grenepipe (Czech and Exposito-Alonso, 2021). Read trimming was performed using Trimmomatic (Bolger et al., 2014), mapping was performed using BWA-MEM 2 (Vasimuddin et al., 2019), and variant calling was performed using GATK haplotypecaller (McKenna et al., 2010).

Exon drop-out variants were identified by assessing read coverage of exons in the reference genome annotation, based on the mapping generated before the sequence variant calling step. Exons which had no reads mapped to them from a given sample were inferred to be missing from that sample's genome. To reduce incidence of false discoveries of exon drop-outs, only samples with an average read-depth across all exons above a certain threshold were considered. From our analyses, an average read-depth of 10x is sufficient to bring the false-discovery rate for most exons to a vanishingly small value, well below 10^{-5}, however lower average read-depths perform considerably worse, so a threshold of 10x was used (see Appendix E.1 for details of the analysis).

Transcript drop-out variants were identified in a similar fashion to exons. For further analyses, only a single representative transcript for each unique pattern of presence across the samples within each gene was used. That is, if a transcript shared identical presence/absence information with another transcript from the same gene which had already been included, then the former would be discarded from subsequent analyses. In this way, transcript drop-outs approximate gene drop-outs.

6.2.6 Population structure analysis

Population structure was assessed based on the sequence variant data using Structure_threader (Pina-Martins et al., 2017) and fastStructure (Raj et al., 2014) for *k* values ranging from 1–12. Principal component analysis was also performed on the sequence variant data. Clustering based on the hamming distance from sequence variant data and transcript drop-out variant data was also done using the neighbour-joining algorithm.

6.2.7 Genome-wide association study of migratory direction

To determine if any variants are associated with migratory direction, we performed a genome-wide association study (GWAS) on spring migratory direction. To perform the analysis, spring migrants were classified as either western (Chetwynd, Hay, Moorak, and Wentworth samples) or northern (Mt Kaputar samples). Summer aestivating moths (Mt Gingera, Mt Kosciuszko, and Ken Green Bogong samples) were not included in the GWAS, as it is not known from which direction they migrated in the preceding spring. Association analysis was performed with univariate linear mixed models using GEMMA (Zhou and Stephens, 2012). P-values were calculated using a likelihood ratio test and a significance threshold of 0.05 with Bonferroni correction was applied. To limit the study to variants which could have large effect sizes, we set a strict minimum minor allele frequency threshold of 0.3.

6.3 Results

6.3.1 DNA sequencing and quality control

Of the 92 Bogong moth samples selected for sequencing, 77 passed all quality control steps (refer to Table 6.1 for the locations of the successful samples). The failed samples were excluded due to high levels of contamination or sample drop-out during sequencing. We believe the sample drop-outs were caused by a technical error which resulted in one column (8 samples) of the library preparation failing. Summaries of the results of quality control on the main sequencing run are shown in Appendix E.2 Fig. E.2.

From the Kraken 2 read classification analysis, we identified three species of bacteria which occurred in large quantities or in multiple samples. These were *Providencia rettgeri*, *Alcaligenes faecalis*, and *Serratia* *marcescens*. Interestingly, *P. rettgeri* is known to be pathogenic to insects, and is carried by insect-parasitic nematodes of the genus *Heterorhabditis* (Jackson et al., 1995). Adult Bogong moths are known to be parasitised by two species of mermithid nematodes (Common, 1954; Welch, 1963), although it is not known if these are also carriers of *P. rettgeri*.

6.3.2 Variant calling

A total of 186,622,107 sequence variants were discovered, representing approximately one variant for every 3.18 bases in the reference genome, exceeding the notable levels of variation recently reported in the diamondback moth *Plutella xylostella*, which has approximately one variant every 6 bases (You et al., 2020). Of the discovered sequence variants, 153,363,045 were SNPs and 33,259,062 were INDELs. Approximately half of the sequence variants were singletons (74,336,046 SNPs and 19,300,939 INDELs occurred in only a single sample).

A total of 11,252 exon drop-out variants and 846 transcript drop-out variants were discovered. Of the transcript drop-out variants, 258 were singletons. We performed gene ontology enrichment analysis on these 846 transcripts and found a small number of terms which were significantly enriched (greater number of occurrences than would be expected by chance), and a single term which was significantly purified (fewer occurrences than would be expected by chance), when compared to the rest of the Bogong moth gene set ($p < 0.05$, Fisher's exact test, Bonferroni corrected). These terms are presented in Table 6.2.

Table 6.2: Gene ontology terms which were significantly enriched or purified in the set of transcripts which were not present in every Bogong moth sample sequenced ($p < 0.05$, Fisher's exact test, Bonferroni corrected).

Term	Name	p
Enriched terms		
Biological process		
GO:0015074	DNA integration	3.58×10^{-26}
GO:0006915	apoptotic process	1.73×10^{-6}
GO:0006313	transposition, DNA-mediated	9.28×10^{-9}
GO:0006310	DNA recombination	2.81×10^{-10}
Cellular component		
GO:0000786	nucleosome	4.73×10^{-25}
Molecular function		
GO:0004803	transposase activity	2.06×10^{-9}
GO:0046982	protein heterodimerization activity	6.94×10^{-30}
GO:0003676	nucleic acid binding	3.28×10^{-20}
GO:0003677	DNA binding	1.74×10^{-19}
Purified terms		
Molecular function		
GO:0005515	protein binding	6.25×10^{-11}

6.3.3 Population structure

No evidence of population substructure was found based on sequence variant (Fig. 6.2a–e) or transcript drop-out variant (Fig. 6.2f) data. Structure analysis showed that the data are best explained by a single panmictic population (Fig. 6.2a–c). This was supported by principal component analysis (Fig. 6.2d) which failed to produce any meaningful sub-population clusters. The neighbour-joining algorithm applied to sample-pair-wise hamming distances of sequence or transcript drop-out variants also failed to indicate any meaningful population structure (Fig. 6.2e–f).

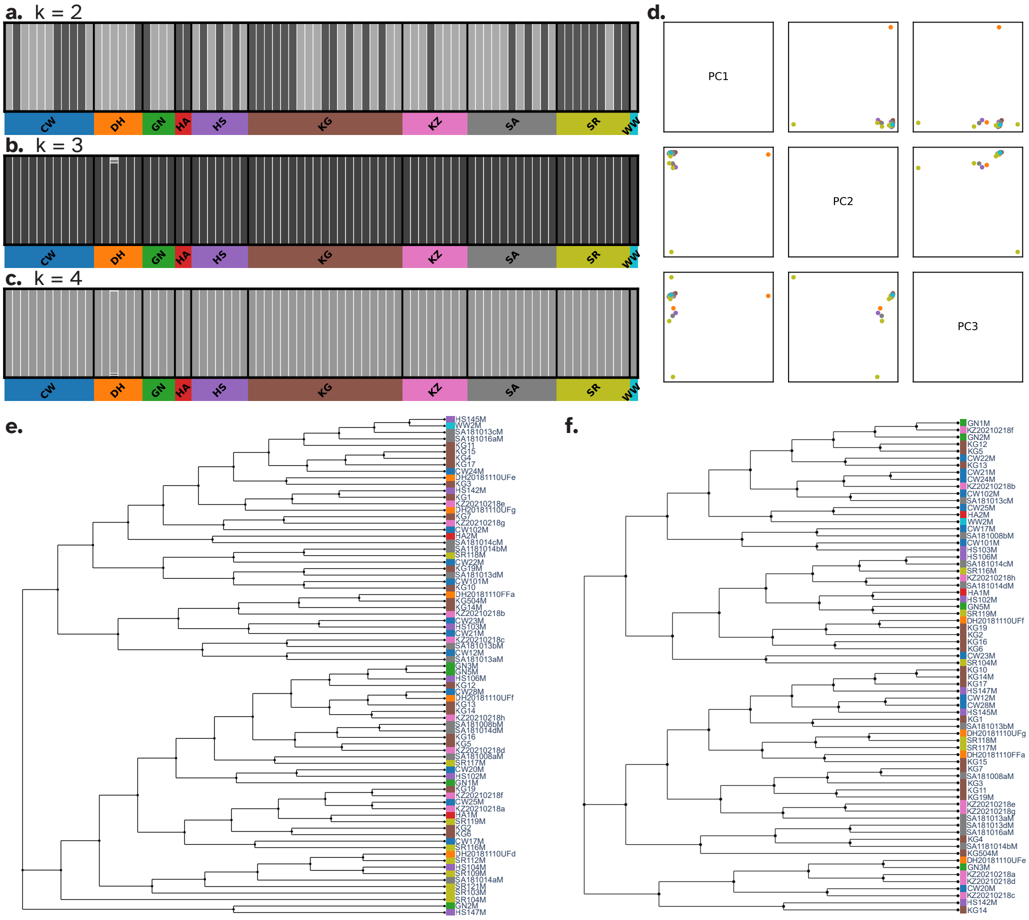


Figure 6.2: No evidence of population structure was detected in variant data from the largest 31 scaffolds of the Bogong moth reference genome, amongst samples collected from across the Bogong moth breeding grounds and summer aestivation range. **a–c.** Structure plots with values of k set to 2, 3, and 4 (k values of 1,5–12 were also tested, but have been omitted for brevity) produced using fastStructure (Raj et al., 2014). Two-letter geographical location abbreviations are defined in Table 6.1 and shown in Fig. 6.1. The data were found to be best explained using $k = 1$ (i.e. a single panmictic population, although the plot for $k = 1$ is not shown as it is uninformative, since $k = 1$ is the trivial case). **d.** The first three principal components of sequence variants. Samples are coloured by location, as in a–c. We did not observe any clear clustering into Bogong moth subpopulations. **e–f.** Neighbour-joining dendrograms from hamming distance of sequence variant (**e**) and transcript drop-out variant (**f**) data. Samples are coloured by location, as in a–c. No evidence of population stratification was detected in either dataset.

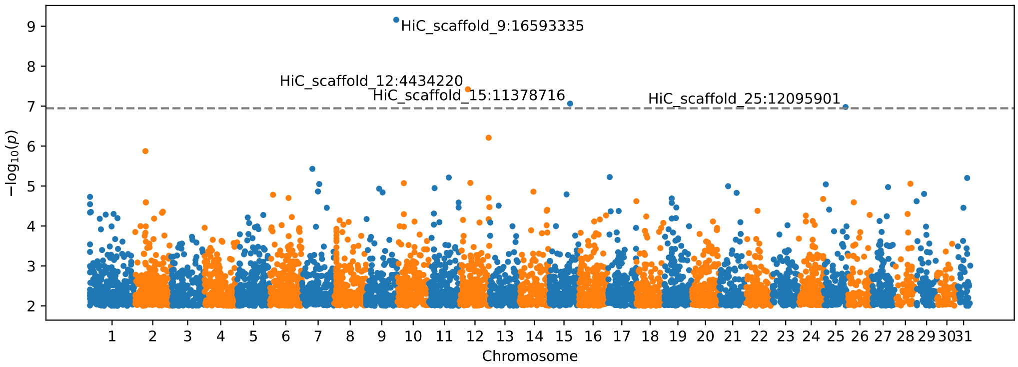


Figure 6.3: Manhattan plot of sequence variant associations with spring migration orientation as a categorical variable. Spring migrants were categorised as either “south-flying” or “east-flying”, depending on the location where they were caught (see Section 6.2). *Dashed line* indicates Bonferroni p-value threshold of 1.13×10^{-7}. The four significantly associated variants are labelled. Loci with p-values greater than 0.01 are omitted. Points are coloured according to which scaffold (chromosome) they appear in, with odd-numbered scaffolds coloured blue and even-numbered scaffolds coloured orange.

6.3.4 GENOME-WIDE ASSOCIATION STUDY

Four sequence variants were found to be significantly associated with migratory direction to the Bonferroni-corrected significance level of $p < 1.13 \times 10^{-7}$ (Fig. 6.3). We refer to these variants as AiSNP_01, AiSNP_02, AiSNP_03, and AiSNP_04. Summaries of the genotype proportions of these significantly-associated loci in each of the sample collection locations are shown in Fig. 6.4. Summaries of the genomic context of each variant, with reference to functional information from orthologs studied in other species, are provided in Table 6.3.

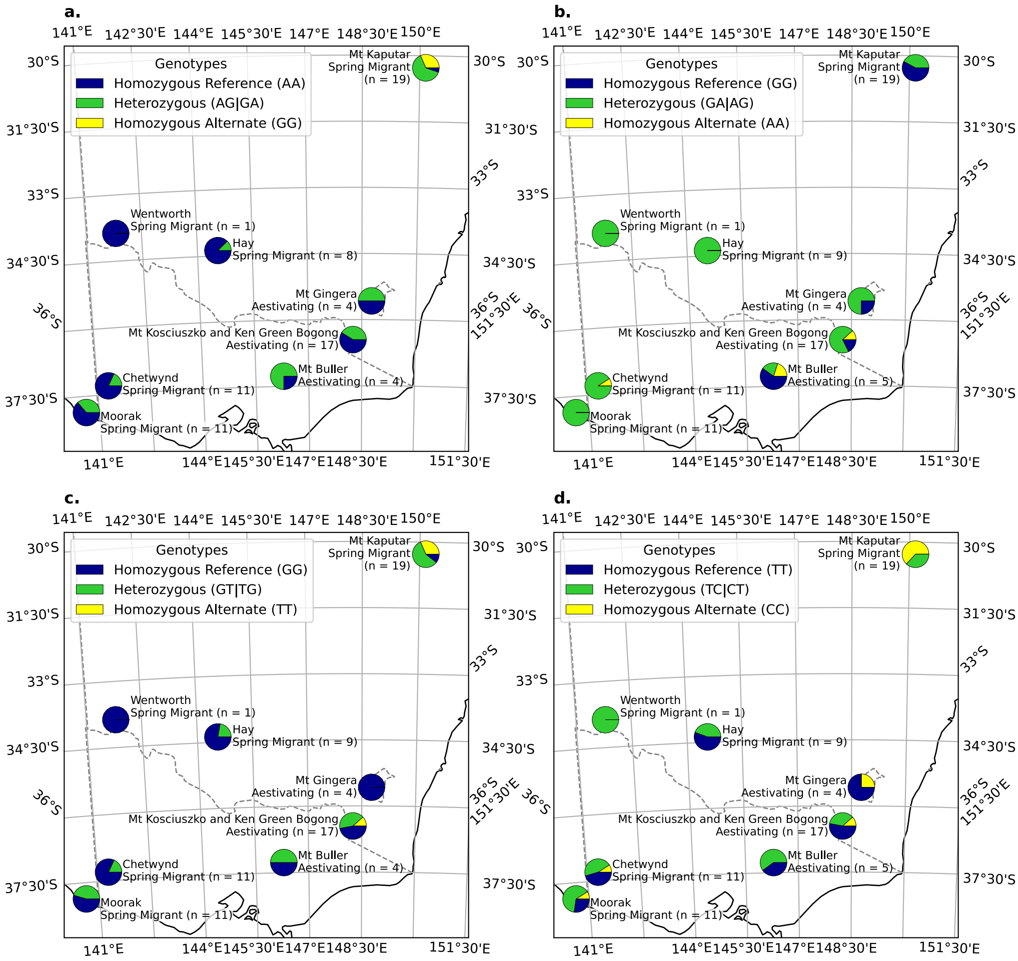


Figure 6.4: Genotype proportions for sequence variants significantly associated ($p < 0.05$, Bonferroni-adjusted $p < 1.13 \times 10^{-7}$) with spring migration orientation according to geographic location across the Bogong moth range. Plotted using Azimuthal Equidistant map projection, observer centred on Mt Kosciuszko. **a.** AiSNP_01 (HiC_scaffold_9:16593335, $p = 6.87 \times 10^{-10}$). **b.** AiSNP_02 (HiC_scaffold_12:4434220, $p = 8.06 \times 10^{-8}$). **c.** AiSNP_03 (HiC_scaffold_15:11378716, $p = 1.73 \times 10^{-8}$). **d.** AiSNP_04 (HiC_scaffold_25:12095901, $p = 2.14 \times 10^{-7}$).

Table 6.3: Summary of loci which are significantly correlated with spring migratory direction.

ID/ Scaffold: Coordinate	Description
AiSNP_01/ HiC_scaffold_9: 16593335	$p = 6.87 \times 10^{-10}$. Located in the 3' exon of an uncharacterised gene, which is expressed in eye and brain tissue in the Bogong moth. The SNP maps to the third position of a tyrosine codon, and is a synonymous substitution (TAT → TAC).
AiSNP_02/ HiC_scaffold_12: 4434220	$p = 8.06 \times 10^{-8}$. Appears to be in a non-coding region, with the closest gene being approximately 10 kb away.
AiSNP_03/ HiC_scaffold_15: 11378716	$p = 1.73 \times 10^{-8}$. Located in an intron, towards the 5' end of the OUTD7B gene (OTU domain-containing protein 7B), which is involved in ubiquination, and ultimately immune regulation (Hu et al., 2016, 2013), DNA repair, and the EGFR/MAPK (Lei et al., 2019) and mTORC2 (Wang et al., 2017) pathways.
AiSNP_04/ HiC_scaffold_25: 12095901	$p = 2.14 \times 10^{-7}$. Located in the coding region of the NFXL1 gene (NF-X1-type zinc finger protein NFXL1). NFXL1 is thought to mediate DNA binding and function as an E3 ubiquitin ligase in plants (Müssig et al., 2010), and has been shown to be expressed in human embryonic stem cells, but down-regulated during differentiation to myelinated oligodendrocytes (Chaerkady et al., 2011). Variants in the human NFXL1 gene have been associated with risk of diagnosis with Specific Language Impairment during childhood (Villanueva et al., 2015). The SNP maps to the third position of a glutamic acid codon, and is a synonymous substitution (GAA → GAG).

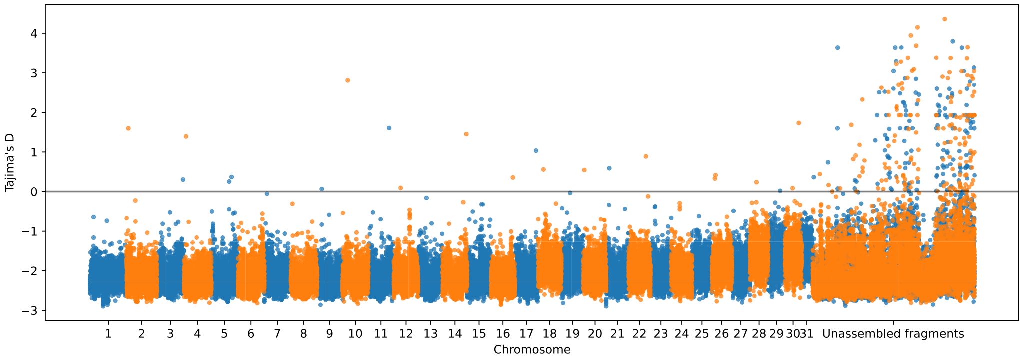


Figure 6.5: Genome-wide scan of Tajima's D (Tajima, 1989), calculated in non-overlapping 10 kb bins. Colouring convention follows that of Fig. 6.3.

6.3.5 Regions under selection

Tajima's D (Tajima, 1989) is a test statistic frequently used to assess the mode of selection acting on nucleotide sequences. It evaluates the null hypothesis that a sequence of DNA has evolved under neutral selection by comparing estimates of nucleotide diversity based on the number of polymorphic sites with estimates based on the allele frequencies of those sites. Values of Tajima's D which deviate negatively from 0 indicate an over-abundance of rare variants, with respect to the expectation under neutral selection, suggesting a recent selective sweep or population expansion. Values of Tajima's D which deviate from 0 in the positive direction indicate an over-abundance of common variants, with respect to the neutral expectation, suggesting the presence of balancing selection or a population contraction.

We calculated Tajima's D in 10 kb windows across the Bogong moth genome using VCF-kit 0.2.9 (https://github.com/andersenlab/VCF-kit). A marked depression of Tajima's D was observed across the genome (Fig. 6.5), implying an over-abundance of rare variants.

We identified genes which are likely to have recently been acted on by balancing selection or positive selection, by selecting genes which overlapped with the top or bottom 1% of the Tajima's D bins, respectively. A full list of these genes is presented in Appendix E.3. A total of 376 genes were identified in the top 1% of the Tajima's D bins (full list: Table E.1), and 256 genes were in the bottom 1% (full list: Table E.2). To determine if selection was acting on particular biological processes or molecular functions, we performed a gene ontology enrichment analysis on both groups of identified genes and found multiple significantly enriched terms in each ($p < 0.05$, Fisher's exact test, Bonferroni corrected). These enriched terms are presented in Table 6.4.

Table 6.4: Enriched gene ontology terms in genes co-located with the top (**a**) and bottom (**b**) 1% of Tajima's D values, calculated in 10 kb bins across the first 31 scaffolds in the Bogong moth genome. Terms shown were found to be significantly enriched ($p < 0.05$, Fisher's exact test, Bonferroni corrected). GO Root terms shown are Molecular Function (MF) and Biological Process (BP).

GO Term	Description	GO Root	p
a)	**Top 1% of Tajima's D values**		
GO:0051033	RNA transmembrane transporter activity	MF	1.14×10^{-5}
GO:0033227	dsRNA transport	BP	1.14×10^{-5}
b)	**Bottom 1% of Tajima's D values**		
GO:0003700	DNA-binding transcription factor activity	MF	1.52×10^{-7}
GO:0043565	sequence-specific DNA binding	MF	1.75×10^{-5}
GO:0006355	regulation of transcription, DNA-templated	BP	1.81×10^{-7}
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	BP	1.65×10^{-5}

6.4 Discussion

Our analyses show that the Bogong moth population contains a large amount of (mostly rare) variation, and that this variation does not correspond to the geographic distribution of its winter breeding grounds, nor its summer range. In fact, the variation appears to be entirely unstructured in the population. Therefore, we conclude Bogong moths form a single panmictic population, a result which agrees with previous attempts to characterise Bogong moth population structure using SNP array (Peter Kriesner, personal communication) and RNA-Seq data (Chapter 5). A similar result has recently been reported in the Bogong moth's diurnal Northern Hemisphere migratory counterpart, the monarch butterfly *Danaus plexippus* (Talla et al., 2020).

Panmixia of the Bogong moth population requires high levels of gene flow between the far reaches of their breeding grounds. This suggests that individual Bogong moths do not, in general, return to the specific region in the breeding grounds from which they originated once they have completed their high-altitude aestivation—at least lineages of moths do not consistently return to a specific region over multiple generations. That is, it is likely that Bogong moths have low migratory connectivity.

Despite the suitability of panmixia as a model of Bogong moth population structure, we were able to discover a small number of variants which were significantly associated with geographic location, and, by inference, migratory direction. Three of these variants occur within genes, however the functional consequences of the variants are not immediately obvious. Two of the variants cause synonymous substitutions and one is located in an intron, so putative function is likely to be conferred through some regulatory process, rather than a modification of a gene product *per se*. Nevertheless, further investigation into the functional significance of each of these variants is warranted.

We are left with a curious combination of conclusions. On the one hand, it seems likely that Bogong moth population and migratory connectivity is low, meaning the moths are not predestined to return to the specific region they hatched in order to breed. On the other hand, there are some some genetic variants associated with geographic location in the breeding grounds. An obvious question arises—how are genetic associations with geographic location established and/or maintained in a panmictic population? An enticing possible explanation is that these location-associated variants are subject to location-dependent selection pressure, which would reduce the capacity of gene flow to eliminate the location-dependent variation. Either a geographically-imprecise return migration of the Bogong moths to their breeding grounds or a degree of mixing and mating prior to commencing the return migration[1] could then facilitate enough gene flow to remove signs of population structure across the broader genome.

An imprecise return migration could be considered adaptive, as it would enable rapid re-colonisation of breeding areas rendered temporarily unfavourable by climatic events such as drought (Farrow and McDonald, 1987), or even untimely or excessive rainfall, which can affect noctuid pupal

---

[1] Observations made by Common (1954) at Mt. Gingera led him to conclude that mating does not occur during Bogong moth aestivation. However, in the early hours of the morning on 29th December, 2019, we (JRAW and EJW) observed a pair of Bogong moths copulating in an aestivation site on Mt. Kosciuszko. Whether this is normal or not—merely representing an exception to the usual post-return-migration mating of the moths—remains unclear.

survival (Murray and Zalucki, 1990a, 1990b; Sims, 2008). Indeed, rapid recolonisations by Bogong moths (and other noctuid moths) do occur (Farrow and McDonald, 1987). The offspring of these colonists would benefit from some degree of flexibility in their putative inherited migratory direction, as they migrate from a different starting area from that of their parents, who migrated the previous year.

As with any association study, the associations we have discovered in our genome-wide scan for location-associated genetic variants are simply correlations, and causality needs to be confirmed with reverse genetics experiments. Such experiments should include direct measurement of the phenotype of interest, namely, migratory direction. This could be achieved using a Frost-Mouritsen flight simulator, which has an established use for studying Bogong moth migration (Dreyer et al., 2021). This would enable the experiment to disentangle potentially confounded phenotypic and behavioural responses, such as timing of migration and migratory direction. Such confounding factors may complicate the present study, owing to the variation in collection time for samples from different areas (Table 6.1). However, for the purposes of our aims, they are not of great concern, as genetic correlations to migratory timing are interesting in their own right, and would still require reverse-genetic confirmation.

Reverse genetics experiments that measure a behavioural phenotype are necessarily laborious, and therefore only tractable for studying a small number of genes. Equipped with the results from this large-scale sequencing effort, we now know where to look for genes which are putatively involved in controlling long-distance navigation, opening the door to previously intractable experiments which could shed light on the fascinating phenomenon of directed animal migration.

The whole-genome scan for signals of selection yielded evidence of a marked over-abundance of rare variants, with respect to our expectation under the null hypothesis that the genome sequence evolved under neutral selection and fixed population size. This indicates that the Bogong moth population has likely recently recovered from a past genetic bottleneck, resulting in variation being dominated by mutations rather than genetic-drift-mediated partial fixation. It is known that Bogong moth population size can vary dramatically from year to year (Green et al., 2021), although it is thought that this has been underpinned by a slow downward trend over the last half-century (*ibid.*).

Interestingly, regions of the genome with high Tajima's D values appear to be preferentially co-located with genes involved in RNA transmembrane transporter activity (Table 6.4a). High values of Tajima's D are often interpreted as evidence of balancing selection, or recent population contraction. The latter seems unlikely because most of the genome has strongly negative Tajima's D values (Fig. 6.5). That leaves us with the conclusion that balancing selection is acting on these regions to increase their diversity in the population. RNA transport is important for regulation of gene expression, which is a fundamental process, so changes in how it operates could have far-reaching consequences for the organism. It is therefore not hard to imagine that this type of selection acting on these regions could have important ecological implications.

On the other hand, genomic regions with low Tajima's D values appear to be preferentially co-located with genes involved in sequence-specific transcription regulation, as well as homophilic cell adhesion (Table 6.4b). Low values of Tajima's D are often interpreted as evidence of a recent selective sweep (i.e. positive selection) or recent population expansion. Since we are looking only at regions which have the bottom 1% of Tajima's D values, we can be fairly confident that selective sweeps are good explanations for the low values near these genes, albeit in a background of population expansion. The enrichment of positive selection acting on sequence-specific transcription regulation is particularly interesting in the context of the Bogong moth, as these are the type of potential molecular events that can bring about important behavioural changes already seen in other organisms (Rittschof et al., 2014).

It is reasonable to assume that selection acts strongly on migration and migratory direction in the Bogong moth. Directed migration in lepidoptera is known to be remarkably fragile (Tenger-Trolander et al., 2019), and would probably disappear quickly without the continual action of selection. It is therefore plausible that migratory direction itself is the phenotype under location-dependent selection, leading to genotypic selection on the genomic regions and the location-dependent variants we identified.

Perhaps the most promising location-dependent variant is AiSNP_03, which is located in an intron towards the 5' end of the OUTD7B gene. OUTD7B is involved in immune regulation (Hu et al., 2016, 2013), and the EGFR/MAPK (Lei et al., 2019) and mTORC2 (Wang et al., 2017) pathways, which are central to regulatory networks in the cell, and act in a context-dependent manner. One could speculate that AiSNP_03 interferes with an intronic enhancer leading to changes in a regulatory module, with broader implications for gene expression.

A common theme has emerged from our analyses of 77 Bogong moth genomes: genetic regulation. Animal behavioural properties (such as navigational skills) are largely emergent and depend on selective events at molecular or physiological levels. It is therefore conceivable that subtle changes to regulation—mediated by modulation of important biochemical pathways such as EGFR/MAPK and mTORC2—could have profound implications for behavioural control. Indeed, regulatory networks must be extremely important for differentiating between the migration and aestivation phases in the Bogong moth life cycle, as deep differences in the transcriptional profiles of moths in these two behavioural states have been observed in sensory and brain tissues (Chapter 5).

Bogong moths' remarkable behaviour is a distributed property of many brain networks controlling sensory organs, muscles, metabolic flux, etc. that respond to both external and internal cues. For example, the encoding of either north/south or east/west directionality could be described as an inherited "value", selected through evolutionary and somatic processes (Friston et al., 1994). The Bogong moths' preference for cooler areas have likely strengthened certain neuronal connections and created this value (e.g. during spring, south is better than west for the Bogong moth, much like dark is better than light for many animals). When this value is realised, these connections are reinforced and dominate amongst numerous neuronal connections (Sporns et al., 2000). When external or internal conditions change—say, the temperature increases, or the moth's brain reaches a certain stage of maturation—these neurons fire together, driving the moth's desire to move towards cooler regions, which they achieve through integration of other navigationally-informative cues, (e.g. landmarks and the geomagnetic field, Dreyer et al., 2018; the night sky, Adden, 2020).

To implement such a value system, the Bogong moth must inherit the capacity to sense relevant thresholds in the cues which inform its migration (e.g. temperature, day length, celestial cues, geomagnetism, etc.). For example, there must be a set of thresholds that either allow or prevent the onset of migration, as a premature or delayed migration could have severe consequences. Similarly, the Bogong moth must have a way of measuring compass directions, including—but not limited to—detecting the geomagnetic field (Dreyer et al., 2018), possibly using a sensor similar to cryptochrome 4, which is thought to be the magnetoreceptor in night-migratory songbirds (Xu et al., 2021). Such capacities are established via a developmental program that has evolved some Bogong moth-specific features, but overall, cannot be hugely different from its close non-migratory—or non-directionally-migratory—relatives.

To become heritable, a behavioural novelty must be reflected in the genome. For example, the migratory direction-associated SNPs we discovered in this study could represent variations which provide a selective advantage for the migratory behaviour of the Bogong moth. SNPs in non-coding (intronic or intergenic) regions are of special interest because they are more likely to affect gene expression. We have discussed the intronic AiSNP_03 variant above, and another of the SNPs we discovered is also in a non-coding region (AiSNP_02; Table 6.3). The remaining two SNPs are in the coding region of genes expressed in eye and brain tissue, and one of which (AiSNP_04) is likely to function as a transcription factor. Of course, there are many more SNPs that may not show in statistical tests, but could still influence gene regulation.

An important mechanism for far-reaching regulatory control is epigenomic modification, mediated through the covalent bonding of a methyl group to a base (typically a cytosine). Even small genomic changes often involving epigenomic modifiers have been shown to have a massive effect on brain development and function (reviewed by O'Donnell and Meaney, 2020). Importantly, there are a number of ways that epigenomic modifications can be inherited. First, there appears to be a lot of heterogeneity in the Bogong moth genome (which is, by definition, heritable) on which selection could act at the epigenomic level. For instance, in the honeybee genome, there are over 220,000 SNPs that potentially could change the number of methylated sites, either by changing a CpG dinucleotide (the typical target of DNA methylation) to another (e.g. ApG), or by creating a new target (e.g. by changing a TpG to CpG) (Wedd et al., 2016). It is likely that there are even more such SNPs in the Bogong moth genome. Second, we also now understand how acquired (epigenetic) features could be transferred to the next generation, for example, by microRNAs in sperm (reviewed by Chen et al., 2016). Therefore, there are molecular mechanisms by which the entire gene-regulatory network topology can be modified and rewired to generate novel gene expression patterns, and these can likely be passed from one generation to another.

Thus, our analysis of genetic variation in the Bogong moth is yet another example of how epigenetic mechanisms, bound by genetic constraints, are prime drivers of brain plasticity arising from both developmental and experience-dependent events. Furthermore, our results expose a multitude of interesting and exciting avenues of further research into the molecular basis of insect migration, and help establish the Bogong moth as an illuminating emerging model, not only for the study of nocturnal migration and navigation, but also for the study of the fundamental biomolecular processes that contribute to complex animal behaviour.

6.5 Author contributions

JRAW, EJW, and RM conceived the project and designed the experiment. JRAW made the majority of sample collections, coordinated the remainder of the collections, performed the lab work, analysed the data, and wrote the draft version of the manuscript. All authors interpreted the results, contributed text, and edited the manuscript.

6.6 Acknowledgements

EJW and JRAW are grateful for funding from the European Research Council (Advanced Grant No. 741298 to EJW), and the Royal Physiographic Society of Lund (to JRAW). JRAW is thankful for the support of an Australian Government Research Training Program (RTP) Scholarship. EJW holds Scientific Permits for collection and experimental manipulations of Bogong moths in several alpine national parks and nature reserves (NSW:

Permit SL100806; Vic: Permit 10008966). The computations and data handling were enabled in part by resources in projects SNIC 2018/8-364 and SNIC 2021/23-74 provided to EJW and JRAW by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

We thank Ken Green, Dean Heinze, David Szakal, Adrien Lefèvre and Benjamin Mathews-Hunter for assistance with making collections in the field, Niccy Aitken and Justin Borevitz for laboratory resources and guidance, and Jochen Zeil, Robert Kucharski, Gavin Huttley and Andy Bachler for useful comments relating to the analyses.

Part III

Conclusion and appendices

Chapter 7

Conclusion

Long-distance migrations are some of the most impressive spectacles in the animal kingdom. The mere mention of animal migration evokes—in the minds of many—Attenborough-esque images of vast wildebeest herds on the Serengeti, innumerable red crabs on Christmas Island, newly hatched sea turtles struggling their way to the ocean, salmon leaping up rapids while narrowly avoiding the claws of a bear, seabirds landing on tiny rock islands to roost, and billions of Monarch butterflies clustering together like orange and black leaves in the densely-forested Mexican Central Highlands. But not all migratory animals are so ostentatious. Some, like the Australian Bogong moth, travel only in the cover of darkness, conceal their vibrant colours in a cloak of melanin (Stavenga et al., 2020), and hide away in only the most secluded mountainous caves.

Each year, billions of Bogong moths perform an incredibly impressive navigational feat. Their innate ability to find a specific set of caves and crevices—which for some is over 1000 km from where they hatched—has drawn increasing interest from neuroethologists, conservationists, artists, and politicians[1] alike. Indeed, open questions relating to the mechanisms used by the Bogong moth to achieve this feat are currently some of the most important in long-distance navigation research (Mouritsen, 2018).

This thesis has provided meaningful progress in our quest to answer these questions. In particular, it presents a number of interesting results, leading us to new insights into Bogong moth behaviour. These insights have arisen from the hard-fought integration and synthesis of substantial volumes of novel data, which were produced using an eclectic variety of new and existing methodologies. These data and methodologies lay clearly-defined paths for further research that will answer these open mechanistic questions in years to come.

7.1 What have we learned?

Although this thesis has yielded multiple different insights, a few are particularly worthy of being highlighted. For example, from high-throughput measurements of the behaviour of Bogong moths—presented in Part I—we have learned that the laboratory-based behavioural results of Dreyer et al. (2018) hold in the wild. Namely, Bogong moths fly in directions relative to the azimuth of prominent visual landmarks. This behaviour facilitates the controlled movement of Bogong moths up and down elevation gradients, and probably gives the moths an opportunity to calibrate their navigational machinery—taking in celestial and geomagnetic cues—while identifying tall

---

[1] In addition to being cultural icons, Bogong moths evidently wield political power, as the regularly stop by Parliament House in Canberra *en route* to the mountains. One can only assume this is to conduct important meetings with the Prime Minister.

mountains on the horizon, and thus potentially favourable aestivation sites. Once at such a mountain, the Bogong moth could then continue, using a strategy of methodical spiral ascent, to traverse the contours of the mountain while checking for suitable caves and crevices, using mechanisms still unknown to science (and possibly involving olfactory cues, as suggested by Warrant et al., 2016). Such a procedure may turn out to explain—at least in part–one of the most important open mechanistic questions in long-distance navigation research (Mouritsen, 2018): how does the pinpointing-the-goal phase work in a Bogong moth?

Another noteworthy result from this thesis comes from Part II. Namely, the Bogong moth population is panmictic, but contains location-discriminating variation at a restricted set of genomic loci. Although necessarily inconclusive (owing to the genome-wide-associating study-design used to obtain it), these results have far-reaching consequences for our understanding of Bogong moth navigation, and indeed, its conservation. In particular, the few loci which are associated with breeding ground location—and therefore migratory direction—are promising putative contributors to the emergent orientation behaviour of Bogong moths. Meanwhile, the overarching lack of genetic structure in the population suggests that Bogong moths probably exhibit a low degree of migratory connectivity, increasing their robustness to temporary occlusions in their breeding range, caused by drought,[2] fire,[3] flood,[4] or pesticide use (Green et al., 2021).

---

[2] e.g. the drought of 2017–2019, which affected large portions eastern Australia.

[3] e.g. the devastating Australian bushfire season of 2019–2020, which also affected large portions eastern Australia.

[4] e.g. the unprecedented floods currently occurring across Queensland and NSW (March 2022; although these particular floods may not considerably affect the Bogong moths, since this year's return migration has yet to begin in earnest).

7.2 Where to from here?

The research presented in this thesis begets a wide array of possible avenues of further research into the ecology, the emergent behaviour, and the physiological and molecular basis of the navigational abilities of Bogong moths. To conclude this thesis, I will briefly emphasise just two.

First, the Bogong moth should continue to be monitored in a long-term, and temporally- and spatially-dense way, possibly using a combination of the long-term approach used in Chapter 3 and the behaviourally-informative, high-frequency approach used in Chapter 4 (which may require a more robust and permanent camera infrastructure than the wildlife cameras used in those chapters). Such a program will provide one of the richest quantitative datasets of the behaviour and abundance of any insect species, allowing for robust measures of the impact of conservation efforts for the Bogong moth into the future, and possibly its use as a sort of "canary in the coal mine" for the health of the broader eastern Australian ecosystem. Furthermore, the resulting data will enable detailed modelling of the behaviour of Bogong moths in response to weather and climatic factors. This will facilitate the elucidation of further unknown aspects of the Bogong moth's migration strategies, and will improve predictions of the outcomes of conservation interventions and a changing climate on the size of the Bogong moth population.

Second, the results from our genome-wide association study of migratory direction indicate that the next step in the investigation of whether there are specific genetic drivers of the Bogong moth's heritable migratory direction can be taken. Namely, a reverse-genetic paradigm could be employed, using the gene-editing system described by Ran et al. (2013), in conjunction with the orientation-recording apparatus described by Dreyer et al. (2021), to test whether the loci identified in Chapter 6 have an influence on Bogong moth orientation behaviour. If successful, these experiments would represent a huge milestone in our understanding of the mechanistic basis of Bogong moth navigation, and of the interaction of genomes with complex emergent behaviour more generally.

And so ends this thesis.

Appendix A

Appendix to Chapter 2

A.1 Weighted Intersection over Minimum (IoM)

It is common for automatic image annotation procedures to produce multiple candidate annotations for a single object of interest (in our case, motion blurs of flying insects). It is therefore necessary to perform non-maximum suppression on the automatically generated candidate annotations (where only the best candidate annotation for each object is kept).

In order to perform non-maximum suppression on candidate annotations, we need a way of matching annotations which refer to the same object. This is typically done by defining some measure of similarity between two annotations, and then applying this measure to each pair of annotations within an image. Then, by setting an appropriate threshold, the program can decide which pairs of annotations require non-maximum suppression to be applied.

In the case of our method, it is common for the automatic annotation procedure to produce multiple annotations of different sizes for each motion blur. This is likely due to the fact that the motion blurs themselves can vary greatly in length and in number of wingbeats, which causes some level of confusion for the automatic annotation model. Therefore, we need to use a similarity measure which is invariant to the size of annotations, and produces a high similarity for annotations which are (roughly) contained within each other. This motivates our definition of similarity of candidate annotations for the purposes of non-maximum suppression. Namely, weighted intersection over minimum (IoM).

Suppose we have two sets $A = \{a_0, a_1, ..., a_n\}$ and $B = \{b_0, b_1, ..., b_m\}$ with corresponding weights $X = \{x(a_0), x(a_1), ..., x(a_n)\}$ and $Y = \{y(b_0), y(b_1), ..., y(b_m)\}$. We define the intersection over minimum of the two sets as

$$IoM(A, B) = \frac{\sum_{z \in A \cap B} \min(x(z), y(z))}{\min(\sum X, \sum Y)}$$

In camfi, we apply this definition by setting the weights X and Y as the segmentation mask values from two candidate annotations respectively. In this case, $A = B$ are the coordinates of every pixel in the image.

A.2 Bounding-box Intersection over Union (IoU)

To validate the quality of an automatic annotation system, we would like to compare the annotations produced by the system to annotations produced by a human. To do this, we need to have a way of matching pairs of annotations. This can be done by measuring the similarity between two annotations, and if they are similar enough, matching them.

The bounding-box intersection over union (IoU) is a commonly used similarity measure for object detection on images. It is defined as per it's name. That is, we find the bounding box of two annotations, then calculate the ratio of the intersection of the two boxes with the union of the two boxes.

The mean bounding-box intersection over union $\overline{IoU}$ is the arithmetic mean of all IoU values across the automatic annotations which were successfully matched to a manual ground-truth annotation. Since matches were made for annotations with $IoU > 0.5$, it must also hold that $\overline{IoU} > 0.5$.

A.3 Hausdorff distance

Our method for measuring wingbeat frequencies depends on accurate annotations of flying insect motion blurs, so it is important to know the accuracy of the annotations produced by our method for automatic annotation.

Suppose we have an automatically generated polyline annotation, and a corresponding polyline annotation made by a human which we would like to validate the automatic annotation against. We would like to know how accurately the automatic annotation recreates the human annotation. We proceed by calculating the Hausdorff distance between the two annotations. First, we define two sets A and B which contain all the points on the respective polyline from each of the two annotations.

The Hausdorff distance $d_H(A, B)$ is defined as

$$d_H(A, B) = \max\left\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\right\},$$

where sup is the supremum, inf is the infimum, and $d(a, b)$ is the Euclidean distance between points a and b. In other words, the Hausdorff distance is the maximum distance between a point in one of the sets, to the closest point in the other. For the purpose of validating automatic annotations, we see that smaller Hausdorff distances between the automatic and manual annotations are better than larger ones.

The mean Hausdorff distance $\overline{d_H}$ is the arithmetic mean of all values of d_H across the automatic polyline annotations which were successfully matched to polyline annotations in the manual ground-truth dataset.

A.4 Signed Length Difference

Another way to assess the accuracy of the automatic polyline annotations against the manually produced annotations is signed length difference ΔL. This is motivated by the fact that our method for calculating wingbeat frequency is fairly sensitive to the length of the polyline annotation. Suppose we have an automatically generated polyline annotation with length L_A and a corresponding ground-truth manual annotation with length L_G. Then the signed length difference is defined as $\Delta L = L_A - L_G$. The closer the signed length difference is to zero, the better.

The mean signed length difference $\overline{\Delta L}$ is the arithmetic mean of all values of ΔL across the automatic polyline annotations which were successfully matched to polyline annotations in the manual ground-truth dataset. The standard deviation of signed length difference $\sigma_{\Delta L}$ is the standard deviation of these values.

A.5 Precision-Recall curve

With regard to object detection, precision is the proportion of detections which correspond to annotations present in the ground-truth dataset. Recall is the proportion of objects in the ground-truth dataset which are detected by the automatic annotation system. In our case, the ground-truth dataset is the set of manual annotations. We match automatic annotations with ground-truth annotations if they have an IoU greater than 0.5.

Each candidate annotation is given a confidence score between 0.0 and 1.0 by the annotation model. This score can be used to filter the candidate annotations (e.g. by removing all annotations with a score less than 0.9). By varying the score threshold, we obtain different precision and recall values for the system.

A precision-recall curve is the curve drawn on a plot of precision vs. recall by varying the score threshold. The closer the curve goes towards the point (1, 1), the better.

A.6 Average precision

The average precision AP_{50} is calculated from the precision-recall curve. It is simply the average (arithmetic mean) of the precision values at the following recall values: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.

A.7 Measurement of rolling shutter line rate

For the purposes of measuring the wingbeat frequency of the moths in the images captured by the wildlife cameras, it is important to know the line rate of the cameras' rolling shutters. This was measured by mounting one of the cameras so that its lens pointed at a rotating white line (in this case, a strip of paper taped to a cardboard tube attached to the blades of a small electric fan), ensuring that the centre of the white line and its centre of rotation were coincident. The apparatus for measuring the rolling shutter line rate of the cameras is shown in Fig. A.1.

The exact rotational velocity of the line was measured by synchronising a strobe light from a smart phone application (Strobily, 2019) to the period of rotation of the line. Synchronisation is achieved when the line appears stationary under the strobe.

A photograph was taken using the camera and the corners of the motion blur traced by the rotation line were marked using the free and open-source VGG Image Annotator (VIA) (Dutta and Zisserman, 2019) (VIA is a simple and standalone manual annotation software for image, audio and video, and is available from https://www.robots.ox.ac.uk/~vgg/software/via/). The marked corners correspond to the positions where the exposure of the rotating line began and ended (see Fig. A.1). Since we know the rotational velocity of the line, we can then calculate the rolling shutter line rate from the coordinates of these corners. Namely,

$$R = \frac{V_a\,(r_B - r_A)}{\angle ACB - \pi} \quad , \tag{A.1}$$

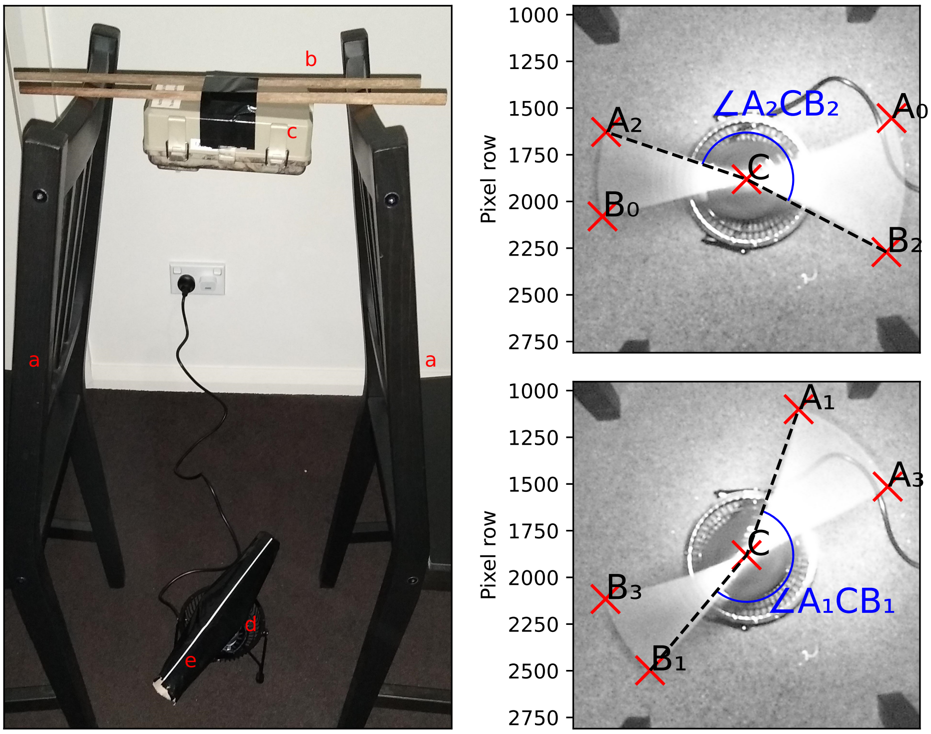


Figure A.1: Measurement of rolling shutter line rate of the wildlife cameras used in this study. **Left:** Apparatus for measuring rolling shutter line rate. **a.** Camera supports (chairs). **b.** Camera mount (long chopsticks with duct tape). **c.** Camera to be measured. **d.** Rotor motor (desk fan with front cover removed). **e.** Rotor (cardboard tube) with white reference line (paper masked with black duct tape). **Right:** Two rolling shutter line rate measurement images, with reference annotations marked, and angles $\angle A_1CB_1$ and $\angle A_2CB_2$ shown. Using the coordinates of these reference annotations, and if the rotational velocity of the white reference line is known, the rolling shutter line rate of the camera can be calculated using Eq. A.1.

where R is the rolling shutter line rate, V_a is the rotational velocity of the white line (in rad s^{-1}), A and B are the coordinates of two opposite corners of the motion blur traced by the rotating line (with A being the corner with the lower pixel-row index), C is the centre of rotation of the line, and r_A and r_B are the pixel-row indices of A and B respectively. Following the convention for digital images, the first pixel-row is at the top of the image (see Fig. A.1).

Rolling shutter line rate was calculated for both the start of exposure (Fig. A.1: A_0 and B_0, A_1 and B_1) and end of exposure (Fig. A.1: A_2 and B_2, A_3 and B_3), for two separate images. These four calculated values of R were very similar, so it was assumed that this particular model of camera has one constant rolling shutter line rate. This rate was taken to be the average of the two measured values from the second image, or 9.05×10^4 lines s^{-1}, which was used for all subsequent analyses, since only one model of camera was used in this study. If a different model is used, we would recommend repeating this measurement.

Appendix B

Appendix to Chapter 3

B.1 Automatic annotation evaluation

Automatic annotation performance was evaluated using a test set of 200 images, as well as the full set of 33780 manually annotated images. Evaluation metrics for both sets are presented in Table B.1. Each metric was similar across both image sets, indicating that the annotation model has not suffered from over-fitting. This is also supported by the contour plots of prediction score vs. IoU, polyline Hausdorff distance, and polyline length difference (Fig. B.1b,c,d, respectively). These plots show similar performance on both the full image set (33780 images) and the test set (200 images). Furthermore, they show that prediction scores for matched annotations (automatic annotations which were successfully matched to annotations in the manual ground-truth dataset) tended to be quite high, as did the IoU of those annotations, while both polyline Hausdorff distance d_H and polyline length difference ΔL clustered relatively close to zero. The precision-recall curves of the automatic annotator (Fig. B.1e) show similar performance between the image sets, and show a drop in precision for recall values above 0.4, possibly due to poorer performance for images which were taken in "day mode" (without infra-red flash). Training took 17414 iterations and completed in less than 2 h (Fig. B.1a) on a machine with two 8-core Intel Xeon E5-2660 CPUs running at 2.2GHz and a Nvidia T4 GPU.

Table B.1: Automatic annotation performance metrics for 2019–2021 study when tested against the full manually-annotated image set (33780 images), and the test set (200 images). Performance metrics calculated are average precision AP_{50}, mean bounding-box intersection over union $\overline{IoU}$, mean Hausdorff distance of polyline annotations $\overline{d_H}$, mean signed length difference of polyline annotations $\overline{\Delta L}$, and the standard deviation of signed length difference of polyline annotations $\sigma_{\Delta L}$. Definitions of these metrics follow those of Wallace et al. (2021). $^{\dagger}AP_{50}$ was calculated on the set of images with at least one manual annotation, rather than the full set of 33780 images.

Image set	AP_{50}	$\overline{IoU}$	$\overline{d_H}$	$\overline{\Delta L}$	$\sigma_{\Delta L}$
Full set	$0.426^{\dagger}$	0.835	25.9	-2.93	39.3
Test set	0.420	0.844	21.5	-0.80	28.5

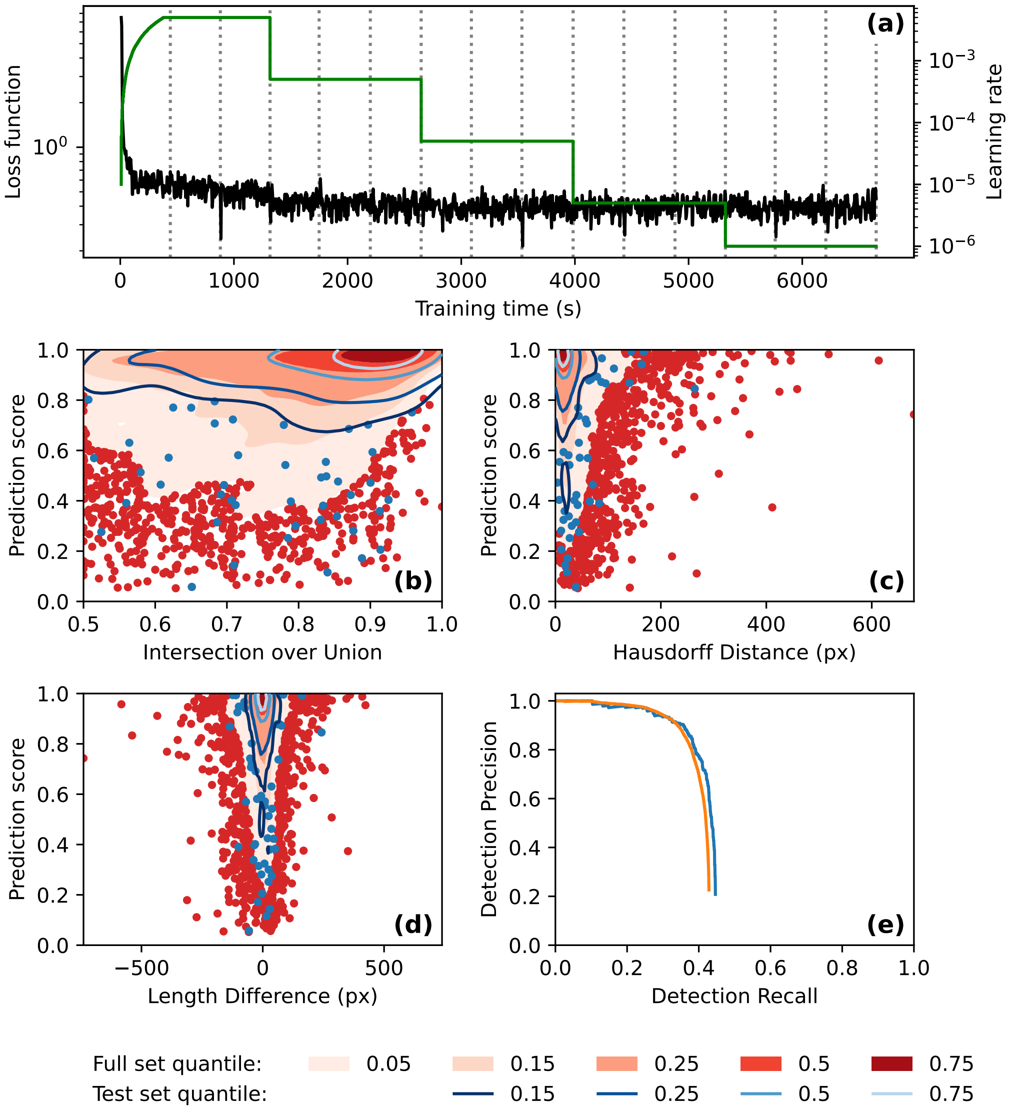


Figure B.1: Automatic annotation evaluation plots for 2019–2021 study. **(a)** Automatic annotation model training learning rate schedule (*green*) and loss function (*black*) over the course of training. Epochs (complete training data traversal) are shown with dotted vertical lines. **(b)**-**(d)** Gaussian kernel density estimate contour plots of prediction score vs. **(b)** bounding box intersection over union, **(c)** polyline Hausdorff distance, and **(d)** polyline length difference, for both image sets. Contours are coloured according to density quantile (key at bottom of figure). “Full set” refers to the set of 33780 images which were manually annotated. “Test set” refers to the set of 200 images with at least one annotation which were not used during model training. In each plot, data which lie outside of the lowest density quantile contour are displayed as points. **(e)** Motion blur detection precision-recall curve, generated by varying prediction score threshold. The precision-recall curve for the test set (200 images) is shown in *blue*, and the precision-recall curve for the set of 4223 images which had at least one manual annotation is shown in *orange*.

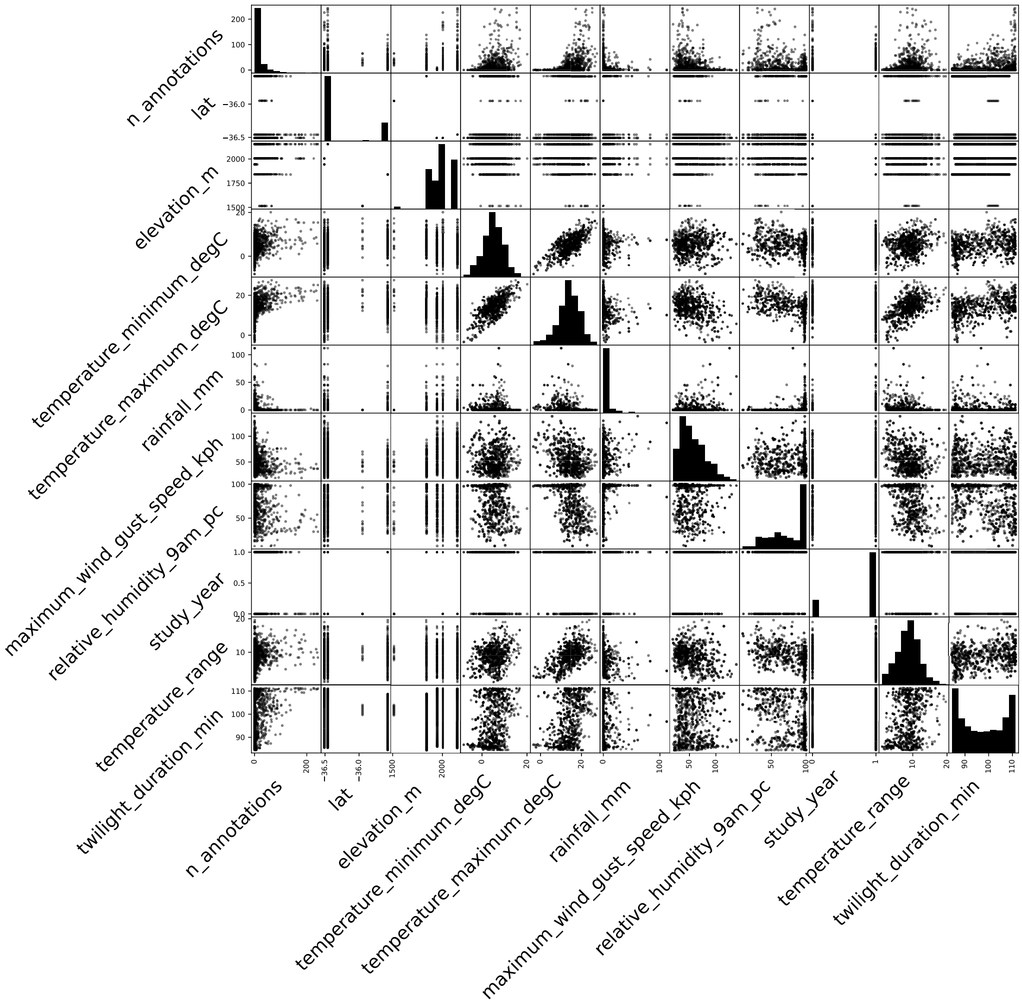


Figure B.2: Scatter matrix of Bogong evening twilight flight covariates.

B.2 Bogong evening twilight flight covariates

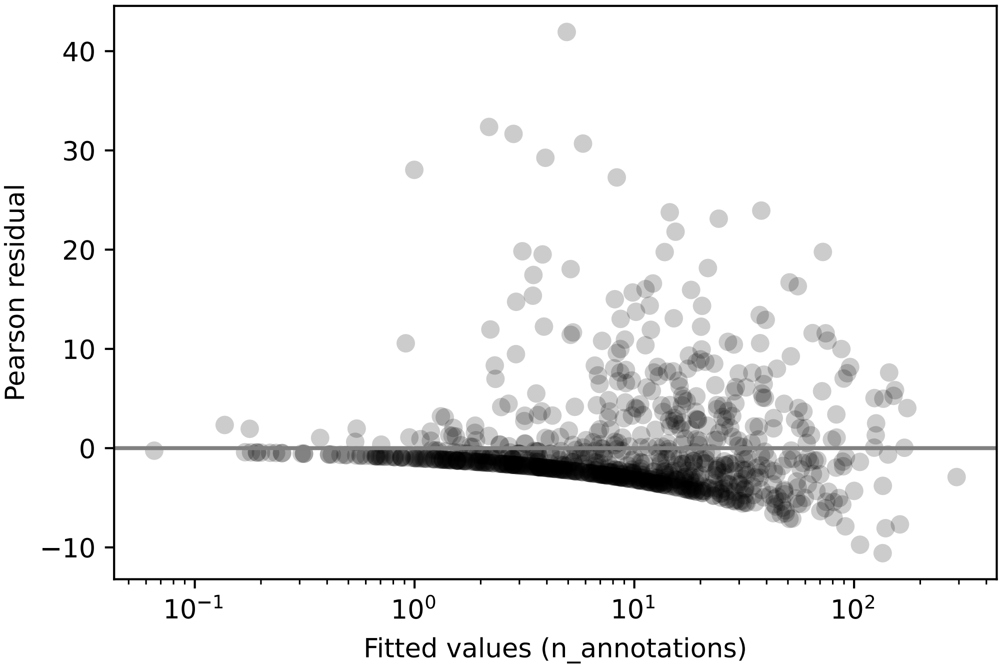


Figure B.3: Pearson residuals versus predicted evening twilight detection count for Poisson GLM of detections against (in order of effect size); elevation, maximum daily temperature, day length, maximum wind speed, study year, temperature range, 9 am relative humidity, latitude, minimum temperature, and rainfall.

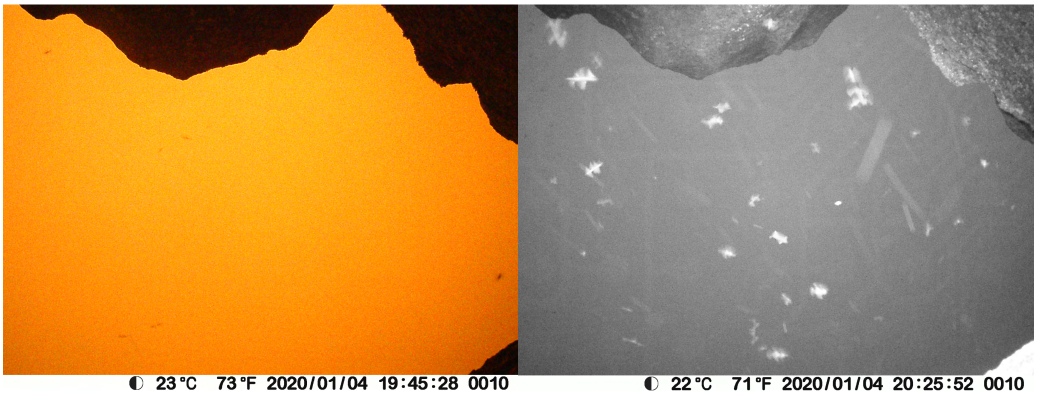


Figure B.4: Bogong moths flying during bushfire outside aestivation cave near the top of Ken Green Bogong on 4th January 2020. **Left:** Photograph taken by camera, shortly before switching to "night mode". The air is thick with smoke, leading to the orange colour. Dark specks in the air are likely Bogong moths. **Right:** Photograph taken by the same camera, once it had switched to "night mode", with infra-red flash. Flying Bogong moths are clearly visible.

B.3 Bogong moths flying during bushfire

Appendix C

Appendix to Chapter 4

Table C.1: Circular distribution models and corresponding output parameters selected using Akaike's information criterion (AIC), computed using the CircMLE R package (Fitak and Johnsen, 2017) on flying insect detections at the respective camera locations (Loc.; for brevity, "kosci_" prefixes are removed from each location name). Model selection was performed on the models defined by Schnute and Groot (1992). Models appearing in table: 2B = "symmetric modified unimodal", 5A = "homogeneous bimodal", 5B = "bimodal". Models are mixtures of von Mises distributions with two components i ($i = 1, 2$). φ_i denotes the mean direction of component i (in radians), κ_i the von Mises concentration parameter of component i, and λ the proportion assigned to the first component. θ is the azimuth of the summit of Mt. Kosciuszko (the nearest and highest peak) from the respective location. †Parameter fixed by model ($\lambda = 0.5$, $\kappa_2 = 0$). ‡Concentration parameters are assumed equal by model ($\kappa_1 = \kappa_2$).

Loc.	Model	φ_1	κ_1	λ	φ_2	κ_2	θ
north0	5B	5.782	16.056	0.365	3.286	0.584	1.993
north1	5B	3.941	6.111	0.302	0.234	1.103	1.972
north2	5B	6.001	2.079	0.490	3.965	3.589	2.037
north3	5B	5.960	0.628	0.487	3.592	13.697	2.303
north4	5B	5.929	8.383	0.250	1.342	1.166	1.493
south0	2B	2.019	46.502	$0.5^\dagger$	-	$0^\dagger$	6.148
south1	5B	1.859	2.235	0.547	4.457	5.203	6.189
south2	5B	1.870	9.389	0.320	4.567	0.642	6.255
south3	5B	2.826	31.053	0.265	5.060	1.037	0.045
south4	5A	2.356	$1.920^\ddagger$	0.495	5.263	$1.920^\ddagger$	0.292

C.1 Model selection tables

Table C.2: Model selection table for track directions at each camera location. †Parameter fixed by model. ‡Parameter depends on another parameter in model (i.e. $\varphi_2 = \varphi_1 + \pi(\bmod 2\pi)$, or $\kappa_1 = \kappa_2$). Models for each location are sorted by the model selection criterion, ΔAIC. All other parameters follow the conventions of Table C.1.

Loc.	Model	φ_1	κ_1	λ	φ_2	κ_2	ΔAIC
south0	2B	2.019	46.502	$0.5^\dagger$	-	$0^\dagger$	0
	4B	1.907	12.105	0.750	$5.049^\ddagger$	0.263	60.081
	2C	2.038	8.767	0.749	-	$0^\dagger$	80.870
	5B	2.067	7.325	0.749	5.682	1.415	163.506
	4A	2.074	9.009	0.749	$5.215^\ddagger$	$9.009^\ddagger$	262.845
	5A	2.129	8.085	0.750	5.567	$8.085^\ddagger$	283.271
	3B	5.182	0.001	$0.5^\dagger$	$2.040^\ddagger$	8.993	378.630
	2A	1.996	3.157	$1^\dagger$	-	-	674.318
	3A	2.044	9.742	$0.5^\dagger$	$5.185^\ddagger$	$9.742^\ddagger$	734.091
	1	-	$0^\dagger$	$1^\dagger$	-	-	2026.713
south1	5B	1.859	2.235	0.547	4.457	5.203	0
	5A	1.839	3.226	0.500	4.449	$3.226^\ddagger$	4.818
	4B	1.861	12.173	0.353	$5.002^\ddagger$	0.983	66.794
	3B	4.614	4.286	$0.5^\dagger$	$1.473^\ddagger$	1.961	71.716
	3A	4.713	2.740	$0.5^\dagger$	$1.572^\ddagger$	$2.740^\ddagger$	74.366
	4A	4.713	2.740	0.501	$1.572^\ddagger$	$2.740^\ddagger$	76.365
	2C	1.913	13.357	0.250	-	$0^\dagger$	131.445
	2A	3.185	0.445	$1^\dagger$	-	-	200.587
	2B	3.215	0.592	$0.5^\dagger$	-	$0^\dagger$	214.103
	1	-	$0^\dagger$	$1^\dagger$	-	-	237.114
south2	5B	1.870	9.389	0.320	4.567	0.642	0
	5A	2.022	2.132	0.578	4.924	$2.132^\ddagger$	1.406
	3B	5.061	1.357	$0.5^\dagger$	$1.920^\ddagger$	3.031	2.179
	4B	5.003	0.573	0.681	$1.862^\ddagger$	9.228	2.307
	4A	1.939	2.084	0.581	$5.081^\ddagger$	$2.084^\ddagger$	3.195
	3A	1.917	2.086	$0.5^\dagger$	$5.058^\ddagger$	$2.086^\ddagger$	4.896
	2C	1.919	7.364	0.250	-	$0^\dagger$	13.050
	2B	2.215	0.874	$0.5^\dagger$	-	$0^\dagger$	35.526
	2A	2.532	0.294	$1^\dagger$	-	-	37.602
	1	-	$0^\dagger$	$1^\dagger$	-	-	41.429
south3	5B	5.060	1.037	0.735	2.826	31.05	0
	5A	2.802	2.877	0.468	5.262	$2.877^\ddagger$	5.122
	4B	5.947	0.561	0.749	$2.805^\ddagger$	32.11	20.765
	3A	2.441	2.259	$0.5^\dagger$	$5.582^\ddagger$	$2.259^\ddagger$	27.993

Loc.	Model	φ_1	κ_1	λ	φ_2	κ_2	ΔAIC
	4A	2.429	2.259	0.464	$5.571^{\ddagger}$	$2.259^{\ddagger}$	29.531
	3B	2.390	1.869	$0.5^{\dagger}$	$5.532^{\ddagger}$	2.657	29.593
	2A	4.201	0.564	$1^{\dagger}$	-	-	38.757
	2C	2.866	8.024	0.250	-	$0^{\dagger}$	38.851
	2B	4.368	0.827	$0.5^{\dagger}$	-	$0^{\dagger}$	43.539
	1	-	$0^{\dagger}$	$1^{\dagger}$	-	-	51.566
south4	5A	2.356	1.920	0.495	5.263	$1.920^{\ddagger}$	0
	3A	5.375	1.876	$0.5^{\dagger}$	$2.234^{\ddagger}$	$1.876^{\ddagger}$	0.291
	5B	2.354	1.616	0.532	5.261	2.339	1.139
	3B	5.362	2.096	$0.5^{\dagger}$	$2.220^{\ddagger}$	1.713	1.755
	4A	5.375	1.875	0.506	$2.233^{\ddagger}$	$1.875^{\ddagger}$	2.267
	4B	5.350	2.310	0.466	$2.208^{\ddagger}$	1.560	3.379
	1	-	$0^{\dagger}$	$1^{\dagger}$	-	-	29.814
	2A	4.022	0.164	$1^{\dagger}$	-	-	30.603
	2C	5.054	2.929	0.251	-	$0^{\dagger}$	30.681
	2B	4.180	0.280	$0.5^{\dagger}$	-	$0^{\dagger}$	31.121
north0	5B	5.782	16.056	0.365	3.286	0.584	0
	4B	2.641	0.432	0.643	$5.783^{\ddagger}$	16.685	32.498
	2C	5.770	20.846	0.300	-	$0^{\dagger}$	62.367
	3B	2.655	0.653	$0.5^{\dagger}$	$5.796^{\ddagger}$	12.975	78.820
	2B	5.773	15.838	$0.5^{\dagger}$	-	$0^{\dagger}$	160.635
	5A	5.873	1.994	0.597	3.228	$1.994^{\ddagger}$	212.473
	4A	2.840	1.781	0.391	$5.982^{\ddagger}$	$1.781^{\ddagger}$	278.382
	3A	2.873	1.860	$0.5^{\dagger}$	$6.014^{\ddagger}$	$1.860^{\ddagger}$	306.605
	2A	5.312	0.487	$1^{\dagger}$	-	-	348.669
	1	-	$0^{\dagger}$	$1^{\dagger}$	-	-	452.970
north1	5B	3.941	6.111	0.302	0.234	1.103	0
	5A	4.010	1.881	0.446	0.345	$1.881^{\ddagger}$	208.187
	4B	3.850	6.079	0.288	$0.709^{\ddagger}$	0.885	298.063
	4A	3.711	1.693	0.443	$0.570^{\ddagger}$	$1.693^{\ddagger}$	484.484
	3A	3.746	1.665	$0.5^{\dagger}$	$0.605^{\ddagger}$	$1.665^{\ddagger}$	510.439
	3B	3.753	1.740	$0.5^{\dagger}$	$0.612^{\ddagger}$	1.636	511.982
	2A	5.555	0.366	$1^{\dagger}$	-	-	672.127
	2C	5.562	0.436	0.749	-	$0^{\dagger}$	691.654
	2B	5.675	0.602	$0.5^{\dagger}$	-	$0^{\dagger}$	715.270
	1	-	$0^{\dagger}$	$1^{\dagger}$	-	-	889.151
north2	5B	6.001	2.079	0.490	3.965	3.589	0
	5A	6.052	2.763	0.439	3.989	$2.763^{\ddagger}$	6.493
	2A	4.810	0.914	$1^{\dagger}$	-	-	504.94
	2C	4.876	1.080	0.749	-	$0^{\dagger}$	602.203
	4B	1.482	0.456	0.250	$4.624^{\ddagger}$	1.200	703.936

Loc.	Model	φ_1	κ_1	λ	φ_2	κ_2	ΔAIC
	2B	4.640	1.451	$0.5^\dagger$	-	$0^\dagger$	730.418
	3B	1.301	0.001	$0.5^\dagger$	$4.443^\ddagger$	2.001	774.017
	4A	1.403	1.045	0.250	$4.545^\ddagger$	$1.045^\ddagger$	886.929
	3A	0.432	1.491	$0.5^\dagger$	$3.574^\ddagger$	$1.491^\ddagger$	1218.499
	1	-	$0^\dagger$	$1^\dagger$	-	-	1403.655
north3	5B	5.960	0.628	0.487	3.592	13.697	0
	3B	3.599	15.007	$0.5^\dagger$	$0.458^\ddagger$	0.362	69.521
	4B	0.459	0.347	0.506	$3.600^\ddagger$	15.204	71.375
	2C	3.612	18.224	0.447	-	$0^\dagger$	95.057
	2B	3.610	16.834	$0.5^\dagger$	-	$0^\dagger$	105.805
	5A	6.001	3.072	0.313	3.487	$3.072^\ddagger$	361.557
	4A	0.224	2.508	0.304	$3.366^\ddagger$	$2.508^\ddagger$	668.046
	3A	0.212	2.633	$0.5^\dagger$	$3.354^\ddagger$	$2.633^\ddagger$	907.343
	2A	3.847	0.881	$1^\dagger$	-	-	926.746
	1	-	$0^\dagger$	$1^\dagger$	-	-	1534.394
north4	5B	5.929	8.383	0.250	1.342	1.166	0
	5A	1.685	1.743	0.487	6.231	$1.743^\ddagger$	60.988
	2A	0.749	0.931	$1^\dagger$	-	-	93.429
	2C	0.691	1.222	0.750	-	$0^\dagger$	114.046
	2B	0.683	1.599	$0.5^\dagger$	-	$0^\dagger$	150.333
	3B	3.800	0.001	$0.5^\dagger$	$0.659^\ddagger$	1.432	154.245
	4B	0.643	1.012	0.750	$3.784^\ddagger$	0.661	170.505
	4A	0.680	1.154	0.749	$3.821^\ddagger$	$1.154^\ddagger$	196.787
	3A	2.839	0.731	$0.5^\dagger$	$5.981^\ddagger$	$0.731^\ddagger$	377.408
	1	-	$0^\dagger$	$1^\dagger$	-	-	380.019

Table C.3: Model selection table for track directions relative to the azimuth of the summit of Mt. Kosciuszko. Follows conventions of Table C.2.

Model	φ_1	κ_1	λ	φ_2	κ_2	ΔAIC
5B	2.055	0.741	0.654	4.480	4.584	0
5A	4.463	1.650	0.561	1.899	$1.650^\ddagger$	385.854
4B	4.466	5.483	0.301	$1.325^\ddagger$	0.393	693.199
2C	4.369	7.320	0.251	-	$0^\dagger$	897.322
3B	4.549	2.412	$0.5^\dagger$	$1.408^\ddagger$	0.987	1012.238
4A	4.607	1.429	0.582	$1.465^\ddagger$	1.429	1238.613
2A	3.645	0.410	$1^\dagger$	-	-	1368.769
3A	1.537	1.483	$0.5^\dagger$	4.679	1.483	1387.302
2B	3.851	0.802	$0.5^\dagger$	-	$0^\dagger$	1442.646
1	-	$0^\dagger$	$1^\dagger$	-	-	2274.647

C.2 Luminance recordings

Table C.4: Luminance recordings from various locations inside and outside a Bogong moth aestivation cave, in February, 2021. Recordings were taken using a digital photometer (Hagner, model ERP-105).

Location	Time	Reading	Multiplier
Inside cave entrance	20:22 18/02	100	0.01
(recorded off standard)	20:35 18/02	78	0.01
	20:56 18/02	14	0.01
Outside cave,	20:22 18/02	442	10
(recorded off standard)	20:35 18/02	250	1
	20:56 18/02	490	0.01
Main front wall inside cave	20:30 18/02	17	0.01
(no moths)	20:33 18/02	41	0.01
	06:51 19/02	461	0.01
Cave, far right	20:30 18/02	6	0.01
(opposite moths)	06:51 19/02	22	0.01
Cave, right, left of rock	20:31 18/02	2	0.01
(opposite moths)	06:52 19/02	14	0.01
Cave, back, shaded	20:32 18/02	1	0.01
(no moths)	06:52 19/02	34	0.01
Cave, far left (no moths)	20:32 18/02	12	0.01
Cave, left (opposite moths)	06:53 19/02	9	0.01

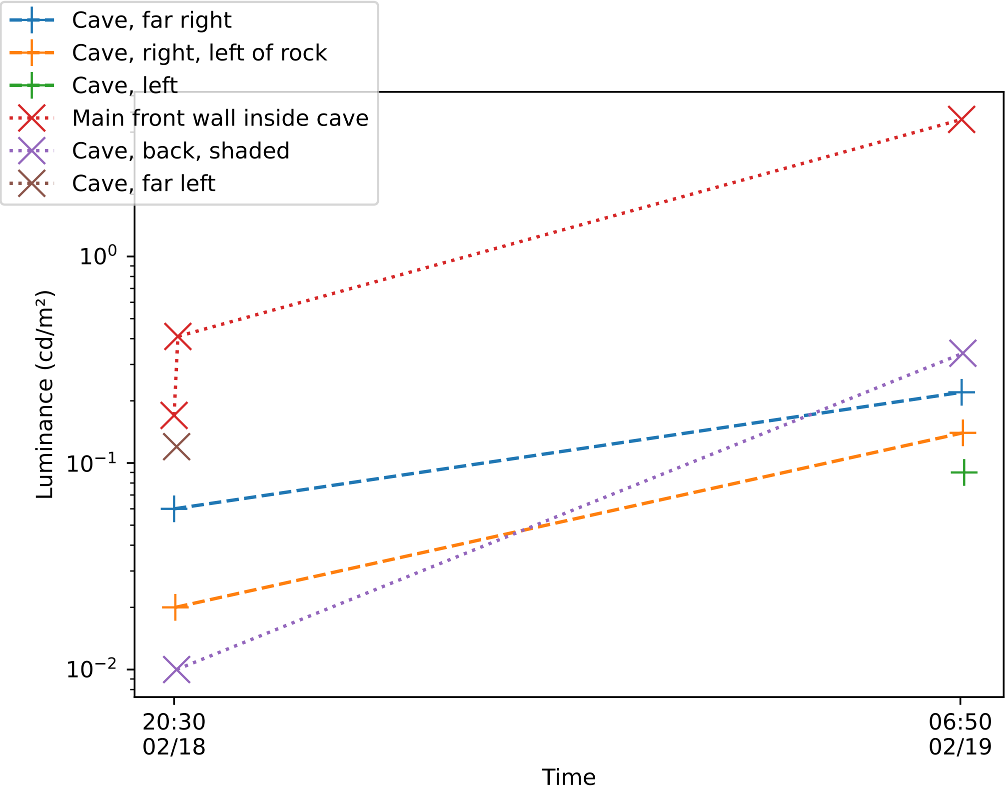


Figure C.1: Occupied positions within a Bogong moth aestivation cave on Mt. Kosciuszko (“+” markers, connected with *dashed lines*) tend to be darker than unoccupied positions (“×” markers, connected with *dotted lines*), particularly during the day (right side of plot), as measured by digital photometer (Hagner, model ERP-105).

C.3 Flight track orientations

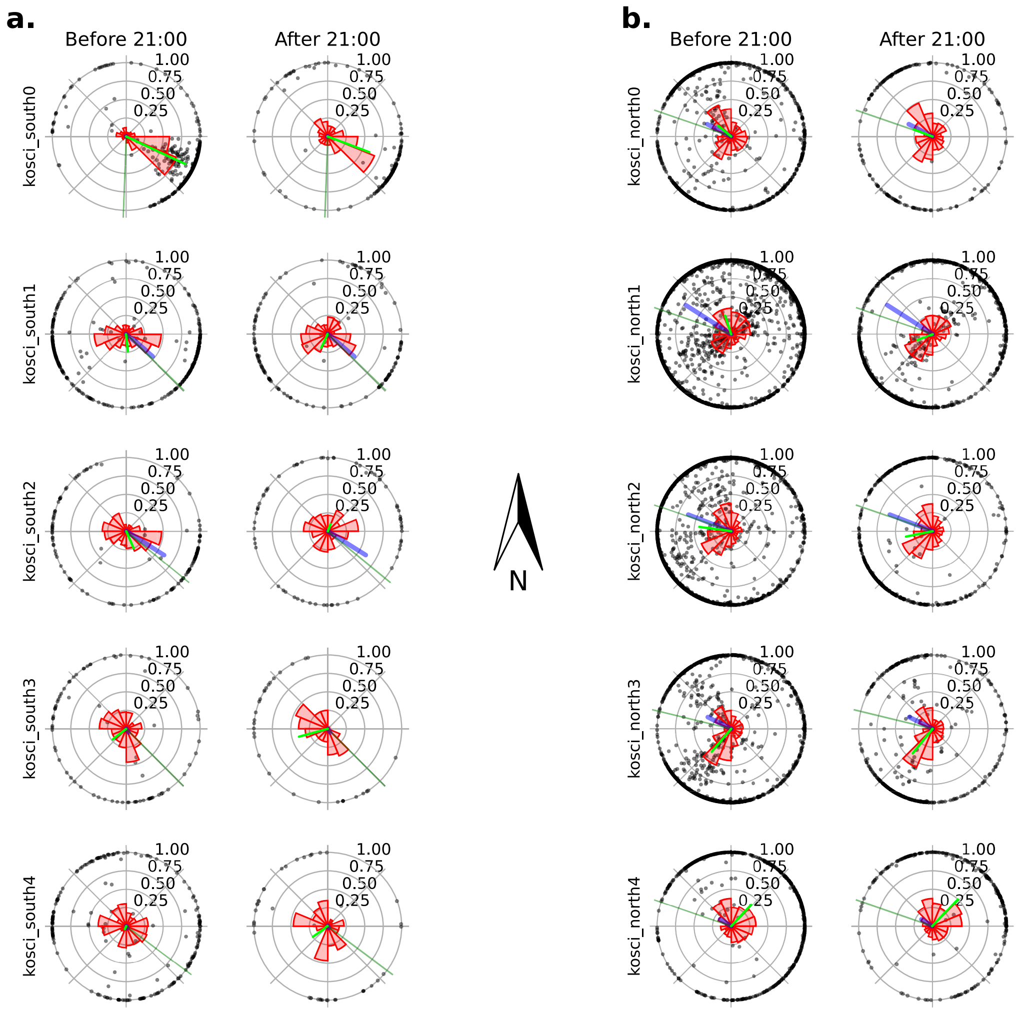


Figure C.2: Trajectories of detected insects during nautical twilight (before 21:00 AEDT) and after for both transects. Columns indicate time period and rows indicate location. **a.** `kosci_south` transect. **b.** `kosci_north` transect. *Black dots*: Track (direction of displacement) of detected insect trajectories. Radius indicates the straightness of the trajectory, calculated as distance travelled divided by displacement (in pixel units). *Red bars*: Circular histogram of detected insect trajectories. The bars are equiareal (area—not height—indicates proportion of detections contained within each bin). *Lime green line*: Mean track (direction of displacement) of detected insect trajectories. Radius indicates circular mean vector length (with values closer to one indicating more concentrated tracks). *Blue line*: Fall line of the slope at the position of the camera. The direction indicates the direction of maximum gradient (perpendicular to topographic lines), and the radius indicates the gradient itself. *Dark green line*: Indicates the bearing of the base of the camera.

Appendix D

Appendix to Chapter 5

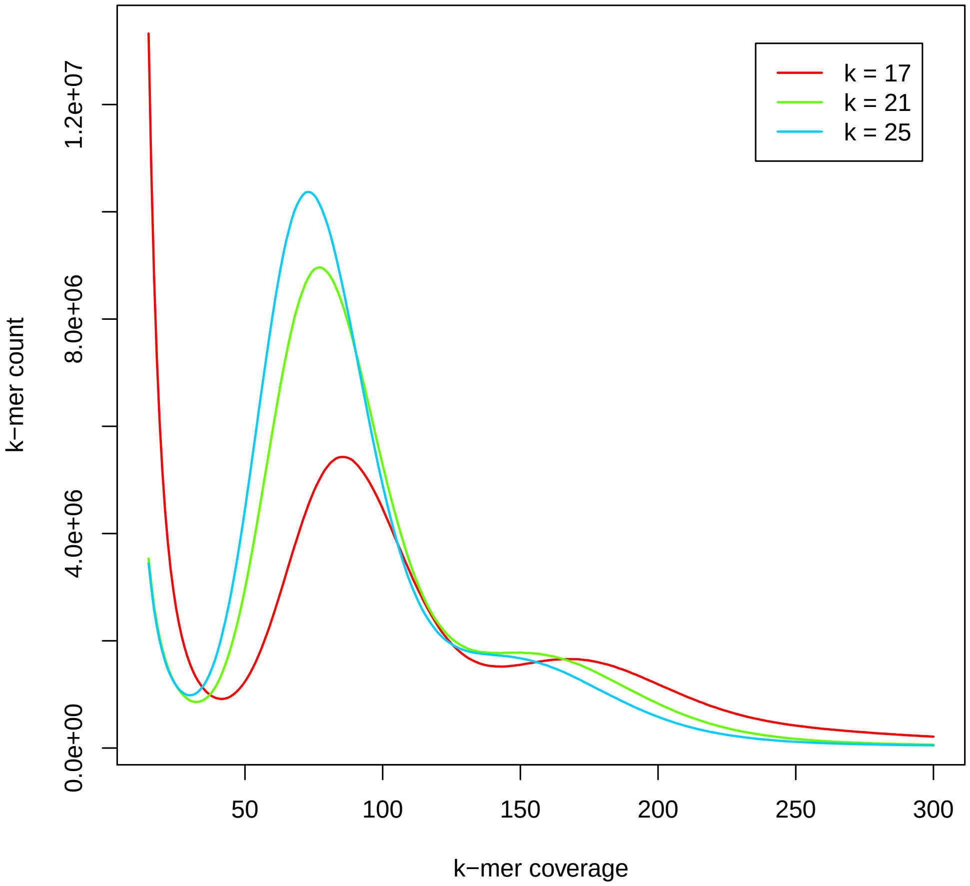


Figure D.1: k-mer histograms from Bogong moth Chromium 10x sequencing data.

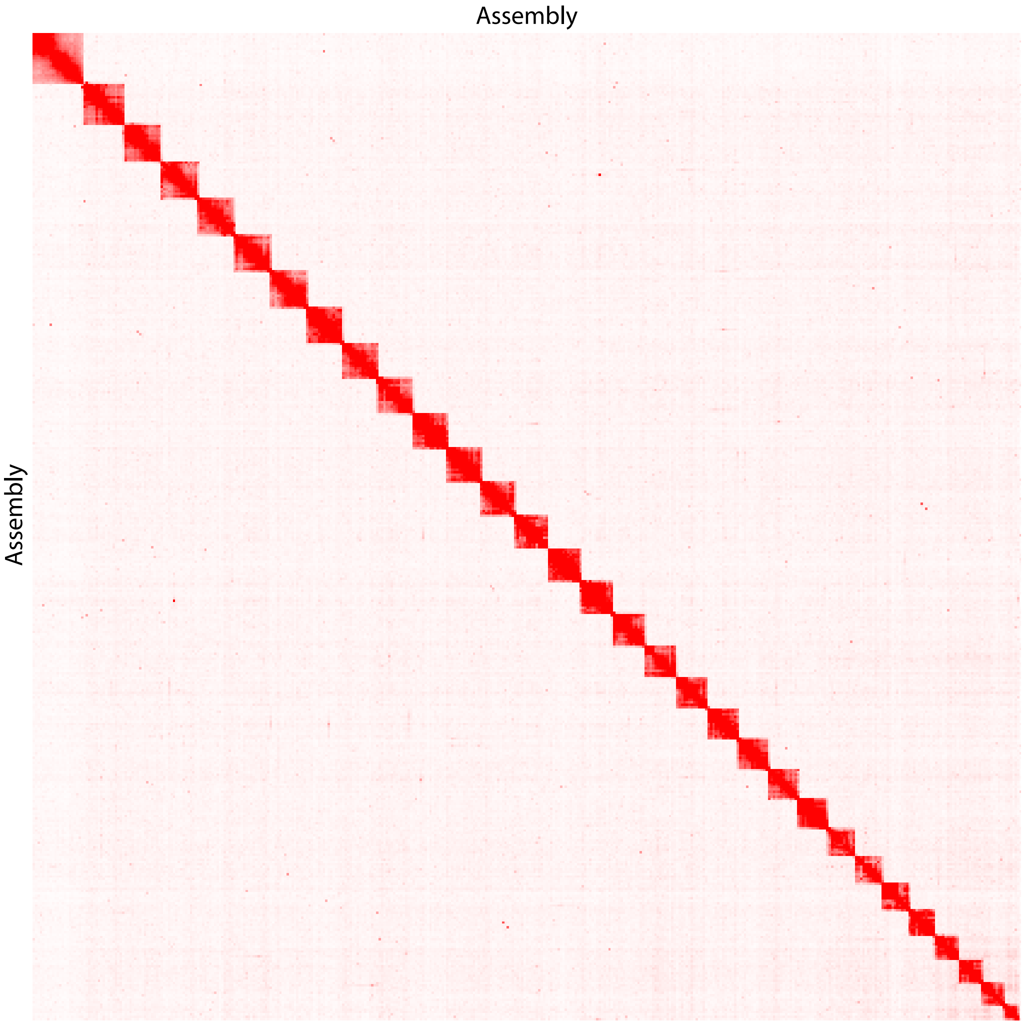


Figure D.2: HiC map of 31 megascaffolds/chromosomes produced using Juicebox Assembly Tools (Durand et al., 2016a).

D.1 Windel: Long-read assembly indel correction using short reads

When applied to the task of genome assembly, long-read technologies from Pacific Biosciences (PacBio) and Oxford Nanopore (ONT) have many advantages over more mature short read sequencing technologies. Importantly, long reads are able to span long complex genomic regions, enabling the assembly of sequences which are impossible to assemble using short reads alone. Long-read technologies do, however, suffer from high error rates on the order of 5-15% (Watson and Warr, 2019). Attempts have been made to correct these errors by incorporating higher quality (in terms of sequence error rate) short-read data (e.g. Pilon, Walker et al., 2014; Racon, Vaser et al., 2017). Nevertheless, some of these errors do propagate through to final assemblies.

Perhaps the most concerning types of error commonly observed in long-read assemblies are insertions and deletions (INDELs) (Watson and Warr, 2019), which can alter the interpretation of translated regions in genes by introducing frameshifts and premature stop codons. This type of error is most pronounced in collapsed consensus assemblies of diploid (and presumably polyploid) samples (Koren et al., 2019), which represent the majority of *de novo* assembly projects in the past few years. So far, the cause of the high indel error rate in these types of assemblies has been characterised only in general terms, which state they are a result of the high error rates of long-read sequencing platforms, combined with the challenge of assembling a collapsed consensus reference from a diploid sample. A more detailed characterisation of these errors is warranted, in the hope they might be fixed.

We propose that a simple algorithm, inspired by a qualitative assessment of an alignment of whole-genome short-read data to a Pilon-polished long-read assembly, is sufficient to correct many of the indel errors in the assembly which are missed by other methods.

During the Bogong moth genome assembly project, we assessed the impact of long-read INDEL errors on the quality of the genome assembly by aligning short reads to the assembly, which had already undergone multiple rounds of Pilon-polishing. Using a genome browser (Thorvaldsdóttir et al., 2013), we then looked at loci which had multiple reads supporting the introduction of a small indel. We observed that at many of these loci, most, if if not all, of the short reads aligned at that locus supported an edit (an insertion or deletion) of a particular length, but the exact coordinate of the required edit was not preserved across the reads. In fact, approximately half of the reads supported an edit at one coordinate, and the other half supported an edit at another, a few bases apart. We argue that in these cases, an edit should be made to remove the indel from the assembly at one of the coordinates, perhaps leaving a heterozygous single-nucleotide variant (SNV) at the locus.

This problem is not confined to the Bogong moth, and probably affects most recent diploid or polyploid genome assemblies, which use a combination of long and short reads. To illustrate this, we aligned short whole-genome-shotgun reads from *Homo sapiens* (NCBI accession ERR194147) to a state-of-the-art short-read-polished long-read assembly of the same sample by Koren et al. (2019), and indeed saw the same issue (Fig. D.3a).

Pilon (Walker et al., 2014) fails to make these edits because it requires indels which appear at different coordinates to be equivalent in order for them all to be considered as evidence supporting a single edit. If a locus is heterozygous for a SNV, then indel edits at different coordinates at that locus would not be equivalent, and Pilon would therefore not make an edit.

For the Bogong moth genome assembly, applying two additional rounds of Pilon, followed by one round of Windel and a final round of Pilon increased the number of BUSCOs in an already Pilon-polished assembly from 1580 (95.3% BUSCO complete) to 1598 (96.4% BUSCO complete).

D.1.1 IMPLEMENTATION

We developed an algorithm which considers evidence for indel edits in a sliding window across a short-read pileup on a draft assembly (Fig. D.3a, *red box*). Our algorithm does not require that the edits supported by the reads are equivalent, only that they are of the same type and length. It can therefore combine evidence from multiple lineages in a diploid or polyploid sample at heterozygous SNV loci, enabling corrections to the assembly which are missed by other tools, such as Pilon, which considers only a single column in an alignment pileup at a time (e.g. Fig. D.3a, *green and blue boxes*). The algorithm is implemented in a new software named Windel, which is written in Python, and is available at https://github.com/J-Wall/windel.

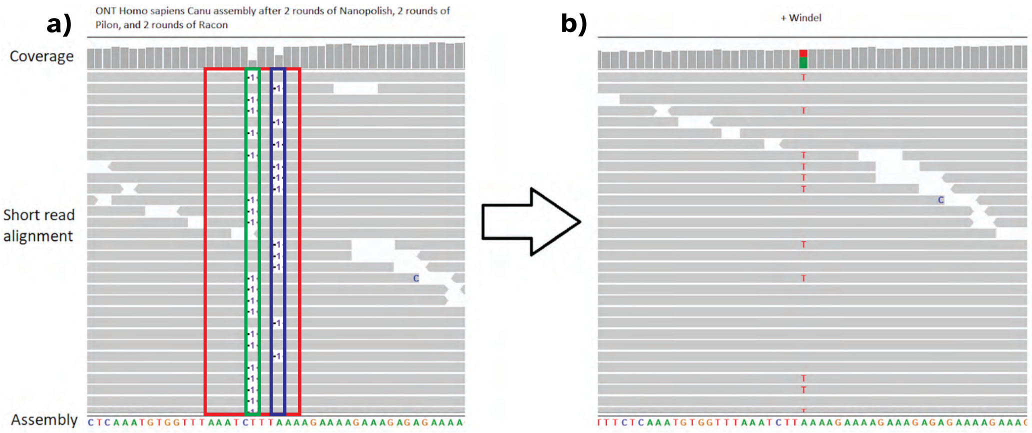
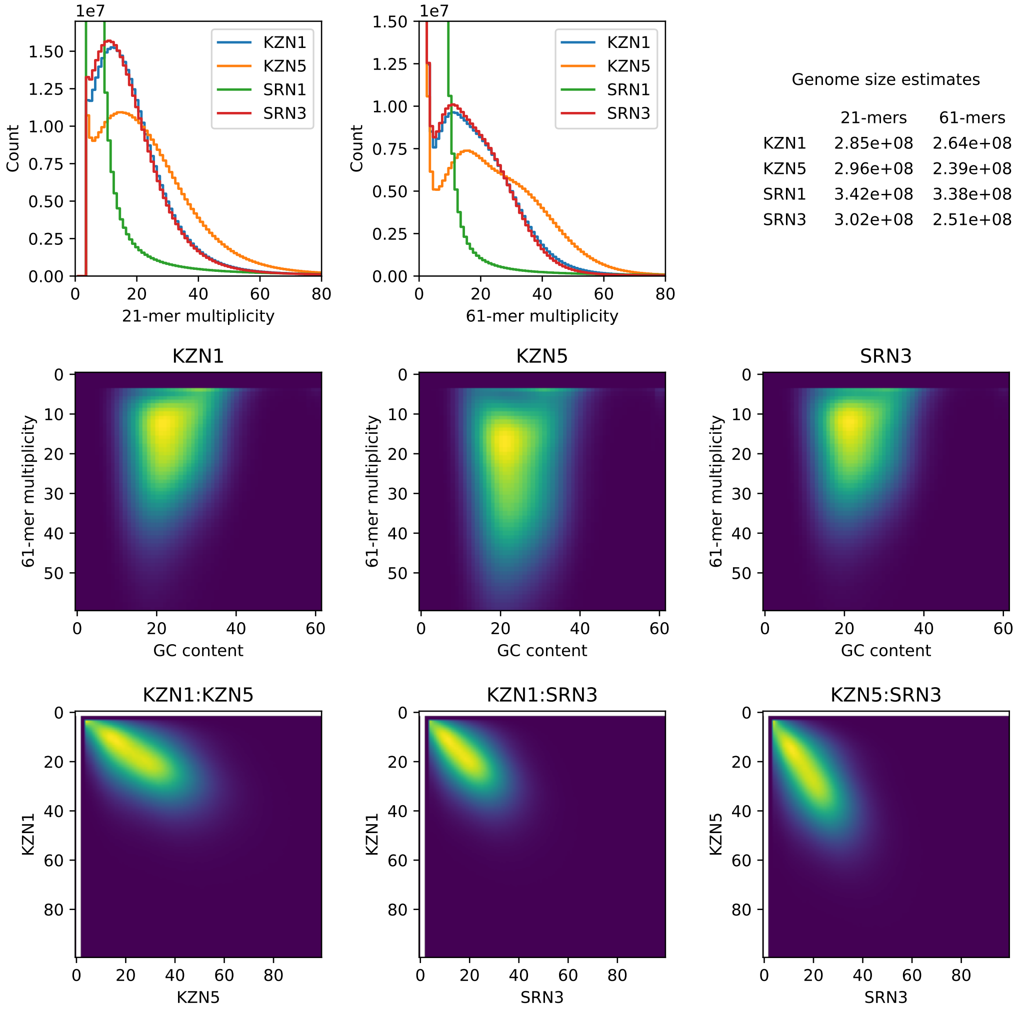


Figure D.3: Example of edit performed by Windel. **a)** Short-read alignment pileup viewed in IGV (Thorvaldsdóttir et al., 2013), showing disagreement between the reads as to the location of a deletion error (*green and blue boxes*). *Red box* shows sliding window considered by Windel when deciding whether or not to make an edit to the genome assembly. The assembly is from Oxford Nanopore long reads, and has already been polished using 2 rounds of Nanopolish, 2 rounds of Pilon, and 2 rounds of Racon. **b)** The result of applying Windel. The short reads now correctly reflect a heterozygous SNP at the edited locus.

D.2 *k*-mer analysis of *A. bogongae* samples

k-mer histograms of each sequencing library show evidence of contamination in one of the samples (SRN1; Fig. D.4, *top row*), but not in the other samples (Fig. D.4). Based on the location of peaks in the *k*-mer histograms of the non-contaminated samples, the size of the genome of *A. bogongae* was estimated to be just under 300 Mb (Fig. D.4, *top right*).



Genome size estimates

	21-mers	61-mers
KZN1	2.85e+08	2.64e+08
KZN5	2.96e+08	2.39e+08
SRN1	3.42e+08	3.38e+08
SRN3	3.02e+08	2.51e+08

Figure D.4: k-mer analysis of *A. bogongae* sequencing libraries. **Top row:** k-mer histograms and genome size estimates for each sample, with $k = 21$ and $k = 61$. SRN1 shows clear signs of contamination, with no discernible peak above multiplicity > 1. **Middle row:** 2D histograms of 61-mer multiplicity against GC content, for each sample (excluding SRN1). No relationship between GC content and 61-mer multiplicity is present, and GC content is equal across samples, suggesting that the reads come from a single species' genome. **Bottom row:** 2D 61-mer histograms of each pair of the uncontaminated samples show unimodal linear correlation of multiplicities, further supporting the notion that the reads come from a single species' genome.

Appendix E

Appendix to Chapter 6

E.1 Expected false discovery rate of missing genomic features from shotgun sequencing

Consider an experiment with the aim of determining whether a particular feature present in a species' reference genome exists in the genome of an individual sample of that species. In this experiment, the only data available are the reference genome, an annotation of the feature of interest, and whole genome shotgun sequencing reads from the sample. After mapping the reads to the reference, we decide if the feature is present by checking whether there are any reads which map to it. Shotgun sequencing reads are by definition randomly sampled from the genome of the sample. Therefore, the above-mentioned protocol could, by chance, lead to a false inference that the feature is missing from the sample genome. We wish to know how likely these false discoveries are. Equivalently, we wish to know the probability of shotgun sequencing reads missing an extant genomic feature by chance.

For the sake of tractability, we start with a number of assumptions:

1. Read sampling is unbiased.
2. Mapping is perfect.
3. Chromosomes do not have ends (e.g. they are circular. This is roughly equivalent to the feature of interest being far from the ends of a linear chromosome).
4. The feature is a single contiguous genomic region (e.g. a prokaryotic gene or eukaryotic exon).
5. The feature is considered present if any aligned read overlaps the feature by at least m nucleotide bases.

In reality, we expect 1 and 2 to hold (at least approximately) for genomic regions which are not redundant or highly repetitive. However, the validity of 2 may also be affected by the particular mapping software used, particularly in the context of sequences which exhibit substantial divergence between the sample and the reference. In prokaryotes, 3 holds, and it also approximately holds in eukaryotes with long chromosomes.

Let the length of the reference genome be G, the number of reads n, the read-length k, and the length of the genomic feature l. Then, given assumptions 1 and 3, the probability p, that a particular read overlaps the feature by m bases is given by

$$p = \frac{k + l - 2m + 1}{G} \quad , \quad m < k, m < l.$$

Therefore, the probability, which we will call Q, that n reads miss the region is given by

$$Q = (1 - p)^n \,.$$

Note that the average sequencing depth D is given by

$$D = \frac{nk}{G}.$$

Combining the above equalities gives

$$Q = \left(1 - \frac{k + l - 2m + 1}{G}\right)^{\frac{D}{k}G} . \tag{E.1}$$

In practice, and particularly when working with eukaryotic genomes, G is very large. It is therefore reasonable to use the asymptotic approximation of Eq. E.1 as $G \to \infty$. We begin with a change of variables, letting $a = k + l - 2m + 1$, and $b = \frac{D}{k}$. Then

$$\begin{aligned}
\lim_{G\to\infty} Q &= \lim_{G\to\infty} \left(1 - \frac{a}{G}\right)^{bG} \\
&= \lim_{G\to\infty} \exp\left(\log\left(1 - \frac{a}{G}\right)^{bG}\right) \\
&= \exp\left(b \lim_{G\to\infty} \frac{\log\left(1 - \frac{a}{G}\right)}{1/G}\right).
\end{aligned}$$

Applying l'Hôpital's rule, we obtain

$$\begin{aligned}
\lim_{G\to\infty} Q &= \exp\left(b \lim_{G\to\infty} \frac{\frac{\delta}{\delta G}\log\left(1-\frac{a}{G}\right)}{\frac{\delta}{\delta G}1/G} \right) \\
&= \exp\left(b \lim_{G\to\infty} \frac{aG}{a-G} \right) \\
&= \exp\left(ab \lim_{G\to\infty} \frac{1}{\frac{a}{G}-1} \right) \\
&= \exp\left(\frac{ab}{0-1} \right).
\end{aligned}$$

Substituting in the values for a and b gives us the asymptotic estimate of the probability of missing a feature by chance alone, given the above-mentioned assumptions. Namely,

$$Q \simeq e^{-(k+l-2m+1)\frac{D}{k}} \quad . \tag{E.2}$$

Fig. E.1 shows Q plotted against l for $k = 150$, $m = 1$, and various values of D.

The primary caveat of this result is the first two assumptions, namely that read sampling is unbiased and mapping is perfect. Clearly, deviations from either of these may result in the true value of Q going up or down. Nevertheless, when they do hold, even approximately, and read-depth is sufficient, we see that Q rapidly vanishes with increasing feature length. Naturally, the specificity required for detecting missing genomic features will vary depending on the research question, however a reasonable rule of thumb seems to be that any average read depth of about 10x and read length of 150 bp will give adequately low (approximately $<< 10^{-5}$) values of Q for features over about 100 bp.

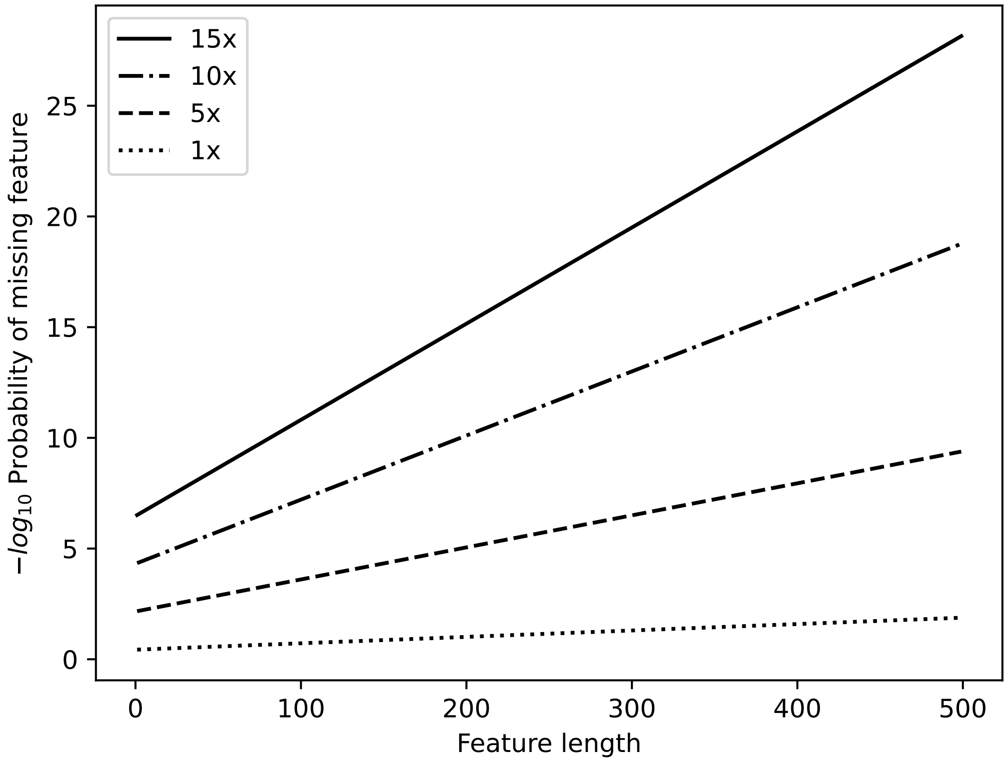


Figure E.1: Example plots of the probability of shotgun sequencing missing a genomic feature against feature length for various read-depths (1x, 5x, 10x and 15x). Plotted values were calculated based on 150 bp unpaired reads and a minimum read-feature overlap of 1 bp.

E.2 Sequencing quality control

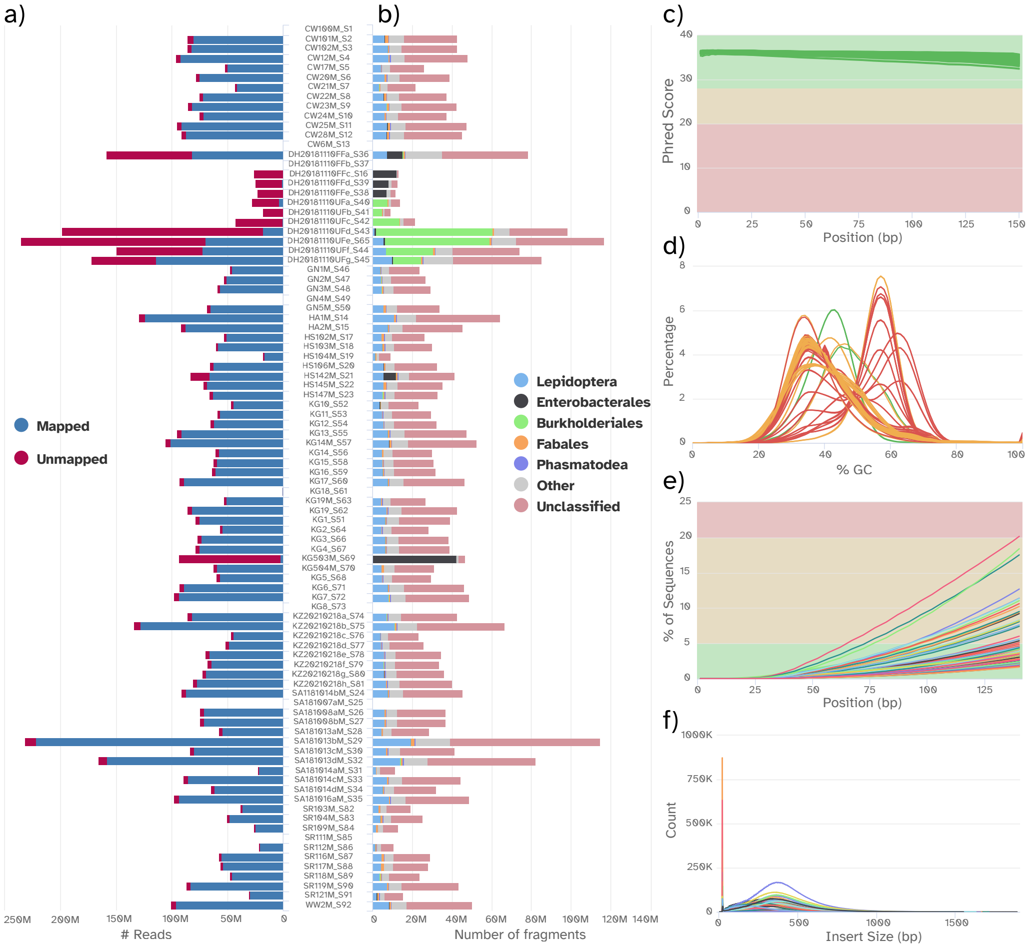


Figure E.2: Quality control plots for the whole-genome resequencing experiment. All plots were generated using MultiQC (Ewels et al., 2016). **a.** Total number of reads obtained from each Bogong moth sample, coloured by whether or not the read successfully mapped to the Bogong moth reference genome using BWA-MEM 2 (Vasimuddin et al., 2019). **b.** Number of reads falling into the five most common orders in the dataset, classified by Kraken 2 (Wood et al., 2019) using the NCBI nt database. Non-bogong DNA contamination was present in KG503M and a number of the DH samples (also evident in the mapping rates for those samples shown in **a**). Most reads are unclassified, since the nt database did not include Bogong moth data. **c–e.** Plots of quality-control metrics calculated using FastQC. **c.** Mean quality (Phred score) by read position for each sample. **d.** Distribution of GC content of reads for each sample. Most samples deviated from the expected distribution (red and orange traces). **e.** Sequencing adapter content by read position for each sample. **f.** Distribution of insert sizes of mapped reads, predicted by Picard tools "collectinsertsizemetrics".

E.3 Genes under selection

Table E.1: Genes co-located with top 1% of Tajima's D bins

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000000064/ HiC_scaffold_1:2369268-2395652	Protein RCC2 homolog	
AGING00000000801/ HiC_scaffold_2:5914087-5944489	WD repeat and HMG-box DNA-binding protein 1	GO:0005515
AGING00000000805/ HiC_scaffold_2:6033051-6071896	PDZ domain-containing protein 2	GO:0005515
AGING00000001133/ HiC_scaffold_2:13906487-14003790	Microspherule protein 1	GO:0071339, GO:0031011, GO:0002151, GO:0005515
AGING00000001134/ HiC_scaffold_2:13948874-13952480	Microspherule protein 1	GO:0071339, GO:0031011, GO:0002151
AGING00000001404/ HiC_scaffold_3:3848976-3882563	Unconventional myosin-XV	GO:0007605, GO:0003779, GO:0016459
AGING00000001405/ HiC_scaffold_3:3882578-3928713	Unconventional myosin-XV	GO:0005856, GO:0005524, GO:0016459, GO:0003774, GO:0005515
AGING00000001543/ HiC_scaffold_3:8709379-8966818	hypothetical protein	
AGING00000001734/ HiC_scaffold_3:14319458-14340513	Thrombospondin type-1 domain-containing protein 7A	
AGING00000002431/ HiC_scaffold_5:1228572-1241992	hypothetical protein	
AGING00000002432/ HiC_scaffold_5:1253812-1269683	hypothetical protein	
AGING00000002437/ HiC_scaffold_5:1288773-1297033	Centrobin	GO:0007099, GO:1902017
AGING00000002438/ HiC_scaffold_5:1298083-1312350	WD repeat-containing protein 44	GO:0005515
AGING00000002775/ HiC_scaffold_5:13689241-13702276	Protein SMG8	GO:0000184
AGING00000002776/ HiC_scaffold_5:13693255-13695581	hypothetical protein	GO:0003676
AGING00000002847/ HiC_scaffold_5:14754880-14829431	Adenomatous polyposis coli protein	GO:0008013, GO:0016055, GO:0030178, GO:0005515
AGING00000002852/ HiC_scaffold_5:14960822-14967058	Putative transporter svop-1	GO:0055085
AGING00000002853/ HiC_scaffold_5:14967900-14973306	hypothetical protein	
AGING00000003411/ HiC_scaffold_6:11091558-11092329	hypothetical protein	
AGING00000003412/ HiC_scaffold_6:11093555-11095269	Long-chain fatty acid transport protein 4	
AGING00000003511/ HiC_scaffold_6:15027588-15036782	Caspase-1	GO:0006508, GO:0004197, GO:0008234
AGING00000003543/ HiC_scaffold_6:16216555-16220815	hypothetical protein	GO:0005509
AGING00000003555/ HiC_scaffold_6:16379551-16384274	Enhancer of mRNA-decapping protein 4	
AGING00000003556/ HiC_scaffold_6:16385664-16386880	hypothetical protein	
AGING00000003564/ HiC_scaffold_6:16524555-16575625	Potassium voltage-gated channel subfamily H member 2, hypothetical protein	
AGING00000003567/ HiC_scaffold_6:16589983-16611266	THAP domain-containing protein 6	GO:0003676
AGING00000003574/ HiC_scaffold_6:16668050-16708580	Sodium/potassium-transporting ATPase subunit beta-1	GO:0006814, GO:0005890, GO:0006813
AGING00000003657/ HiC_scaffold_7:1775658-1797121	Retrovirus-related Pol polyprotein from transposon 17.6	
AGING00000003877/ HiC_scaffold_7:8911391-8961499	hypothetical protein	GO:0007156, GO:0005509, GO:0007155, GO:0005886, GO:0016020
AGING00000004193/ HiC_scaffold_8:1076499-1100777	Ionotropic receptor 75a	GO:0004970, GO:0016020, GO:0015276
AGING00000004194/ HiC_scaffold_8:1088325-1100777	hypothetical protein	
AGING00000004492/ HiC_scaffold_8:11605758-11621549	Adenylyltransferase and sulfurtransferase MOCS3	GO:0005829, GO:0002143, GO:0004792, GO:0008641
AGING00000004866/ HiC_scaffold_9:8214010-8256524	hypothetical protein	GO:0005576, GO:0042742
AGING00000005240/ HiC_scaffold_10:3427307-3620870	hypothetical protein	GO:0005509, GO:0019722
AGING00000005477/ HiC_scaffold_10:9428905-9446428	hypothetical protein	
AGING00000005478/ HiC_scaffold_10:9431492-9433905	hypothetical protein	
AGING00000005707/ HiC_scaffold_11:1066270-1087251	E3 ubiquitin-protein ligase listerin	GO:1990116, GO:1990112, GO:0061630
AGING00000005774/ HiC_scaffold_11:3225046-3329276	Calcium-activated potassium channel slowpoke	GO:0006813, GO:0016020, GO:0060072

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000005959/HiC_scaffold_11:9201787-9217898	CTP synthase	GO:0003883, GO:0006221, GO:0006241
AGING00000006284/HiC_scaffold_12:4513547-4827055	Sodium channel protein 60E	GO:0006814, GO:0001518, GO:0006811, GO:0005248, GO:0016020, GO:0055085, GO:0005515, GO:0005216
AGING00000006488/HiC_scaffold_12:10729908-10741299	Probable Ufm1-specific protease 1	
AGING00000006489/HiC_scaffold_12:10746507-10812278	Dynein heavy chain 10, axonemal	GO:0007018, GO:0003777, GO:0030286
AGING00000006490/HiC_scaffold_12:10815800-10834987	Dynein heavy chain 10, axonemal	
AGING00000006491/HiC_scaffold_12:10845046-10847722	Dynein heavy chain 10, axonemal	GO:0007018, GO:0003777
AGING00000006492/HiC_scaffold_12:10863151-10869094	Dynein heavy chain 10, axonemal, hypothetical protein	
AGING00000006493/HiC_scaffold_12:10870300-10871461	hypothetical protein	
AGING00000006494/HiC_scaffold_12:10873767-10875453	Dynein heavy chain 10, axonemal	
AGING00000006495/HiC_scaffold_12:10878024-10886928	hypothetical protein	
AGING00000006708/HiC_scaffold_13:1478657-1479534	hypothetical protein	GO:0006508, GO:0004252
AGING00000006816/HiC_scaffold_13:3930946-3957090	Pyruvate carboxylase, mitochondrial	GO:0004736, GO:0005524, GO:0006094, GO:0046872, GO:0006090, GO:0003824, GO:0009374
AGING00000007293/HiC_scaffold_14:4257100-4260378	hypothetical protein	
AGING00000007711/HiC_scaffold_14:14844098-15362665	Neuroligin-4, X-linked	
AGING00000007786/HiC_scaffold_15:1580018-1597946	Ral GTPase-activating protein subunit beta	GO:0005096, GO:0051056
AGING00000007841/HiC_scaffold_15:3820589-3842797	Histone-lysine N-methyltransferase ash1	GO:0005515, GO:0018024, GO:0005634, GO:0003682
AGING00000007963/HiC_scaffold_15:8878416-8885982	hypothetical protein, Unconventional myosin-IXb	GO:0005515
AGING00000007971/HiC_scaffold_15:9507085-9871837	Phorbol ester/diacylglycerol-binding protein unc-13, Protein unc-13 homolog A	GO:0019992, GO:0035556, GO:0007268
AGING00000007975/HiC_scaffold_15:9776289-9786157	hypothetical protein	
AGING00000008111/HiC_scaffold_15:14964668-15021121	Sorting nexin-27	GO:0006886, GO:0032266, GO:0035091, GO:0005515, GO:0007165
AGING00000008612/HiC_scaffold_16:13491822-13577258	hypothetical protein	GO:0008061, GO:0005576, GO:0006030, GO:0005515
AGING00000008903/HiC_scaffold_17:6952193-6957433	Zinc finger protein 658	GO:0003676
AGING00000008904/HiC_scaffold_17:6958290-6964039	Zinc finger protein 91	GO:0003676
AGING00000009101/HiC_scaffold_17:14333738-14346746	Mini-chromosome maintenance complex-binding protein	GO:0003676
AGING00000009105/HiC_scaffold_17:14389400-14438714	F-box only protein 32	GO:0005515
AGING00000009217/HiC_scaffold_18:2670389-3112613	Patj homolog	GO:0005515
AGING00000009275/HiC_scaffold_18:4817896-4846673	hypothetical protein	GO:0003676
AGING00000009276/HiC_scaffold_18:4846944-4893684	Endoplasmic reticulum metallopeptidase 1	
AGING00000009284/HiC_scaffold_18:5180457-5192374	MIP18 family protein galla-1	GO:0106035
AGING00000009285/HiC_scaffold_18:5192387-5193640	Autophagy-related protein 13 homolog	GO:0000045, GO:0006914, GO:1990316
AGING00000009286/HiC_scaffold_18:5196913-5198245	Cleavage and polyadenylation specificity factor subunit 1	GO:0005515
AGING00000009287/HiC_scaffold_18:5198724-5245122	Cleavage and polyadenylation specificity factor subunit 1	GO:0005515, GO:0005634, GO:0003676
AGING00000009301/HiC_scaffold_18:6324254-6391336	PRKCA-binding protein, hypothetical protein	GO:0019904, GO:0005515
AGING00000009315/HiC_scaffold_18:6788347-7083184	Down syndrome cell adhesion molecule-like protein Dscam2	GO:0005515
AGING00000009327/HiC_scaffold_18:7619824-7752028	ATP-dependent RNA helicase DHX30, ATP-dependent DNA/RNA helicase DHX36	GO:0004386
AGING00000009420/HiC_scaffold_18:11712422-12085055	hypothetical protein	
AGING00000009560/HiC_scaffold_19:2231781-2240513	Coiled-coil domain-containing protein 63	
AGING00000009751/HiC_scaffold_19:6086648-6098440	TBC1 domain family member 22A	
AGING00000009883/HiC_scaffold_19:9612444-9833171	Ras-like protein family member 10B	GO:0003924, GO:0005525
AGING00000009897/HiC_scaffold_19:10122536-10123950	hypothetical protein	

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000009954/HiC_scaffold_19:11218059-11236075	Probable ATP-dependent RNA helicase kurz	GO:0005524, GO:0004386, GO:0003676
AGING00000009979/HiC_scaffold_19:11856861-11862379	Alpha-tocopherol transfer protein-like	
AGING00000010213/HiC_scaffold_20:7086861-7111679	Protein UBASH3A homolog	
AGING00000010472/HiC_scaffold_20:13654144-13710748	Protein RRP5 homolog	GO:0006397, GO:0006396, GO:0003676, GO:0005515, GO:0005634
AGING00000010474/HiC_scaffold_20:13700870-13707809	hypothetical protein	
AGING00000010490/HiC_scaffold_20:14112474-14129247	Ribosome-releasing factor 2, mitochondrial	GO:0003924, GO:0005525
AGING00000010494/HiC_scaffold_20:14223632-14255248	Integrator complex subunit 1	GO:0034474, GO:0032039
AGING00000010495/HiC_scaffold_20:14256484-14270930	Integrator complex subunit 1	GO:0034474, GO:0032039
AGING00000010562/HiC_scaffold_21:1639004-2089934	Potassium voltage-gated channel protein Shaw	GO:0006813, GO:0008076, GO:0005249, GO:0016020, GO:0055085, GO:0006811, GO:0005216
AGING00000010977/HiC_scaffold_22:584372-1051012	Complexin	GO:0006836, GO:0019905
AGING00000010993/HiC_scaffold_22:1498661-2123841	hypothetical protein	GO:0007186, GO:0004930, GO:0016021
AGING00000010996/HiC_scaffold_22:1864835-1970394	hypothetical protein	
AGING00000011207/HiC_scaffold_22:10210215-10266241	hypothetical protein	GO:0005576, GO:0019731
AGING00000011208/HiC_scaffold_22:10217926-10278651	hypothetical protein	GO:0005576, GO:0019731
AGING00000011266/HiC_scaffold_22:12727183-12754763	Histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific	GO:0018024, GO:0005634, GO:0005515
AGING00000011267/HiC_scaffold_22:12727200-12736494	hypothetical protein	
AGING00000011632/HiC_scaffold_23:12084135-12103137	Nucleolar complex protein 3 homolog	
AGING00000011689/HiC_scaffold_23:14219322-14228278	Pancreatic lipase-related protein 3	GO:0052689
AGING00000011697/HiC_scaffold_24:958-16927	Zinc finger protein 112, Zinc finger and BTB domain-containing protein 24	GO:0008270, GO:0005634, GO:0003676
AGING00000011699/HiC_scaffold_24:75703-78971	Protein ALP1-like	
AGING00000011700/HiC_scaffold_24:76257-79501	hypothetical protein	
AGING00000011777/HiC_scaffold_24:2491006-2508802	DNA topoisomerase 3-beta-1	GO:0003916, GO:0003917, GO:0003677, GO:0006265
AGING00000011849/HiC_scaffold_24:5402074-5410323	Deoxycytidylate deaminase	GO:0016787, GO:0006220, GO:0008270, GO:0004132, GO:0003824
AGING00000011850/HiC_scaffold_24:5410466-5416624	Deoxycytidylate deaminase	GO:0016787, GO:0006220, GO:0008270, GO:0004132, GO:0003824
AGING00000011851/HiC_scaffold_24:5420190-5430696	N-alpha-acetyltransferase 15, NatA auxiliary subunit	
AGING00000012097/HiC_scaffold_25:1410867-1423128	Pancreatic lipase-related protein 2	GO:0052689, GO:0006629
AGING00000012098/HiC_scaffold_25:1425942-1439551	Pancreatic triacylglycerol lipase (Fragment)	GO:0052689, GO:0006629
AGING00000012141/HiC_scaffold_25:3352857-3363741	Vacuolar protein sorting-associated protein 35	GO:0015031, GO:0030906, GO:0042147
AGING00000012144/HiC_scaffold_25:3386256-3417900	Vacuolar protein sorting-associated protein 35	GO:0015031, GO:0030906, GO:0042147
AGING00000012164/HiC_scaffold_25:3932725-3933927	Prolow-density lipoprotein receptor-related protein 1	GO:0005509
AGING00000012165/HiC_scaffold_25:3933483-3933923	hypothetical protein	
AGING00000012166/HiC_scaffold_25:3934397-3937507	Vitellogenin receptor	
AGING00000012167/HiC_scaffold_25:3948895-3960886	Putative vitellogenin receptor	GO:0005515
AGING00000012168/HiC_scaffold_25:3961485-3964195	Putative vitellogenin receptor	GO:0005509
AGING00000012169/HiC_scaffold_25:3964968-3972561	Putative vitellogenin receptor	
AGING00000012205/HiC_scaffold_25:4680858-4771860	Glutathione synthetase	GO:0016874, GO:0006750, GO:0004363, GO:0005524
AGING00000012208/HiC_scaffold_25:4710314-4712935	snRNA-activating protein complex subunit 1	
AGING00000012209/HiC_scaffold_25:4713609-4717650	hypothetical protein	
AGING00000012409/HiC_scaffold_25:10311836-10384678	Dystonin	GO:0005856, GO:0007010, GO:0005509, GO:0008092, GO:0008017, GO:0005515
AGING00000012455/HiC_scaffold_25:11882977-11949816	hypothetical protein	

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000012539/ HiC_scaffold_26:1919703-1985091	hypothetical protein	
AGING00000012554/ HiC_scaffold_26:2370931-2466275	hypothetical protein	
AGING00000012629/ HiC_scaffold_26:4840353-5099393	hypothetical protein	GO:0007339, GO:0016020, GO:0005515
AGING00000012637/ HiC_scaffold_26:5389549-5407735	hypothetical protein	GO:0007339, GO:0016020, GO:0005515
AGING00000012693/ HiC_scaffold_26:9435046-9843365	Heterogeneous nuclear ribonucleoprotein L	GO:0006397, GO:0005634, GO:0003723, GO:0003676
AGING00000012726/ HiC_scaffold_26:11391457-11440614	EH domain-binding protein 1	GO:0005515
AGING00000012819/ HiC_scaffold_27:1369369-1404412	Solute carrier family 12 member 9	GO:0022857, GO:0016020, GO:0055085, GO:0006811
AGING00000012886/ HiC_scaffold_27:3339280-3352401	Histone lysine demethylase PHF8	
AGING00000012954/ HiC_scaffold_27:6759153-6814105	Catenin delta-2	GO:0098609, GO:0032956, GO:0010172, GO:0005913, GO:0045296, GO:0005515
AGING00000012969/ HiC_scaffold_27:8254999-8452338	Ankyrin-2, Ankyrin-3, hypothetical protein	GO:0007165, GO:0005515
AGING00000012980/ HiC_scaffold_27:9357490-9377086	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2	GO:0005515
AGING00000012981/ HiC_scaffold_27:9377607-9381992	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1	GO:0005515
AGING00000012987/ HiC_scaffold_27:9543520-9579901	Inositol 1,4,5-trisphosphate receptor	GO:0005216, GO:0055085, GO:0016020, GO:0006811
AGING00000013018/ HiC_scaffold_27:10419996-10544298	Voltage-dependent calcium channel type D subunit alpha-1	GO:0005245, GO:0070588, GO:0016020, GO:0055085, GO:0006811, GO:0005216, GO:0005891
AGING00000013057/ HiC_scaffold_28:24505-65543	Putative fatty acyl-CoA reductase CG5065	GO:0080019
AGING00000013066/ HiC_scaffold_28:233085-276361	hypothetical protein	
AGING00000013067/ HiC_scaffold_28:234233-240666	hypothetical protein	
AGING00000013071/ HiC_scaffold_28:282126-308606	hypothetical protein	
AGING00000013085/ HiC_scaffold_28:479343-497887	DNA-binding protein Ets97D	GO:0005634, GO:0003700, GO:0006355, GO:0043565
AGING00000013091/ HiC_scaffold_28:648382-654155	hypothetical protein	
AGING00000013113/ HiC_scaffold_28:1135691-1171459	Cullin-2	GO:0031461, GO:0006511, GO:0031625
AGING00000013114/ HiC_scaffold_28:1183493-1212330	Protein son of sevenless, hypothetical protein	GO:0005085, GO:0007264
AGING00000013115/ HiC_scaffold_28:1206851-1229983	hypothetical protein	
AGING00000013116/ HiC_scaffold_28:1213621-1216147	Protein son of sevenless	GO:0005085, GO:0007264
AGING00000013117/ HiC_scaffold_28:1218633-1234571	Protein son of sevenless	GO:0005085, GO:0007264, GO:0046982
AGING00000013118/ HiC_scaffold_28:1235410-1255202	SID1 transmembrane family member 1	GO:0033227, GO:0051033, GO:0016021
AGING00000013119/ HiC_scaffold_28:1258905-1275902	SID1 transmembrane family member 1	GO:0033227, GO:0051033, GO:0016021
AGING00000013135/ HiC_scaffold_28:1799160-1821343	Codanin-1	
AGING00000013148/ HiC_scaffold_28:2113808-2129225	tRNA pseudouridine(38/39) synthase	GO:0003723, GO:0009982, GO:0009451, GO:0001522
AGING00000013149/ HiC_scaffold_28:2137089-2196359	Krueppel-like factor 3	GO:0003676
AGING00000013158/ HiC_scaffold_28:2294165-2302381	Leucine-rich repeat transmembrane protein FLRT3	
AGING00000013160/ HiC_scaffold_28:2344661-2362381	Brain-specific homeobox protein	GO:0003677, GO:0043565, GO:0006355
AGING00000013162/ HiC_scaffold_28:2426172-2439452	Ankyrin repeat domain-containing protein 50	
AGING00000013165/ HiC_scaffold_28:2549014-2564891	Cleavage stimulation factor subunit 3, Protein suppressor of forked	GO:0006397, GO:0006396, GO:0005634, GO:0005515
AGING00000013175/ HiC_scaffold_28:2703438-2780577	Anoctamin-4	GO:0046983
AGING00000013178/ HiC_scaffold_28:2831116-2847285	Anoctamin-3, Anoctamin-5	GO:0046983
AGING00000013184/ HiC_scaffold_28:2976098-3036975	Protein diaphanous	GO:0007292, GO:0030036, GO:0003779, GO:0007015, GO:0017048, GO:0016043
AGING00000013186/ HiC_scaffold_28:3062238-3075629	hypothetical protein	GO:0007517, GO:0046983, GO:0006355
AGING00000013196/ HiC_scaffold_28:3233000-3249321	Isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial (Fragment)	GO:0055114, GO:0016616, GO:0004449, GO:0006099
AGING00000013197/ HiC_scaffold_28:3249583-3267588	Transcriptional repressor protein YY1	GO:0003676

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000013199/ HiC_scaffold_28:3276984-3323450	Mitogen-activated protein kinase-binding protein 1	GO:0005515
AGING00000013200/ HiC_scaffold_28:3278996-3292947	hypothetical protein	
AGING00000013203/ HiC_scaffold_28:3369281-3625740	Mitogen-activated protein kinase-binding protein 1	GO:0005515
AGING00000013210/ HiC_scaffold_28:3793510-3819055	DNA-binding protein Ets97D	GO:0005634, GO:0003700, GO:0006355, GO:0043565
AGING00000013212/ HiC_scaffold_28:3825720-3857577	hypothetical protein	
AGING00000013213/ HiC_scaffold_28:3849231-3851331	hypothetical protein	
AGING00000013214/ HiC_scaffold_28:3857644-3870591	hypothetical protein	
AGING00000013217/ HiC_scaffold_28:3879164-4002070	hypothetical protein	
AGING00000013218/ HiC_scaffold_28:3923240-4007466	hypothetical protein	
AGING00000013219/ HiC_scaffold_28:4013586-4014954	hypothetical protein	
AGING00000013220/ HiC_scaffold_28:4017844-4035065	Sorting and assembly machinery component 50 homolog A	GO:0019867
AGING00000013226/ HiC_scaffold_28:4121143-4134692	Protein son of sevenless	GO:0005085, GO:0007264
AGING00000013229/ HiC_scaffold_28:4159153-4190827	SID1 transmembrane family member 1	GO:0033227, GO:0051033, GO:0016021
AGING00000013239/ HiC_scaffold_28:4540142-4545274	Fatty acyl-CoA reductase wat	GO:0080019
AGING00000013240/ HiC_scaffold_28:4546604-4554934	Fatty acyl-CoA reductase 1	GO:0080019
AGING00000013243/ HiC_scaffold_28:4628253-4635102	Putative fatty acyl-CoA reductase CG5065, Fatty acyl-CoA reductase wat	GO:0080019
AGING00000013244/ HiC_scaffold_28:4636038-4637645	hypothetical protein	
AGING00000013245/ HiC_scaffold_28:4638746-4639973	Putative fatty acyl-CoA reductase CG5065	GO:0080019
AGING00000013262/ HiC_scaffold_28:5007142-5107769	Neurofibromin	GO:0043087, GO:0007165
AGING00000013271/ HiC_scaffold_28:5208414-5793564	hypothetical protein	
AGING00000013273/ HiC_scaffold_28:5857716-5857853	hypothetical protein	
AGING00000013274/ HiC_scaffold_28:5859039-5862707	hypothetical protein	GO:0008270, GO:0003676
AGING00000013304/ HiC_scaffold_28:7311446-7313892	hypothetical protein	
AGING00000013305/ HiC_scaffold_28:7349730-7429361	TBC1 domain family member 1	GO:0005515
AGING00000013335/ HiC_scaffold_28:8055574-8089064	Protein PAT1 homolog 1	GO:0000290
AGING00000013336/ HiC_scaffold_28:8067876-8068489	Uncharacterized protein PFD1115c	
AGING00000013341/ HiC_scaffold_28:8214912-8261956	Probable multidrug resistance-associated protein lethal(2)03659	GO:0005524, GO:0042626, GO:0055085, GO:0016887, GO:0016021
AGING00000013342/ HiC_scaffold_28:8216980-8217720	hypothetical protein	
AGING00000013343/ HiC_scaffold_28:8278944-8310683	Acetyl-CoA carboxylase	GO:0016874, GO:0005524, GO:0003989, GO:0006633
AGING00000013344/ HiC_scaffold_28:8311621-8358753	Acetyl-CoA carboxylase 2, Acetyl-CoA carboxylase 1	GO:0005524, GO:0003989, GO:0046872, GO:0006633
AGING00000013345/ HiC_scaffold_28:8318555-8326763	Odorant receptor 4	GO:0007608, GO:0005549, GO:0016020, GO:0004984
AGING00000013347/ HiC_scaffold_28:8359787-8403760	Acetyl-CoA carboxylase 2	
AGING00000013368/ HiC_scaffold_28:8806046-8831172	Glycogen [starch] synthase	GO:0004373, GO:0005978
AGING00000013369/ HiC_scaffold_28:8838080-8846745	Glycogen [starch] synthase	GO:0004373, GO:0005978
AGING00000013370/ HiC_scaffold_28:8911550-8961688	Carcinine transporter	GO:0022857, GO:0055085, GO:0016021
AGING00000013386/ HiC_scaffold_28:9281375-9349625	Multidrug resistance protein homolog 49	GO:0005524, GO:0042626, GO:0055085, GO:0016887, GO:0016021
AGING00000013408/ HiC_scaffold_28:10164130-10372182	hypothetical protein	GO:0008080
AGING00000013410/ HiC_scaffold_28:10383636-10476138	Copper-transporting ATPase 1	GO:0005507, GO:0030001, GO:0046872
AGING00000013411/ HiC_scaffold_28:10477650-10480624	Copper-transporting ATPase 1	GO:0000166
AGING00000013412/ HiC_scaffold_28:10483911-10493493	Copper-transporting ATPase 2	GO:0016021
AGING00000013413/ HiC_scaffold_28:10519055-10525219	hypothetical protein	
AGING00000013414/ HiC_scaffold_28:10526559-10533329	hypothetical protein	

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000013420/HiC_scaffold_28:10591380-10833369	Inactive serine protease scarface	GO:0006508, GO:0004252
AGING00000013427/HiC_scaffold_28:10928137-10982427	hypothetical protein	
AGING00000013428/HiC_scaffold_28:10951778-11056164	hypothetical protein	
AGING00000013441/HiC_scaffold_28:11434977-11444613	hypothetical protein	
AGING00000013442/HiC_scaffold_28:11445375-11460392	Scavenger receptor class B member 1	GO:0016020
AGING00000013449/HiC_scaffold_28:11601412-11615865	hypothetical protein	GO:0008080
AGING00000013450/HiC_scaffold_28:11616709-11737235	GTPase-activating protein CdGAPr	GO:0007165, GO:0005515
AGING00000013451/HiC_scaffold_28:11761431-11787625	SID1 transmembrane family member 1	GO:0033227, GO:0051033, GO:0016021
AGING00000013454/HiC_scaffold_28:11886610-11890246	hypothetical protein	
AGING00000013462/HiC_scaffold_29:14752-22064	hypothetical protein	
AGING00000013463/HiC_scaffold_29:21990-31745	Glucose-6-phosphatase	GO:0004346
AGING00000013464/HiC_scaffold_29:73289-90221	Segmentation protein even-skipped	GO:0003677, GO:0043565, GO:0006355
AGING00000013465/HiC_scaffold_29:99456-121951	Putative ammonium transporter 3	GO:0016020, GO:0008519, GO:0015696, GO:0072488
AGING00000013494/HiC_scaffold_29:1417547-1501117	Cullin-1	GO:0006511, GO:0031625
AGING00000013495/HiC_scaffold_29:1427315-1431467	Transposon Ty3-G Gag-Pol polyprotein	GO:0015074, GO:0003676
AGING00000013497/HiC_scaffold_29:1481661-1485785	LHFPL tetraspan subfamily member 2a protein	
AGING00000013498/HiC_scaffold_29:1501013-1547060	Cullin-1	GO:0031461, GO:0006511, GO:0031625
AGING00000013499/HiC_scaffold_29:1547510-1591420	hypothetical protein	GO:0003824
AGING00000013502/HiC_scaffold_29:1651853-1667364	hypothetical protein	
AGING00000013505/HiC_scaffold_29:1695716-1801697	hypothetical protein	
AGING00000013522/HiC_scaffold_29:2219550-2384153	Bromodomain adjacent to zinc finger domain protein 2B	GO:0005634, GO:0003677, GO:0005515
AGING00000013537/HiC_scaffold_29:2710808-2712867	Serine protease filzig	GO:0006508, GO:0004252
AGING00000013538/HiC_scaffold_29:2750487-2753431	hypothetical protein	
AGING00000013539/HiC_scaffold_29:2754327-2892805	hypothetical protein	
AGING00000013541/HiC_scaffold_29:2925237-2938802	Spliceosome-associated protein CWC27 homolog	GO:0017176, GO:0006457, GO:0000413, GO:0003755, GO:0016021, GO:0006506
AGING00000013543/HiC_scaffold_29:2950575-2988650	Probable chitinase 2	GO:0004553, GO:0005975, GO:0008061
AGING00000013550/HiC_scaffold_29:3029905-3109406	Serine proteinase stubble	GO:0006508, GO:0004252
AGING00000013596/HiC_scaffold_29:4455069-4673870	Transcription factor CP2	GO:0006357, GO:0003700
AGING00000013597/HiC_scaffold_29:4464250-4496377	hypothetical protein	
AGING00000013602/HiC_scaffold_29:4683031-4713346	Unconventional myosin-Vb	GO:0051015, GO:0005524, GO:0016459, GO:0003774
AGING00000013603/HiC_scaffold_29:4714623-4758698	Unconventional myosin-Va	GO:0005515, GO:0005524, GO:0016459, GO:0003774
AGING00000013607/HiC_scaffold_29:4852660-4864373	hypothetical protein	
AGING00000013608/HiC_scaffold_29:4866359-4870736	hypothetical protein	GO:0003676
AGING00000013609/HiC_scaffold_29:4871342-4908474	Zinc finger protein 718	GO:0003676
AGING00000013611/HiC_scaffold_29:4909061-4932466	hypothetical protein	GO:0003676
AGING00000013612/HiC_scaffold_29:4938489-4941576	Guanine nucleotide-binding protein G(f) subunit alpha	GO:0007186, GO:0019001, GO:0003924, GO:0031683, GO:0007165
AGING00000013613/HiC_scaffold_29:4944931-4945341	Histone H2B.1, hypothetical protein	GO:0046982, GO:0003677, GO:0000786
AGING00000013614/HiC_scaffold_29:4946490-4946900	hypothetical protein	GO:0046982, GO:0003677, GO:0000786
AGING00000013623/HiC_scaffold_29:5105863-5122772	Aspartate–tRNA ligase, cytoplasmic	GO:0006422, GO:0006418, GO:0004815, GO:0005524, GO:0000166, GO:0004812, GO:0005737
AGING00000013672/HiC_scaffold_29:6312963-6346721	hypothetical protein	

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000013697/ HiC_scaffold_29:6968169-7114178	Talin-1	GO:0005856, GO:0030276, GO:0051015, GO:0007155, GO:0006897, GO:0005925, GO:0007016, GO:0005200, GO:0001726, GO:0003779
AGING00000013703/ HiC_scaffold_29:7199776-7307130	Aldo-keto reductase AKR2E4	GO:0055114, GO:0016491
AGING00000013707/ HiC_scaffold_29:7264339-7290689	Talin-1	GO:0030276, GO:0006897, GO:0005925, GO:0007016, GO:0005200, GO:0001726, GO:0003779
AGING00000013708/ HiC_scaffold_29:7305032-7316381	D-galacturonate reductase	GO:0055114, GO:0016491
AGING00000013709/ HiC_scaffold_29:7319055-7320253	Aldo-keto reductase AKR2E4	GO:0055114, GO:0016491
AGING00000013744/ HiC_scaffold_29:8047504-8086800	Nucleolar GTP-binding protein 1	GO:0005730, GO:0005525
AGING00000013755/ HiC_scaffold_29:8253375-8303953	Mediator of RNA polymerase II transcription subunit 16	GO:0016592, GO:0005515
AGING00000013756/ HiC_scaffold_29:8276378-8476018	Neogenin	GO:0005887, GO:0045664, GO:0045663, GO:0048598, GO:0007224
AGING00000013757/ HiC_scaffold_29:8480846-8486153	Netrin receptor DCC, hypothetical protein	GO:0005515, GO:0005042, GO:0007411, GO:0006915, GO:0007399
AGING00000013758/ HiC_scaffold_29:8487058-8490739	Receptor-type tyrosine-protein phosphatase delta	GO:0005515
AGING00000013760/ HiC_scaffold_29:8498893-8500915	hypothetical protein	GO:0005515
AGING00000013761/ HiC_scaffold_29:8505522-8519725	hypothetical protein	
AGING00000013773/ HiC_scaffold_29:8876092-8992183	Protein encore	GO:0003676
AGING00000013774/ HiC_scaffold_29:8903254-8945883	R3H domain-containing protein 1	GO:0003676
AGING00000013820/ HiC_scaffold_29:10196642-10253557	Methionine–tRNA ligase, cytoplasmic	GO:0006431, GO:0005524, GO:0006418, GO:0000166, GO:0005515, GO:0004812, GO:0004825
AGING00000013821/ HiC_scaffold_29:10257006-10263074	Probable NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	GO:0016020, GO:0009055, GO:0008137
AGING00000013846/ HiC_scaffold_29:10683326-10800069	Ornithine decarboxylase antizyme 1	GO:0008073
AGING00000013849/ HiC_scaffold_29:10815832-11048006	Diacylglycerol kinase 1	GO:0005509, GO:0003951, GO:0016301, GO:0035556, GO:0004143, GO:0007165
AGING00000013855/ HiC_scaffold_29:11080726-11109658	hypothetical protein	
AGING00000013859/ HiC_scaffold_29:11172146-11173310	hypothetical protein	
AGING00000013868/ HiC_scaffold_29:11305015-11309188	hypothetical protein	GO:0003676
AGING00000013869/ HiC_scaffold_29:11305839-11309588	hypothetical protein	
AGING00000013870/ HiC_scaffold_30:26153-39877	Galectin-4	GO:0030246
AGING00000013884/ HiC_scaffold_30:501779-560288	Transcription elongation regulator 1	GO:0005515
AGING00000013887/ HiC_scaffold_30:614461-656197	U4/U6 small nuclear ribonucleoprotein Prp3	GO:0000398, GO:0046540
AGING00000013889/ HiC_scaffold_30:656450-677656	U4/U6 small nuclear ribonucleoprotein Prp4	GO:0005515
AGING00000013913/ HiC_scaffold_30:1392831-1649746	Spectrin beta chain, non-erythrocytic 5	GO:0005543, GO:0005515
AGING00000013914/ HiC_scaffold_30:1557579-1631594	Spectrin beta chain, non-erythrocytic 1	GO:0005515
AGING00000013916/ HiC_scaffold_30:1738989-1757073	Leucine-rich repeat serine/threonine-protein kinase 1	GO:0004672, GO:0006468, GO:0005524
AGING00000013917/ HiC_scaffold_30:1758027-1762607	Leucine-rich repeat serine/threonine-protein kinase 1	
AGING00000013918/ HiC_scaffold_30:1765277-1771252	Leucine-rich repeat serine/threonine-protein kinase 1	
AGING00000013919/ HiC_scaffold_30:1767708-1784214	hypothetical protein	
AGING00000013920/ HiC_scaffold_30:1772936-1783896	Leucine-rich repeat serine/threonine-protein kinase 1	GO:0005515
AGING00000013923/ HiC_scaffold_30:1815823-1820966	hypothetical protein	
AGING00000013924/ HiC_scaffold_30:1817780-1826920	Ankyrin-2	GO:0005515
AGING00000013925/ HiC_scaffold_30:1833835-1847404	Translation initiation factor eIF-2B subunit beta	GO:0044237
AGING00000013926/ HiC_scaffold_30:1837480-1848099	hypothetical protein	
AGING00000013947/ HiC_scaffold_30:2365996-2441136	hypothetical protein	

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000013949/HiC_scaffold_30:2488686-2514261	Major facilitator superfamily domain-containing protein 6	
AGING00000013970/HiC_scaffold_30:2948507-2950379	hypothetical protein	
AGING00000013971/HiC_scaffold_30:2953058-2954445	hypothetical protein	
AGING00000013978/HiC_scaffold_30:3111214-3114977	hypothetical protein	
AGING00000013979/HiC_scaffold_30:3153918-3199087	Synaptic vesicular amine transporter	GO:0055085
AGING00000013985/HiC_scaffold_30:3320908-3446379	Mannosyl-oligosaccharide glucosidase, Mannosyl-oligosaccharide glucosidase GCS1	GO:0009311, GO:0004573, GO:0003824
AGING00000014004/HiC_scaffold_30:3727518-3737962	Cilia- and flagella-associated protein 91, hypothetical protein	
AGING00000014020/HiC_scaffold_30:3953101-3958583	Cardioactive peptide	
AGING00000014024/HiC_scaffold_30:4038743-4043790	hypothetical protein	
AGING00000014028/HiC_scaffold_30:4100985-4108655	DNA-dependent protein kinase catalytic subunit	GO:0016301, GO:0005515
AGING00000014029/HiC_scaffold_30:4109132-4119748	hypothetical protein	GO:0006879, GO:0008199, GO:0006826
AGING00000014035/HiC_scaffold_30:4199404-4209116	hypothetical protein	GO:0008270, GO:0005634, GO:0003676
AGING00000014052/HiC_scaffold_30:4339702-4345068	PR domain zinc finger protein 5	GO:0008270, GO:0005634, GO:0003676
AGING00000014084/HiC_scaffold_30:4958353-4975849	Exportin-7-A, Exportin-7-B	GO:0051169
AGING00000014093/HiC_scaffold_30:5169251-5265368	Heat shock 70 kDa protein cognate 5	GO:0006457, GO:0005524, GO:0051082
AGING00000014115/HiC_scaffold_30:5821167-5859598	hypothetical protein	
AGING00000014137/HiC_scaffold_30:6120891-6162678	Serine protease snake, hypothetical protein	GO:0006508, GO:0004252
AGING00000014144/HiC_scaffold_30:6280933-6343352	Putative U5 small nuclear ribonucleoprotein 200 kDa helicase	
AGING00000014145/HiC_scaffold_30:6343304-6357466	Putative U5 small nuclear ribonucleoprotein 200 kDa helicase	
AGING00000014146/HiC_scaffold_30:6367096-6504496	Putative U5 small nuclear ribonucleoprotein 200 kDa helicase	GO:0005524, GO:0003676
AGING00000014147/HiC_scaffold_30:6403294-6710995	Mediator of RNA polymerase II transcription subunit 28	
AGING00000014148/HiC_scaffold_30:6566339-6770959	Phenoloxidase-activating factor 2	GO:0006508, GO:0004252
AGING00000014149/HiC_scaffold_30:6786636-6816634	Dipeptidyl peptidase 8	GO:0006508, GO:0008236
AGING00000014151/HiC_scaffold_30:6841798-6866870	Metal transporter CNNM2, Metal transporter CNNM4	
AGING00000014152/HiC_scaffold_30:6904315-6939053	Metal transporter CNNM2	
AGING00000014173/HiC_scaffold_30:7252497-7277709	hypothetical protein	
AGING00000014186/HiC_scaffold_30:7563424-7685071	Serine protease Hayan	GO:0006508, GO:0004252
AGING00000014188/HiC_scaffold_30:7606610-7655702	Serine protease Hayan, Serine protease snake	GO:0006508, GO:0004252
AGING00000014191/HiC_scaffold_30:7830910-7861208	Poly [ADP-ribose] polymerase tankyrase	GO:0016055, GO:0051225, GO:0003950, GO:0005515, GO:0032212
AGING00000014192/HiC_scaffold_30:7841554-7847880	hypothetical protein	
AGING00000014193/HiC_scaffold_30:7862084-7966030	STE20-like serine/threonine-protein kinase, Serine/threonine-protein kinase 10	GO:0006468, GO:0004672, GO:0005524, GO:0004674
AGING00000014194/HiC_scaffold_30:7908639-7937935	STE20-like serine/threonine-protein kinase	GO:0004674, GO:0004672, GO:0005524, GO:0006468
AGING00000014212/HiC_scaffold_30:8240599-8296788	Ionotropic receptor 25a	GO:0005216, GO:0038023, GO:0004970, GO:0006811, GO:0016020, GO:0015276
AGING00000014213/HiC_scaffold_30:8263932-8313869	Complex I assembly factor ACAD9, mitochondrial	GO:0055114, GO:0016627
AGING00000014214/HiC_scaffold_30:8319449-8364436	Glutamate receptor ionotropic, kainate 2	GO:0004970, GO:0016020, GO:0015276
AGING00000014215/HiC_scaffold_30:8365284-8370873	Glutamate receptor ionotropic, kainate 2 (Fragment)	GO:0016020, GO:0015276
AGING00000014216/HiC_scaffold_30:8385323-8420688	hypothetical protein	GO:0005319, GO:0006869
AGING00000014217/HiC_scaffold_30:8480123-8497296	Glutamate receptor ionotropic, kainate 2, hypothetical protein	GO:0004970, GO:0016020, GO:0015276
AGING00000014222/HiC_scaffold_30:8690426-8799013	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase epsilon-1	GO:0005085, GO:0007264
AGING00000014223/HiC_scaffold_30:8799913-8805964	hypothetical protein	
AGING00000014227/HiC_scaffold_30:8933965-8983251	hypothetical protein	

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000014228/HiC_scaffold_30:8957108-8974008	hypothetical protein	
AGING00000014248/HiC_scaffold_30:9250196-9278064	Dynactin subunit 1	
AGING00000014264/HiC_scaffold_30:9492662-9525039	hypothetical protein	
AGING00000014296/HiC_scaffold_30:10163402-10262402	Protein mahjong	GO:0016567, GO:0005515
AGING00000014297/HiC_scaffold_30:10176772-10192631	hypothetical protein	
AGING00000014298/HiC_scaffold_30:10218398-10270867	EP300-interacting inhibitor of differentiation 3	GO:0030915, GO:0005634, GO:0006281
AGING00000014304/HiC_scaffold_30:10352543-10377703	hypothetical protein	
AGING00000014314/HiC_scaffold_30:10488933-10493859	hypothetical protein	GO:0030176, GO:0006886
AGING00000014315/HiC_scaffold_30:10499424-10503461	RE1-silencing transcription factor	GO:0003676
AGING00000014325/HiC_scaffold_31:16908-124003	Protein l(2)37Cc, Oxysterol-binding protein-related protein 8	GO:0016020
AGING00000014337/HiC_scaffold_31:469284-597112	SEC23-interacting protein	GO:0046872
AGING00000014338/HiC_scaffold_31:570656-634525	Protein MEMO1	
AGING00000014341/HiC_scaffold_31:668658-945310	Leishmanolysin-like peptidase	GO:0006508, GO:0016020, GO:0007155, GO:0004222
AGING00000014342/HiC_scaffold_31:683092-761642	hypothetical protein	
AGING00000014343/HiC_scaffold_31:762371-773485	Ferrochelatase, mitochondrial	GO:0006783, GO:0004325
AGING00000014345/HiC_scaffold_31:824016-836855	Leishmanolysin-like peptidase	
AGING00000014367/HiC_scaffold_31:1474519-1498486	G2/M phase-specific E3 ubiquitin-protein ligase	
AGING00000014372/HiC_scaffold_31:1544536-1645402	Translational regulator orb2, Cytoplasmic polyadenylation element-binding protein 2	GO:0045182, GO:0003730, GO:0006417, GO:0003676
AGING00000014396/HiC_scaffold_31:1997664-2026210	Zinc finger protein 345	GO:0003676
AGING00000014397/HiC_scaffold_31:2019922-2021905	hypothetical protein	
AGING00000014398/HiC_scaffold_31:2026477-2034330	4-hydroxybenzoate polyprenyltransferase, mitochondrial	GO:0016765, GO:0004659, GO:0016021
AGING00000014406/HiC_scaffold_31:2149290-2216510	Nuclear pore complex protein Nup98-Nup96	GO:0017056, GO:0005643
AGING00000014407/HiC_scaffold_31:2149651-2157314	hypothetical protein	
AGING00000014408/HiC_scaffold_31:2216985-2232168	Zinc finger protein 771	GO:0008270, GO:0005634, GO:0003676
AGING00000014431/HiC_scaffold_31:2558267-2711619	hypothetical protein	GO:0005525, GO:0003924, GO:0005200, GO:0005874, GO:0007017
AGING00000014432/HiC_scaffold_31:2590917-2638817	Zinc finger protein 761, Zinc finger and BTB domain-containing protein 41	GO:0003676
AGING00000014436/HiC_scaffold_31:2750135-2761299	Serine protease persephone, Serine protease snake	GO:0006508, GO:0004252
AGING00000014453/HiC_scaffold_31:3132762-3211446	Gelsolin	GO:0051015
AGING00000014454/HiC_scaffold_31:3147514-3160469	Gelsolin	GO:0051015
AGING00000014455/HiC_scaffold_31:3165043-3172701	L-dopachrome tautomerase yellow-f	
AGING00000014457/HiC_scaffold_31:3220047-3246438	Gelsolin	GO:0051015
AGING00000014458/HiC_scaffold_31:3249870-3272829	L-dopachrome tautomerase yellow-f2	
AGING00000014460/HiC_scaffold_31:3276395-3301145	L-dopachrome tautomerase yellow-f2	
AGING00000014464/HiC_scaffold_31:3446796-3516827	Protein phosphatase 1 regulatory subunit 37	GO:0005515
AGING00000014472/HiC_scaffold_31:3799411-3819648	Cytoplasmic FMR1-interacting protein	
AGING00000014473/HiC_scaffold_31:3817259-3819509	hypothetical protein	
AGING00000014474/HiC_scaffold_31:3820972-3833171	Cytoplasmic FMR1-interacting protein	
AGING00000014481/HiC_scaffold_31:3908603-3974952	hypothetical protein	
AGING00000014496/HiC_scaffold_31:4158352-4166806	hypothetical protein	
AGING00000014497/HiC_scaffold_31:4168644-4168922	hypothetical protein	
AGING00000014498/HiC_scaffold_31:4168916-4176278	hypothetical protein	
AGING00000014500/HiC_scaffold_31:4192221-4213328	Xaa-Pro aminopeptidase 1	GO:0070006, GO:0016787

Gene ID/Coordinate (Top 1% Tajima's D)	Gene product	GO terms
AGING00000014525/HiC_scaffold_31:4624578-4632579	hypothetical protein	GO:0003676
AGING00000014526/HiC_scaffold_31:4636993-4639505	Homeobox protein PKNOX2	GO:0003677, GO:0006355
AGING00000014538/HiC_scaffold_31:5111643-5292070	hypothetical protein	
AGING00000014542/HiC_scaffold_31:5206990-5312104	Zinc finger protein 271	GO:0003676
AGING00000014543/HiC_scaffold_31:5315275-5328529	hypothetical protein	GO:0008270, GO:0003676
AGING00000014544/HiC_scaffold_31:5317020-5324613	hypothetical protein	
AGING00000014546/HiC_scaffold_31:5345335-5359375	Lipase member H	GO:0052689, GO:0006629
AGING00000014559/HiC_scaffold_31:5609991-5638241	BTB/POZ domain-containing protein 7	GO:0005515, GO:0061138
AGING00000014576/HiC_scaffold_31:5961784-5972572	Zinc finger protein 595	GO:0003676
AGING00000014594/HiC_scaffold_31:6576631-6603266	hypothetical protein	
AGING00000014597/HiC_scaffold_31:6652433-6661426	hypothetical protein	GO:0051015
AGING00000014598/HiC_scaffold_31:6665289-6666628	MIP18 family protein galla-2	GO:0106035
AGING00000014599/HiC_scaffold_31:6667122-6671007	hypothetical protein	GO:0008270, GO:0005634
AGING00000014608/HiC_scaffold_31:6793739-6862563	Signal transducer and activator of transcription 5B	GO:0003677, GO:0006355, GO:0003700, GO:0005634, GO:0007165
AGING00000014617/HiC_scaffold_31:7031836-7080530	Zinc finger protein 2 homolog	GO:0055085, GO:0003676

Table E.2: Genes co-located with bottom 1% of Tajima's D bins

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000000033/HiC_scaffold_1:1270943-1469135	Krueppel-like factor 6	GO:0003676
AGING00000000096/HiC_scaffold_1:3184319-3280381	Glutamate receptor 1	GO:0004970, GO:0016020, GO:0015276
AGING00000000238/HiC_scaffold_1:8829776-8882767	Roquin-1	GO:0046872
AGING00000000239/HiC_scaffold_1:8884022-8946502	hypothetical protein	GO:0016021
AGING00000000241/HiC_scaffold_1:8907295-8910305	PiggyBac transposable element-derived protein 3	
AGING00000000246/HiC_scaffold_1:9193187-9197398	hypothetical protein	
AGING00000000247/HiC_scaffold_1:9194509-9195074	hypothetical protein	
AGING00000000248/HiC_scaffold_1:9198810-9199584	hypothetical protein	
AGING00000000249/HiC_scaffold_1:9220512-9306913	Laminin subunit alpha	GO:0007155
AGING00000000259/HiC_scaffold_1:9735237-9774225	Deformed epidermal autoregulatory factor 1	GO:0003677
AGING00000000260/HiC_scaffold_1:9746673-9747837	hypothetical protein	
AGING00000000268/HiC_scaffold_1:10149472-10186129	Mitogen-activated protein kinase kinase kinase 12	GO:0004672, GO:0006468, GO:0005524
AGING00000000269/HiC_scaffold_1:10190591-10191580	Peroxisomal membrane protein PEX16	
AGING00000000270/HiC_scaffold_1:10200333-10240809	ADP-ribosylation factor GTPase-activating protein 1	GO:0005096
AGING00000000273/HiC_scaffold_1:10367348-10368169	Histone deacetylase 8	
AGING00000000275/HiC_scaffold_1:10429187-10461147	hypothetical protein	
AGING00000000301/HiC_scaffold_1:11220779-11232813	6-phosphogluconate dehydrogenase, decarboxylating	GO:0016491, GO:0055114, GO:0004616, GO:0050661, GO:0006098
AGING00000000302/HiC_scaffold_1:11233261-11249870	Repetin	
AGING00000000328/HiC_scaffold_1:11765501-11809875	Immunoglobulin domain-containing protein oig-4	
AGING00000000329/HiC_scaffold_1:11811636-11963135	Potassium channel subfamily T member 1	GO:0006813, GO:0016020
AGING00000000348/HiC_scaffold_1:12703217-12785801	Nuclear hormone receptor FTZ-F1	GO:0003677, GO:0006355, GO:0008270, GO:0003700, GO:0003707, GO:0005634, GO:0043565
AGING00000000474/HiC_scaffold_1:16890969-17120918	SCY1-like protein 2	GO:0006468, GO:0004672, GO:0005524

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000000698/HiC_scaffold_2:2363336-2492235	Acetylcholine receptor subunit alpha-type acr-16, CHRNA7-FAM7A fusion protein, Neuronal acetylcholine receptor subunit alpha-7	GO:0034220, GO:0004888, GO:0005230, GO:0016021, GO:0022848, GO:0045211, GO:0006811, GO:0005216
AGING00000000722/HiC_scaffold_2:3459470-3465892	Peptidyl-alpha-hydroxyglycine alpha-amidating lyase 1	
AGING00000000723/HiC_scaffold_2:3464544-3472079	Protein RRNAD1	
AGING00000000792/HiC_scaffold_2:5765744-5836064	Insulin-like receptor	GO:0007169, GO:0004714, GO:0005524, GO:0006468, GO:0005515, GO:0004672, GO:0016020, GO:0004713
AGING00000000796/HiC_scaffold_2:5855019-5860130	Glycosaminoglycan xylosylkinase homolog	
AGING00000000797/HiC_scaffold_2:5869287-5891095	Short stature homeobox protein 2 (Fragment)	GO:0003677, GO:0043565, GO:0006355
AGING00000000804/HiC_scaffold_2:6019027-6021947	hypothetical protein	
AGING00000000827/HiC_scaffold_2:6322962-6335952	hypothetical protein	GO:0004553, GO:0005975
AGING00000000828/HiC_scaffold_2:6334920-6370736	Transcriptional coactivator YAP1	GO:0005515
AGING00000000829/HiC_scaffold_2:6369652-6376853	Zinc finger protein 266, Zinc finger protein 225, Zinc finger protein 226	GO:0008270, GO:0005634, GO:0003676
AGING00000000830/HiC_scaffold_2:6380511-6399483	Nicotinate phosphoribosyltransferase	GO:0004514, GO:0004516, GO:0009435, GO:0003824
AGING00000000831/HiC_scaffold_2:6405563-6445924	Adenylyl cyclase-associated protein 1	GO:0007010, GO:0000902, GO:0003779
AGING00000000832/HiC_scaffold_2:6442090-6443473	hypothetical protein	
AGING00000000833/HiC_scaffold_2:6445814-6452239	Enteropeptidase	GO:0005044, GO:0005515, GO:0006508, GO:0004252, GO:0016020
AGING00000000834/HiC_scaffold_2:6452361-6455867	hypothetical protein	
AGING00000000835/HiC_scaffold_2:6495204-6519787	hypothetical protein	
AGING00000000836/HiC_scaffold_2:6503671-6516845	hypothetical protein	
AGING00000000857/HiC_scaffold_2:6768238-6777209	Transportin-1	GO:0006606, GO:0005515
AGING00000000858/HiC_scaffold_2:6772215-6777209	hypothetical protein	
AGING00000000859/HiC_scaffold_2:6772808-6784570	Doublesex- and mab-3-related transcription factor 2	GO:0005634, GO:0003700, GO:0006355, GO:0043565
AGING00000000893/HiC_scaffold_2:7258849-7320951	Voltage-dependent calcium channel subunit alpha-2/delta-3	
AGING00000000989/HiC_scaffold_2:10007661-10044628	Protein grainyhead	GO:0006357, GO:0003700
AGING00000000990/HiC_scaffold_2:10054552-10055009	hypothetical protein	
AGING00000001097/HiC_scaffold_2:12844400-13028222	Very low-density lipoprotein receptor	GO:0005509, GO:0005515
AGING00000001103/HiC_scaffold_2:13548222-13625333	Leucine-rich repeat neuronal protein 2	GO:0005515
AGING00000001150/HiC_scaffold_2:14378992-14476888	DNA-binding protein SATB2	GO:0006338, GO:0003677
AGING00000001168/HiC_scaffold_2:15019018-15050150	Tyrosine-protein kinase Dnt	GO:0006468, GO:0004672, GO:0005524
AGING00000001192/HiC_scaffold_2:16411098-16437383	Transcription factor Ken	GO:0003676, GO:0005515
AGING00000001196/HiC_scaffold_2:16476533-16627051	Protein bunched, class 1/class 3/D/E isoforms	GO:0003700, GO:0006355
AGING00000001595/HiC_scaffold_3:9695792-9718527	hypothetical protein	
AGING00000001605/HiC_scaffold_3:10098139-10185636	hypothetical protein	
AGING00000001637/HiC_scaffold_3:11387565-11420530	Meteorin-like protein	
AGING00000001678/HiC_scaffold_3:12946765-12948362	hypothetical protein	
AGING00000001738/HiC_scaffold_3:14414486-14469666	Serine/threonine-protein kinase LMTK2	GO:0006468, GO:0004672, GO:0005524
AGING00000001947/HiC_scaffold_4:4017647-4039689	Teneurin-m	
AGING00000001990/HiC_scaffold_4:4697976-4868109	Protein O-mannosyl-transferase Tmtc3	GO:0005515
AGING00000002028/HiC_scaffold_4:6057525-6060111	Achaete-scute complex protein T3	GO:0006357, GO:0046983, GO:0003677
AGING00000002047/HiC_scaffold_4:6405839-6460770	Rhombotin-1	
AGING00000002078/HiC_scaffold_4:7276753-7306575	Tyrosine-protein kinase Btk29A	GO:0005524, GO:0006468, GO:0035556, GO:0005515, GO:0004672, GO:0004713
AGING00000002079/HiC_scaffold_4:7292130-7296472	hypothetical protein	
AGING00000002094/HiC_scaffold_4:7712626-8019899	Teneurin-m	

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000002151/ HiC_scaffold_4:9252939-9296327	Paired box protein Pax-6	GO:0003677, GO:0006355
AGING00000002252/ HiC_scaffold_4:11631232-11667753	Protein sprouty	GO:0016020, GO:0007275, GO:0009966
AGING00000002306/ HiC_scaffold_4:13470139-13595577	Zwei Ig domain protein zig-8	
AGING00000002328/ HiC_scaffold_4:14751434-14934181	Zinc finger protein jing homolog	GO:0003676
AGING00000002377/ HiC_scaffold_4:17045831-17061655	Troponin T	GO:0005861, GO:0006937
AGING00000002405/ HiC_scaffold_5:360829-376503	hypothetical protein	
AGING00000002406/ HiC_scaffold_5:376514-463428	hypothetical protein	
AGING00000002468/ HiC_scaffold_5:2918890-2985317	CCR4-NOT transcription complex subunit 6	
AGING00000002469/ HiC_scaffold_5:2926796-2936255	hypothetical protein	
AGING00000002504/ HiC_scaffold_5:4145133-4468504	Peripheral plasma membrane protein CASK	GO:0006468, GO:0004672, GO:0005524, GO:0005515
AGING00000002616/ HiC_scaffold_5:8030015-8096175	Lachesin	
AGING00000002681/ HiC_scaffold_5:10559959-10607130	hypothetical protein, Homeobox protein abdominal-A, Homeobox protein abdominal-A homolog (Fragment), Homeobox protein abdominal-A homolog	GO:0003677, GO:0043565, GO:0006355
AGING00000002685/ HiC_scaffold_5:10877496-10877801	Homeotic protein ultrabithorax	GO:0003677, GO:0043565, GO:0006355
AGING00000002686/ HiC_scaffold_5:11130595-11173791	Homeotic protein antennapedia	GO:0003677, GO:0006355, GO:0003700, GO:0005634, GO:0043565
AGING00000002690/ HiC_scaffold_5:11359360-11363228	hypothetical protein	
AGING00000002703/ HiC_scaffold_5:11650947-11654426	Leucine-rich repeat protein soc-2 homolog	GO:0005515
AGING00000002704/ HiC_scaffold_5:11656157-11659812	Ras-related C3 botulinum toxin substrate 1	GO:0003924, GO:0007264, GO:0005525
AGING00000002705/ HiC_scaffold_5:11659997-11663572	Testin	GO:0008270
AGING00000002826/ HiC_scaffold_5:14254957-14270784	Ras GTPase-activating protein-binding protein 2	GO:0003676
AGING00000002827/ HiC_scaffold_5:14274114-14280590	RNA-binding protein squid	GO:0003729, GO:0003676
AGING00000003054/ HiC_scaffold_6:2361037-2547610	Leucine-rich repeat neuronal protein 2	GO:0005515
AGING00000003058/ HiC_scaffold_6:2565082-2624363	Patronin	GO:0005516, GO:0030507, GO:0031175, GO:0008017, GO:0005515
AGING00000003065/ HiC_scaffold_6:2674702-2903167	Protein spire	GO:0016192, GO:0045010, GO:0003779
AGING00000003164/ HiC_scaffold_6:5771126-5852275	Whirlin	GO:0005515
AGING00000003206/ HiC_scaffold_6:6306395-6384090	T-box transcription factor TBX20	GO:0003700, GO:0006355, GO:0005634
AGING00000003208/ HiC_scaffold_6:6433373-6476890	T-box transcription factor TBX20	GO:0003700, GO:0006355, GO:0005634
AGING00000003275/ HiC_scaffold_6:7683741-7967415	Neural-cadherin	GO:0007156, GO:0005509, GO:0007155, GO:0005886, GO:0016020
AGING00000003319/ HiC_scaffold_6:9024563-9032241	F-box/SPRY domain-containing protein 1	GO:0005515
AGING00000003370/ HiC_scaffold_6:10168715-10239500	Neuroligin-1	
AGING00000003404/ HiC_scaffold_6:10758503-10982457	hypothetical protein, Protein muscleblind	GO:0046872
AGING00000003407/ HiC_scaffold_6:11017140-11021470	Muscleblind-like protein 2	GO:0046872
AGING00000003416/ HiC_scaffold_6:11206255-11245716	Protein ultraspiracle homolog	GO:0003677, GO:0006355, GO:0008270, GO:0003700, GO:0003707, GO:0005634, GO:0043565
AGING00000003452/ HiC_scaffold_6:12066614-12198093	LIM domain transcription factor LMO4	
AGING00000003453/ HiC_scaffold_6:12198416-12233612	hypothetical protein	
AGING00000003466/ HiC_scaffold_6:12739385-12818022	Serine/threonine-protein kinase 24	GO:0006468, GO:0004672, GO:0005524
AGING00000003478/ HiC_scaffold_6:13100495-13120014	Protein Wnt-1	GO:0005102, GO:0016055, GO:0005576, GO:0007275
AGING00000003619/ HiC_scaffold_7:50209-50535	hypothetical protein	
AGING00000003620/ HiC_scaffold_7:50209-57909	Palmitoyltransferase app	GO:0016409
AGING00000003621/ HiC_scaffold_7:60359-67079	Solute carrier family 35 member E1 homolog	
AGING00000003622/ HiC_scaffold_7:70276-84848	Endoplasmin	GO:0006457, GO:0005524, GO:0051082

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000003623/ HiC_scaffold_7:78146-83069	hypothetical protein	
AGING00000003624/ HiC_scaffold_7:85881-86291	hypothetical protein	GO:0043248
AGING00000003755/ HiC_scaffold_7:6297999-6383045	GATA-binding factor C, GATA-binding factor 2	GO:0008270, GO:0003700, GO:0006355, GO:0043565
AGING00000003832/ HiC_scaffold_7:7752442-7821711	hypothetical protein	
AGING00000003888/ HiC_scaffold_7:9249980-9264902	hypothetical protein	
AGING00000003889/ HiC_scaffold_7:9265066-9368587	Protein still life, isoform SIF type 1	GO:0005085, GO:0035556, GO:0007165, GO:0005515
AGING00000003968/ HiC_scaffold_7:11081745-11598954	Polypyrimidine tract-binding protein 2	GO:0006397, GO:0005634, GO:0003723, GO:0003676
AGING00000003969/ HiC_scaffold_7:11099741-11100641	hypothetical protein	
AGING00000004086/ HiC_scaffold_7:14936506-14948601	Netrin receptor unc-5	GO:0005042, GO:0038007, GO:0016021
AGING00000004245/ HiC_scaffold_8:2989932-3933889	CUGBP Elav-like family member 3-A	GO:0003676
AGING00000004375/ HiC_scaffold_8:7890509-8005846	Cadherin-related tumor suppressor	GO:0007156, GO:0005509, GO:0007155, GO:0005886, GO:0016020
AGING00000004376/ HiC_scaffold_8:8008470-8009210	Cadherin-related tumor suppressor	GO:0007156, GO:0005509, GO:0007155, GO:0005886, GO:0016020
AGING00000004428/ HiC_scaffold_8:9205476-9331744	Protein O-mannosyl-transferase TMTC2	GO:0005515
AGING00000004447/ HiC_scaffold_8:9919665-10054437	Dachshund homolog 1	
AGING00000004494/ HiC_scaffold_8:11705062-11817461	Cadherin-86C	GO:0005509, GO:0016020
AGING00000004572/ HiC_scaffold_8:14860930-15229671	Protein dachsous	GO:0007156, GO:0005509, GO:0007155, GO:0005886, GO:0016020
AGING00000004734/ HiC_scaffold_9:2006958-2151807	Max dimerization protein 4	GO:0046983
AGING00000004764/ HiC_scaffold_9:3999119-4119562	Leucine zipper putative tumor suppressor 2 homolog	
AGING00000004797/ HiC_scaffold_9:5568280-5582924	hypothetical protein	
AGING00000004807/ HiC_scaffold_9:5820542-5849906	Tyrosine-protein phosphatase	GO:0008138, GO:0016311, GO:0004725, GO:0016791
AGING00000004812/ HiC_scaffold_9:6184290-6240740	Transcription factor collier	GO:0046983, GO:0003700, GO:0003677, GO:0006355
AGING00000004887/ HiC_scaffold_9:8704989-8733255	Glutamate receptor ionotropic, kainate 2	GO:0005216, GO:0038023, GO:0004970, GO:0006811, GO:0016020, GO:0015276
AGING00000004947/ HiC_scaffold_9:9890802-10157813	Lysine-specific histone demethylase 1A	GO:0034720, GO:0003677, GO:0016491, GO:0055114, GO:0006355, GO:0050660, GO:0005515, GO:0005634
AGING00000004949/ HiC_scaffold_9:10240830-10368737	Toll-like receptor 6	GO:0007165, GO:0005515
AGING00000004952/ HiC_scaffold_9:10944353-10948384	Toll-like receptor 6	GO:0007165, GO:0005515
AGING00000004982/ HiC_scaffold_9:12264955-12271509	hypothetical protein	
AGING00000005155/ HiC_scaffold_10:473143-576194	Histone deacetylase 4	
AGING00000005173/ HiC_scaffold_10:998461-1010248	hypothetical protein	GO:0003676
AGING00000005329/ HiC_scaffold_10:5967337-5970450	hypothetical protein	
AGING00000005447/ HiC_scaffold_10:7778818-7870988	hypothetical protein	
AGING00000005449/ HiC_scaffold_10:8042155-8079590	hypothetical protein	GO:0005089
AGING00000005457/ HiC_scaffold_10:8226901-8284811	Autism susceptibility gene 2 protein	
AGING00000005458/ HiC_scaffold_10:8332328-8482994	hypothetical protein	
AGING00000005479/ HiC_scaffold_10:9482380-9589952	Protein alan shepard	GO:1990904, GO:0003723, GO:0003676
AGING00000005527/ HiC_scaffold_10:10339246-10342783	hypothetical protein	GO:0050909, GO:0016021
AGING00000005528/ HiC_scaffold_10:10347317-10352063	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	
AGING00000005551/ HiC_scaffold_10:11565221-11756344	Putative polypeptide N-acetylgalactosaminyltransferase 9	
AGING00000005573/ HiC_scaffold_10:12511707-12674536	POU domain, class 6, transcription factor 2	GO:0003700, GO:0003677, GO:0006355
AGING00000005580/ HiC_scaffold_10:12936620-13004790	BMP and activin membrane-bound inhibitor homolog	GO:0090263, GO:0030512
AGING00000005636/ HiC_scaffold_10:15622838-15642725	Myosin heavy chain, muscle	GO:0051015, GO:0005524, GO:0016459, GO:0003774, GO:0005515

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000005743/ HiC_scaffold_11:1704780-1987809	Centaurin-gamma-1A	GO:0005096, GO:0003924, GO:0005525
AGING00000005866/ HiC_scaffold_11:5799980-5889939	Fibropellin-1	GO:0005509, GO:0005515
AGING00000005889/ HiC_scaffold_11:6575839-6582004	Ras-related protein Rab-8A	GO:0003924, GO:0005525
AGING00000005890/ HiC_scaffold_11:6630134-6733937	Ras-related and estrogen-regulated growth inhibitor-like protein	GO:0003924, GO:0016020, GO:0007165, GO:0005525
AGING00000005935/ HiC_scaffold_11:8070057-8203287	Sterile alpha motif domain-containing protein 5	GO:0005515
AGING00000005947/ HiC_scaffold_11:8525453-8619422	Forkhead box protein O	GO:0003700, GO:0006355, GO:0043565
AGING00000005979/ HiC_scaffold_11:9918688-9960912	Acetylcholine receptor subunit alpha-like, Neuronal acetylcholine receptor subunit alpha-3	GO:0034220, GO:0004888, GO:0005230, GO:0016021, GO:0022848, GO:0045211, GO:0006811, GO:0005216
AGING00000006114/ HiC_scaffold_11:16102664-16116076	hypothetical protein	
AGING00000006142/ HiC_scaffold_12:207521-215708	hypothetical protein	
AGING00000006348/ HiC_scaffold_12:6078755-6175810	Membralin	
AGING00000006349/ HiC_scaffold_12:6182978-6192447	Transposable element Tc1 transposase	
AGING00000006376/ HiC_scaffold_12:7096804-7175738	ADAMTS-like protein 4	GO:0008233
AGING00000006823/ HiC_scaffold_13:4044351-4564213	Uncharacterized protein CG43867	GO:0005856
AGING00000006829/ HiC_scaffold_13:4571167-4670186	Serine/threonine-protein kinase minibrain	GO:0004712, GO:0005524, GO:0046777, GO:0006468, GO:0004672
AGING00000006870/ HiC_scaffold_13:6140435-6268097	Protein sprint	GO:0007165
AGING00000006911/ HiC_scaffold_13:7309038-7313352	Transmembrane protein fend	
AGING00000006920/ HiC_scaffold_13:7747472-7796214	Serine/threonine-protein kinase SIK2	GO:0006468, GO:0004672, GO:0005524, GO:0004674
AGING00000006927/ HiC_scaffold_13:7832601-8081077	Bromo adjacent homology domain-containing 1 protein	GO:0003682
AGING00000006941/ HiC_scaffold_13:8468935-8850132	Semaphorin-2A	GO:0030215, GO:0005515
AGING00000006969/ HiC_scaffold_13:9845113-9979363	Fasciclin-2	GO:0016020, GO:0007155, GO:0005515
AGING00000007045/ HiC_scaffold_13:12888839-13222116	Plexin A3, Acyl-CoA:lysophosphatidylglycerol acyltransferase 1	GO:0017154, GO:0071526, GO:0016746, GO:0005515, GO:0007165
AGING00000007231/ HiC_scaffold_14:1647554-1700281	hypothetical protein	
AGING00000007375/ HiC_scaffold_14:6922557-7080588	Gamma-aminobutyric acid type B receptor subunit 1	GO:0007186, GO:0004930, GO:0004965, GO:0016021
AGING00000007381/ HiC_scaffold_14:7286789-7342469	hypothetical protein	
AGING00000007382/ HiC_scaffold_14:7349327-7370776	Short transient receptor potential channel 5	GO:0005515, GO:0005216, GO:0005262, GO:0070588, GO:0055085, GO:0006811, GO:0016020
AGING00000007397/ HiC_scaffold_14:7643370-7715017	hypothetical protein	
AGING00000007409/ HiC_scaffold_14:7963203-8010279	Host cell factor	GO:0005515
AGING00000007410/ HiC_scaffold_14:7978740-7987753	Rab-like protein 6, hypothetical protein	GO:0003924, GO:0005525
AGING00000007426/ HiC_scaffold_14:8170169-8236754	hypothetical protein	
AGING00000007428/ HiC_scaffold_14:8229637-8270524	Matrix metalloproteinase-14	GO:0004222, GO:0031012, GO:0008270, GO:0008237, GO:0006508
AGING00000007443/ HiC_scaffold_14:8472374-8505218	hypothetical protein	
AGING00000007449/ HiC_scaffold_14:8640947-8657485	Potassium voltage-gated channel subfamily H member 7	GO:0006813, GO:0005216, GO:0005249, GO:0016021, GO:0055085, GO:0006811, GO:0016020
AGING00000007470/ HiC_scaffold_14:8992917-9114046	Pseudouridine-5'-phosphate glycosidase	GO:0016798
AGING00000007471/ HiC_scaffold_14:9003067-9089829	CD82 antigen	GO:0016021
AGING00000007495/ HiC_scaffold_14:9542974-9591553	Fibroblast growth factor receptor 3, hypothetical protein	
AGING00000007510/ HiC_scaffold_14:10068290-10069214	hypothetical protein	
AGING00000007511/ HiC_scaffold_14:10069575-10195768	Cyclic nucleotide-gated cation channel alpha-3	
AGING00000007628/ HiC_scaffold_14:11888194-11890502	Zinc finger protein 670	GO:0003676
AGING00000007629/ HiC_scaffold_14:11893532-11895095	Heparan-sulfate 6-O-sulfotransferase 1	GO:0008146, GO:0016021

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000007630/ HiC_scaffold_14:11896561-11900063	hypothetical protein	
AGING00000007667/ HiC_scaffold_14:12613426-12761356	Division abnormally delayed protein	GO:0045880, GO:0062023, GO:0030513, GO:0009966, GO:0046658, GO:0090263
AGING00000007751/ HiC_scaffold_15:269847-465640	Zwei Ig domain protein zig-8	
AGING00000007798/ HiC_scaffold_15:1878461-1878879	hypothetical protein	
AGING00000007838/ HiC_scaffold_15:3536947-3621265	Furin-like protease 2	GO:0006508, GO:0004252
AGING00000007918/ HiC_scaffold_15:6760191-6782432	Protein giant-lens	
AGING00000007923/ HiC_scaffold_15:7166649-7498144	Small conductance calcium-activated potassium channel protein	GO:0005516, GO:0006813, GO:0016286, GO:0016021, GO:0015269
AGING00000007994/ HiC_scaffold_15:10936157-10973682	hypothetical protein	
AGING00000007995/ HiC_scaffold_15:10973946-10982031	Centrosome-associated zinc finger protein CP190	GO:0005515, GO:0003676
AGING00000008029/ HiC_scaffold_15:11546192-11644399	hypothetical protein	
AGING00000008210/ HiC_scaffold_16:2608475-2638139	Helicase domino	GO:0005524
AGING00000008293/ HiC_scaffold_16:5330070-5555722	Myelin transcription factor 1-like protein	GO:0008270, GO:0003700, GO:0006355, GO:0005634
AGING00000008315/ HiC_scaffold_16:6605568-6942921	RNA-binding protein Musashi homolog Rbp6	GO:0003723, GO:0003676
AGING00000008317/ HiC_scaffold_16:7003947-7080081	hypothetical protein	
AGING00000008355/ HiC_scaffold_16:7623803-7714662	Transmembrane protein 132E, Transmembrane protein 132B	
AGING00000008517/ HiC_scaffold_16:10997924-11003527	hypothetical protein	
AGING00000008545/ HiC_scaffold_16:11713529-11794846	hypothetical protein, Protein phosphatase 1E	GO:0004722, GO:0006470, GO:0043169, GO:0003824
AGING00000008546/ HiC_scaffold_16:11752055-11767744	hypothetical protein	
AGING00000008592/ HiC_scaffold_16:12513393-12631139	Serine/threonine-protein kinase NLK	GO:0004707, GO:0004672, GO:0006468, GO:0005524
AGING00000008593/ HiC_scaffold_16:12593156-12593575	hypothetical protein	
AGING00000008618/ HiC_scaffold_16:13764415-14084891	Protein outspread	
AGING00000008623/ HiC_scaffold_16:14127762-14197576	Mediator of RNA polymerase II transcription subunit 13	GO:0006357, GO:0016592, GO:0003712
AGING00000008764/ HiC_scaffold_17:3391464-3395423	hypothetical protein	
AGING00000008765/ HiC_scaffold_17:3394445-3395127	hypothetical protein	
AGING00000008878/ HiC_scaffold_17:6629563-6642662	F-box/LRR-repeat protein 16	GO:0005515
AGING00000008879/ HiC_scaffold_17:6644713-6657044	Puromycin-sensitive aminopeptidase	GO:0006508, GO:0008237, GO:0008270
AGING00000008880/ HiC_scaffold_17:6658768-6660630	Ubiquitin-conjugating enzyme E2 variant 2	
AGING00000008881/ HiC_scaffold_17:6663983-6664783	hypothetical protein	GO:0003677
AGING00000008982/ HiC_scaffold_17:8620129-8758062	Zinc finger homeobox protein 3	GO:0003677, GO:0006355, GO:0003676, GO:0008270, GO:0043565
AGING00000009027/ HiC_scaffold_17:10728619-10785458	Tyrosine-protein kinase Src64B	GO:0005524, GO:0006468, GO:0005515, GO:0004672, GO:0004713
AGING00000009067/ HiC_scaffold_17:12825690-12909714	Knirps-related protein	GO:0006355, GO:0008270, GO:0003700, GO:0005634, GO:0043565
AGING00000009068/ HiC_scaffold_17:12846432-12874434	hypothetical protein	
AGING00000009108/ HiC_scaffold_17:14738354-14755050	Myc protein	GO:0003700, GO:0046983, GO:0006355, GO:0005634
AGING00000009121/ HiC_scaffold_17:15317923-15338282	Longitudinals lacking protein, isoforms A/B/D/L, Longitudinals lacking protein, isoforms H/M/V	GO:0005515, GO:0003676
AGING00000009134/ HiC_scaffold_17:15645000-15702431	Reticulon-4-interacting protein 1, mitochondrial	GO:0055114, GO:0016491
AGING00000009505/ HiC_scaffold_19:308239-326326	TWiK family of potassium channels protein 18	GO:0005267, GO:0016020, GO:0071805
AGING00000010228/ HiC_scaffold_20:7513284-7571588	Protein pellino	GO:0000209, GO:0061630, GO:0008592
AGING00000010235/ HiC_scaffold_20:7655141-7655611	hypothetical protein	
AGING00000010236/ HiC_scaffold_20:7681434-7715335	hypothetical protein	
AGING00000010415/ HiC_scaffold_20:11798078-12124710	hypothetical protein	
AGING00000010529/ HiC_scaffold_21:44774-56105	Membrane-bound alkaline phosphatase	GO:0003824, GO:0016791

Gene ID/Coordinate (Bottom 1% Tajima's D)	Gene product	GO terms
AGING00000010530/ HiC_scaffold_21:57802-61894	Membrane-bound alkaline phosphatase	GO:0003824, GO:0016791
AGING00000010531/ HiC_scaffold_21:71160-80562	Membrane-bound alkaline phosphatase	GO:0003824, GO:0016791
AGING00000010532/ HiC_scaffold_21:83381-88534	Membrane-bound alkaline phosphatase	GO:0003824, GO:0016791
AGING00000010579/ HiC_scaffold_21:2615192-2671173	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	GO:0008138, GO:0016311, GO:0004725
AGING00000010657/ HiC_scaffold_21:4641488-4677784	Protein jim lovell	GO:0003677, GO:0005515
AGING00000010712/ HiC_scaffold_21:6591398-6721509	Histone-lysine N-methyltransferase MECOM	GO:0003676
AGING00000010746/ HiC_scaffold_21:7352953-7421481	Monocarboxylate transporter 10	GO:0055085
AGING00000010783/ HiC_scaffold_21:8541589-8541858	hypothetical protein	
AGING00000010803/ HiC_scaffold_21:9514182-9556195	Protein bowel	GO:0003676
AGING00000010804/ HiC_scaffold_21:9556627-9568396	Regulator of G-protein signaling loco	GO:0005515
AGING00000010830/ HiC_scaffold_21:10958281-11091461	Kinesin-like protein KIF1B	GO:0003777, GO:0007018, GO:0005524, GO:0008017, GO:0005515
AGING00000010898/ HiC_scaffold_21:13421767-13558683	Stress-activated protein kinase JNK	GO:0004707, GO:0004672, GO:0006468, GO:0005524
AGING00000010925/ HiC_scaffold_21:14041340-14072558	Frizzled-10	GO:0005515, GO:0004888, GO:0016021, GO:0007166, GO:0016020
AGING00000011105/ HiC_scaffold_22:6119576-6147117	Muscle segmentation homeobox	GO:0003677, GO:0043565, GO:0006355
AGING00000011153/ HiC_scaffold_22:7930541-8182806	5-hydroxytryptamine receptor 1	GO:0007186, GO:0004930, GO:0016021
AGING00000011410/ HiC_scaffold_23:4611586-4757248	Protein trachealess	GO:0046983, GO:0006355, GO:0005515
AGING00000011555/ HiC_scaffold_23:9295204-9405696	Trithorax group protein osa	GO:0035060, GO:0006338, GO:0016514, GO:0003677
AGING00000011755/ HiC_scaffold_24:2106903-2111426	hypothetical protein	
AGING00000011856/ HiC_scaffold_24:5608324-5659854	Cyclin-dependent kinase inhibitor 1	GO:0007050, GO:0005634, GO:0004861
AGING00000011860/ HiC_scaffold_24:5868551-5892511	Protein WWC2	GO:0005515
AGING00000011988/ HiC_scaffold_24:10482880-11106160	B-box type zinc finger protein ncl-1	GO:0008270, GO:0005515
AGING00000012023/ HiC_scaffold_24:12855053-13346988	Neurobeachin	GO:0005515
AGING00000012079/ HiC_scaffold_25:1051479-1052807	hypothetical protein	GO:0005525, GO:0003924, GO:0005874, GO:0005200, GO:0007017
AGING00000012221/ HiC_scaffold_25:5038841-5048834	hypothetical protein	GO:0003676
AGING00000012324/ HiC_scaffold_25:7416771-7420628	Elongin-C	GO:0006511
AGING00000012325/ HiC_scaffold_25:7421879-7463754	Ubiquitin carboxyl-terminal hydrolase 20	GO:0016579, GO:0004843, GO:0008270, GO:0036459, GO:0006511
AGING00000012326/ HiC_scaffold_25:7422250-7423803	Zinc finger protein PLAG1, hypothetical protein	GO:0006351, GO:0003676
AGING00000012573/ HiC_scaffold_26:3136398-3174826	Serine/threonine-protein kinase MARK1, Serine/threonine-protein kinase MARK2, Serine/threonine-protein kinase par-1	GO:0006468, GO:0004672, GO:0005524
AGING00000012801/ HiC_scaffold_27:788999-1031718	Calsyntenin-1	GO:0007156, GO:0005509, GO:0016020
AGING00000012901/ HiC_scaffold_27:4598535-4621160	Plasma membrane calcium-transporting ATPase 2, hypothetical protein	GO:0016021, GO:0070588, GO:0005388, GO:0005524
AGING00000012956/ HiC_scaffold_27:7172962-7408200	hypothetical protein	
AGING00000014290/ HiC_scaffold_30:9953180-10043114	Group 3 secretory phospholipase A2, hypothetical protein	GO:0006644, GO:0004623, GO:0050482
AGING00000014291/ HiC_scaffold_30:10008831-10016359	hypothetical protein	

References

Adden, A., 2020. There and back again: The neural basis of migration in the Bogong moth (PhD thesis). Lund University, Faculty of Science.

Adden, A., Dreyer, D., Frost, B., Mouritsen, H., Xu, J., Green, K., Whitehouse, M., Chahl, J., Wallace, J., Foster, J., Heinze, S., Warrant, E., 2020a. The starry night sky provides true compass information for long-distance nocturnal navigation in the Australian Bogong moth, in: Adden, A. (Ed.), There and Back Again: The Neural Basis of Migration in the Bogong Moth (PhD Thesis). Lund University, Faculty of Science.

Adden, A., Wibrand, S., Pfeiffer, K., Warrant, E., Heinze, S., 2020b. The brain of a nocturnal migratory insect, the Australian Bogong moth. Journal of Comparative Neurology 528, 1942–1963. https://doi.org/10.1002/cne.24866

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: 2nd International Symposium on Information Theory. Akadémiai Kiadó Location Budapest, Hungary, pp. 267–281.

Akritas, M.G., Bershady, M.A., 1996. Linear regression for astronomical data with measurement errors and intrinsic scatter. The Astrophysical Journal 470, 706–714. https://doi.org/10.1086/177901

Alerstam, T., 1993. Bird migration. Cambridge University Press.

Athanassiou, C.G., Kavallieratos, N.G., Mazomenos, B.E., 2004. Effect of trap type, trap color, trapping location, and pheromone dispenser on captures of male *Palpita unionalis* (Lepidoptera: Pyralidae). Journal of Economic Entomology 97, 321–329. https://doi.org/10.1093/jee/97.2.321

Beck, J., Linsenmair, K.E., others, 2006. Feasibility of light-trapping in community research on moths: Attraction radius of light, completeness of samples, nightly flight times and seasonality of Southeast-Asian hawkmoths (Lepidoptera: Sphingidae). Journal of Research on the Lepidoptera 39, 18–37.

Bennett, G., 1834. Wanderings in New South Wales, Batavia, Pedir Coast, Singapore, and China: Being the journal of a naturalist in those countries during 1832, 1833, and 1834. Richard Bentley.

Bewick, A.J., Vogel, K.J., Moore, A.J., Schmitz, R.J., 2017. Evolution of DNA methylation across insects. Molecular biology and evolution 34, 654–665.

Boisduval, J.A., 1832. Faune entomologique de l'Océan Pacifique: Avec l'illustration des insectes nouveaux recueillis pendant le voyage. J. Tastu.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Brehm, G., 2017. A new LED lamp for the collection of nocturnal lepidoptera and a spectral comparison of light-trapping lamps. Nota lepidopterologica 40, 87.

Breitenstein, M.D., Reichlin, F., Leibe, B., Koller-Meier, E., Van Gool, L., 2009. Robust tracking-by-detection using a detector confidence particle filter, in: 2009 IEEE 12th International Conference on Computer Vision. IEEE, pp. 1515–1522.

Brydegaard, M., Jansson, S., 2019. Advances in entomological laser radar. The Journal of Engineering 2019, 7542–7545. https://doi.org/10.1049/joe.2019.0598

Brydegaard, M., Malmqvist, E., Jansson, S., Larsson, J., Török, S., Zhao, G., 2017. The Scheimpflug lidar method, in: Lidar Remote Sensing for Environmental Monitoring 2017. International Society for Optics; Photonics, p. 104060I. https://doi.org/10.1117/12.2272939

Buehlmann, C., Mangan, M., Graham, P., 2020. Multimodal interactions in insect navigation. Animal Cognition 23, 1129–1141.

Caley, P., Welvaert, M., 2018. Aestivation dynamics of bogong moths (*Agrotis infusa*) in the Australian Alps and predation by wild pigs (*Sus scrofa*). Pacific Conservation Biology 24, 178–182.

Carron, P., Happold, D., Bubela, T., 1990. Diet of two sympatric Australian subalpine rodents, *Mastacomys fuscus* and *Rattus fuscipes*. Australian Wildlife Research 17, 479–489.

Chaerkady, R., Letzen, B., Renuse, S., Sahasrabuddhe, N.A., Kumar, P., All, A.H., Thakor, N.V., Delanghe, B., Gearhart, J.D., Pandey, A., others, 2011. Quantitative temporal proteomic analysis of human embryonic stem cell differentiation into oligodendrocyte progenitor cells. Proteomics 11, 4007–4020.

Chapman, J.W., Drake, V.A., Reynolds, D.R., 2011. Recent insights from radar studies of insect flight. Annual review of entomology 56, 337–356. https://doi.org/10.1146/annurev-ento-120709-144820

Chapman, J.W., Reynolds, D.R., Mouritsen, H., Hill, J.K., Riley, J.R., Sivell, D., Smith, A.D., Woiwod, I.P., 2008. Wind selection and drift compensation optimize migratory pathways in a high-flying moth. Current Biology 18, 514–518.

Chapman, J.W., Reynolds, D.R., Smith, A.D., Riley, J.R., Pedgley, D.E., Woiwod, I.P., 2002. High-altitude migration of the diamondback moth plutella xylostella to the UK: A study using radar, aerial netting, and ground trapping. Ecological Entomology 27, 641–650.

Chapman, J.W., Reynolds, D.R., Wilson, K., 2015. Long-range seasonal migration in insects: Mechanisms, evolutionary drivers and ecological consequences. Ecology letters 18, 287–302.

Chen, Q., Yan, W., Duan, E., 2016. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. Nature Reviews Genetics

17, 733–743.

Cheung, A., Zhang, S., Stricker, C., Srinivasan, M.V., 2007. Animal navigation: The difficulty of moving in a straight line. Biological cybernetics 97, 47–61.

Collett, T.S., Zeil, J., 2018. Insect learning flights and walks. Current Biology 28, R984–R988.

Colvin, S., others, 2021. PyPi.

Common, I., 1954. A study of the ecology of the adult bogong moth, *Agrotis infusa* (Boisd)(Lepidoptera: Noctuidae), with special reference to its behaviour during migration and aestivation. Australian Journal of Zoology 2, 223–263.

Common, I., 1952. Migration and gregarious aestivation in the bogong moth, *Agrotis infusa*. Nature 170, 981–982.

Common, I.F.B., 1990. Moths of australia. E. J. Brill.

Costa-Luis, C. da, Larroque, S.K., Altendorf, K., Mary, H., richardsheridan, Korobov, M., Yorav-Raphael, N., Ivanov, I., Bargull, M., Rodrigues, N., CHEN, G., Lee, A., Newey, C., James, Coales, J., Zugnoni, M., Pagel, M.D., mjstevens777, Dektyarev, M., Rothberg, A., Alexander, Panteleit, D., Dill, F., FichteFoll, Sturm, G., HeoHeo, Kemenade, H. van, McCracken, J., Nordlund, M., Nechaev, N., 2021. tqdm: A fast, extensible progress bar for Python and CLI. Zenodo. https://doi.org/10.5281/zenodo.5109730

Crouse, D.F., 2016. On implementing 2D rectangular assignment algorithms. IEEE Transactions on Aerospace and Electronic Systems 52, 1679–1696.

Czech, L., Exposito-Alonso, M., 2021. Grenepipe: A flexible, scalable, and reproducible pipeline to automate variant and frequency calling from sequence reads. arXiv preprint arXiv:2103.15167.

Davys, R., 2014. Sculpture: *Bogong Moth Migration*. Yindyamarra Sculpture Walk.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., Notredame, C., 2017. Nextflow enables reproducible computational workflows. Nature biotechnology 35, 316–319.

Dingle, H., 2014. Migration: The biology of life on the move. Oxford University Press, USA.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Drake, V.A., Farrow, R.A., 1985. A radar and aerial-trapping study of an early spring migration of moths (Lepidoptera) in inland New South Wales. Australian journal of Ecology 10, 223–223. https://doi.org/10.1111/j.1442-9993.1985.tb00885.x

Drake, V.A., Reynolds, D.R., 2012. Radar entomology: Observing insect flight and migration. Cabi.

Dreyer, D., Frost, B., Mouritsen, H., Günther, A., Green, K., Whitehouse, M., Johnsen, S., Heinze, S., Warrant, E., 2018. The Earth's magnetic field and visual landmarks steer migratory flight behavior in the nocturnal Australian Bogong moth. Current Biology 28, 2160–2166. https://doi.org/10.1016/j.cub.2018.05.030

Dreyer, D., Frost, B., Mouritsen, H., Lefèvre, A., Menz, M., Warrant, E., 2021. A guide for using flight simulators to study the sensory basis of long-distance migration in insects. Frontiers in Behavioral Neuroscience.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., others, 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95.

Dudchenko, O., Shamim, M.S., Batra, S.S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., St Hilaire, B.G., Yao, W., Stamenova, E., others, 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. BioRxiv 254797.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., Aiden, E.L., 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell systems 3, 99–101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., Aiden, E.L., 2016b. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell systems 3, 95–98.

Durieux, G., Liedvogel, M., 2020. Orientation and navigation in the animal world, in: Morton, Y.J., Diggelen, F. van, Spilker Jr, J.J., Parkinson, B.W., Lo, S., Gao, G. (Eds.), Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications. Wiley Online Library, pp. 1689–1709.

Dutta, A., Zisserman, A., 2019. The VIA annotation software for images, audio and video, in: Proceedings of the 27th ACM International Conference on Multimedia, MM '19. ACM, New York, NY, USA. https://doi.org/10.1145/3343031.3350535

Eastwood, E., 1967. Radar ornithology. Methuen.

Egevang, C., Stenhouse, I.J., Phillips, R.A., Petersen, A., Fox, J.W., Silk, J.R., 2010. Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. Proceedings of the National Academy of Sciences 107, 2078–2081.

el Jundi, B., Foster, J.J., Khaldy, L., Byrne, M.J., Dacke, M., Baird, E., 2016. A snapshot-based mechanism for celestial orientation. Current biology 26, 1456–1462.

Evans, J.D., Schwarz, R.S., Chen, Y.P., Budge, G., Cornman, R.S., De la Rua, P., Miranda, J.R. de, Foret, S., Foster, L., Gauthier, L., others, 2013. Standard methods for molecular research in *apis mellifera*. Journal of Apicultural Research 52, 1–54.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Farrow, R.A., McDonald, G., 1987. Migration strategies and outbreaks of noctuid pests in australia. International Journal of Tropical Insect Science 8, 531–542.

Fitak, R.R., Johnsen, S., 2017. Bringing the analysis of animal orientation data full circle: Model-based approaches with maximum likelihood. Journal of experimental biology 220, 3878–3882.

Flood, J.M., 1996. Moth hunters of the Australian Capital Territory. Lewisham, NSW: Clarendon.

Flood, J.M., 1980. The moth hunters: Aboriginal prehistory of the Australian Alps. Australian Institute of Aboriginal Studies Canberra.

Foley, F., 2001. Sculpture: *Winged Harvest*. Australian National University.

Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., Maleszka, R., 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proceedings of the National Academy of Sciences 109, 4968–4973.

Foster, J.J., Smolka, J., Nilsson, D.-E., Dacke, M., 2018. How animals follow the stars. Proceedings of the Royal Society B: Biological Sciences 285, 20172322.

Freas, C.A., Cheng, K., 2022. The basis of navigation across species. Annual review of psychology 73, 217–241.

Friston, K., Tononi, G., Reeke Jr, G., Sporns, O., Edelman, G.M., 1994. Value-dependent selection in the brain: Simulation in a synthetic neural model. Neuroscience 59, 229–243.

Furlanis, E., Scheiffele, P., 2018. Regulation of neuronal differentiation, function, and plasticity by alternative splicing. Annual review of cell and developmental biology 34, 451–469.

Gao, B., Hedlund, J., Reynolds, D.R., Zhai, B., Hu, G., Chapman, J.W., 2020. The "migratory connectivity" concept, and its applicability to insect migrants. Movement Ecology 8, 1–13.

Gau, J., Gemilere, R., subteam), L.-V. (FM, Lynch, J., Gravish, N., Sponberg, S., 2021. Rapid frequency modulation in a resonant system: Aerial perturbation recovery in hawkmoths. Proceedings of the Royal Society B 288, 20210352.

Gauthreaux Jr, S.A., Belser, C.G., 2003. Radar ornithology and biological conservation. The Auk 120, 266–277. https://doi.org/10.1093/auk/120.2.266

Gertz, E.M., Yu, Y.-K., Agarwala, R., Schäffer, A.A., Altschul, S.F., 2006. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC biology 4, 1–14.

Gibson, R.K., Broome, L., Hutchinson, M.F., 2018. Susceptibility to climate change via effects on food resources: The feeding ecology of the endangered mountain pygmy-possum (*Burramys parvus*). Wildlife Research 45, 539–550.

Gillies, S., others, 2007--. Shapely: Manipulation and analysis of geometric objects.

Glastad, K.M., Hunt, B.G., Goodisman, M.A., 2014. Evolutionary insights into DNA methylation in insects. Current Opinion in Insect Science 1, 25–30.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., others, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29, 644–652.

Graham, P., Fauria, K., Collett, T.S., 2003. The influence of beacon-aiming on the routes of wood ants. Journal of Experimental Biology 206, 535–541.

Green, K., 2014. Growing season air temperature lapse rate in the snowy mountains. Australian Meteorological and Oceanographic Journal 64, 289–291.

Green, K., 2011. The transport of nutrients and energy into the Australian Snowy Mountains by migrating Bogong moths *Agrotis infusa*. Austral Ecology 36, 25–34. https://doi.org/10.1111/j.1442-9993.2010.02109.x

Green, K., 2010b. The aestivation sites of bogong moths, *Agrotis infusa* (Boisduval)(Lepidoptera: Noctuidae), in the Snowy Mountains and the projected effects of climate change. Australian Entomologist, The 37, 93–104.

Green, K., 2010a. Alpine taxa exhibit differing responses to climate warming in the Snowy Mountains of Australia. Journal of Mountain Science 7, 167–175.

Green, K., 2008. Migratory Bogong moths (*Agrotis infusa*) transport arsenic and concentrate it to lethal effect by estivating gregariously in alpine regions of the Snowy Mountains of Australia. Arctic, Antarctic, and Alpine Research 40, 74–80.

Green, K., 2003. Altitudinal and temporal differences in the food of foxes (*Vulpes vulpes*) at alpine and subalpine altitudes in the Snowy Mountains. Wildlife Research 30, 245–253.

Green, K., 1989. Altitudinal and seasonal differences in the diets of antechinus-swainsonii and antechinus-stuartii (marsupialia, dasyuridae) in relation to the availability of prey in the snowy mountains. Wildlife Research 16, 581–592.

Green, K., Caley, P., Baker, M., Dreyer, D., Wallace, J., Warrant, E., 2021. Australian Bogong moths *Agrotis infusa* (Lepidoptera: Noctuidae), 1951–2020: Decline and crash. Austral Entomology. https://doi.org/10.1111/aen.12517

Green, K., Osborne, W., 1981. The diet of foxes, *Vulpes vulpes* (L.), In relation to abundance of prey above the winter snowline in New South Wales. Wildlife Research 8, 349–360.

Gregg, P., Fitt, G., Coombs, M., Henderson, G., 1994. Migrating moths collected in tower-mounted light traps in northern new south wales, australia: Influence of local and synoptic weather. Bulletin of Entomological Research 84, 17–30.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology 9, 1–22.

Hao, Z., Drake, V.A., Taylor, J.R., Warrant, E., 2020. Insect target classes discerned from entomological radar data. Remote Sensing 12, 673.

Harris, C.R., Millman, K.J., Walt, S.J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M.H. van, Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hawke, T., Bates, H., Hand, S., Archer, M., Broome, L., 2019. Dietary analysis of an uncharacteristic population of the Mountain Pygmy-possum (*Burramys parvus*) in the Kosciuszko National Park, New South Wales, Australia. PeerJ 7, e6307.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969. https://doi.org/10.1109/iccv.2017.322

Heinze, S., 2017. Unraveling the neural basis of insect navigation. Current opinion in insect science 24, 58–67.

Heinze, S., Warrant, E., 2016. Bogong moths. Current Biology 26, R263–R265.

Holt, C., Yandell, M., 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics 12, 1–14.

Honkanen, A., Adden, A., Silva Freitas, J. da, Heinze, S., 2019. The insect central complex and the neural basis of navigational strategies. Journal of Experimental Biology 222, jeb188854.

Hu, C., Kong, S., Wang, R., Long, T., Fu, X., 2018. Identification of migratory insects from their physical features using a decision-tree support vector machine and its application to radar entomology. Scientific Reports 8, 1–11.

Hu, H., Brittain, G.C., Chang, J.-H., Puebla-Osorio, N., Jin, J., Zal, A., Xiao, Y., Cheng, X., Chang, M., Fu, Y.-X., others, 2013. OTUD7B controls non-canonical NF-κB activation through deubiquitination of TRAF3. Nature 494, 371–374.

Hu, H., Wang, H., Xiao, Y., Jin, J., Chang, J.-H., Zou, Q., Xie, X., Cheng, X., Sun, S.-C., 2016. Otud7b facilitates T cell activation and inflammatory responses by regulating Zap70 ubiquitination. Journal of experimental medicine 213, 399–414.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. IEEE Annals of the History of Computing 9, 90–95. https://doi.org/10.1109/MCSE.2007.55

Hutchinson, F., 1985. Chemical changes induced in DNA by ionizing radiation. Progress in nucleic acid research and molecular biology 32, 115–154.

Hyduke, D.R., Palsson, B.Ø., 2010. Towards genome-scale signalling-network reconstructions. Nature Reviews Genetics 11, 297–307.

Infusino, M., Brehm, G., Di Marco, C., Scalercio, S., 2017. Assessing the efficiency of UV LEDs as light sources for sampling the diversity of macro-moths (Lepidoptera). European Journal of Entomology 114. https://doi.org/10.14411/eje.2017.004

Jackson, T., Wang, H., Nugent, M.J., Griffin, C.T., Burnell, A.M., Dowds, B.C., 1995. Isolation of insect pathogenic bacteria, *Providencia rettgeri*, from *Heterorhabditis* spp. Journal of applied bacteriology 78, 237–244.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., others, 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kaushik, P.K., Olsson, S.B., 2020. Using virtual worlds to understand insect navigation for bio-inspired systems. Current opinion in insect science 42, 97–104.

Keaney, B., others, 2016. Bogong moth aestivation sites as an archive for understanding the Floral, Faunal and Indigenous history of the Northern Australian Alps (PhD thesis). The Australian National University.

Kemenade, H. van, wiredfool, Murray, A., Clark, A., Karpinsky, A., Baranovič, O., Gohlke, C., Dufresne, J., Crowell, B., Schmidt, D., Kopachev, K., Houghton, A., Mani, S., Landey, S., vashek, Ware, J., Douglas, J., Caro, D., Martinez, U., Kossouho, S., Lahd, R., T, S., Lee, A., Brown, E.W., Tonnhofer, O., Bonfill, M., Rowlands, P., Al-Saidi, F., Novikov, G., Górny, M., 2021. python-pillow/Pillow: 8.3.1. Zenodo. https://doi.org/10.5281/zenodo.5076624

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., team, J. development, 2016. Jupyter notebooks - a publishing format for reproducible computational workflows, in: Loizides, F., Scmidt, B. (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, Netherlands, pp. 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

Knox, A., 2012. Sculpture: *Moth Ascending the Capital.* artsACT.

Knuff, A.K., Winiger, N., Klein, A.-M., Segelbacher, G., Staab, M., 2019. Optimizing sampling of flying insects using a modified window trap. Methods in Ecology and Evolution 10, 1820–1825. https://doi.org/10.1111/2041-210X.13258

Koren, S., Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M., 2019. Reply to "errors in long-read assemblies can critically affect protein prediction." Nature biotechnology 37, 127–128.

Korf, I., 2004. Gene finding in novel genomes. BMC bioinformatics 5, 1–9.

Krauel, J.J., Westbrook, J.K., McCracken, G.F., 2015. Weather-driven dynamics in a dual-migrant system: Moths and bats. Journal of Animal Ecology 84, 604–614.

Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., Zdobnov, E.M., 2019. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research 47, D807–D811.

Krueger, F., Andrews, S.R., 2011. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. bioinformatics 27, 1571–1572.

Laurent, P., Frérot, B., 2007. Monitoring of European corn borer with pheromone-baited traps: Review of trapping system basics and remaining problems. Journal of economic entomology 100, 1797–1807. https://doi.org/10.1093/jee/100.6.1797

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11, 204–220.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLoS computational biology 9, e1003118.

Lei, S., He, Z., Chen, T., Guo, X., Zeng, Z., Shen, Y., Jiang, J., 2019. Long noncoding RNA 00976 promotes pancreatic cancer progression through OTUD7B by sponging miR-137 involving EGFR/MAPK pathway. Journal of Experimental & Clinical Cancer Research 38, 1–15.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2117–2125. https://doi.org/10.1109/cvpr.2017.106

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context, in: European Conference on Computer Vision. Springer, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Lohmann, K.J., Lohmann, C.M., 2019. There and back again: Natal homing by magnetic navigation in sea turtles and salmon. Journal of Experimental Biology 222, jeb184077.

Love, P., 2010. Spatial and temporal characteristics of arsenic in the Bogong moth (*Agrotis infusa*) (PhD thesis). La Trobe University.

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., Bensch, S., 2017. Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evolution Letters 1, 155–168.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Kim, T.-K., 2021. Multiple object tracking: A literature review. Artificial Intelligence 293, 103448.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., Maleszka, R., 2010. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. PLoS biology 8, e1000506.

Lyko, F., Maleszka, R., 2011. Insects as innovative models for functional studies of DNA methylation. Trends in Genetics 27, 127–131.

Magrane, M., others, 2011. UniProt Knowledgebase: A hub of integrated protein data. Database 2011.

Maleszka, R., 2016. Epigenetic code and insect behavioural plasticity. Current opinion in insect science 15, 45–52.

Mansergh, I., Heinze, D., others, 2019. Bogong moths "*Agrotis infusa*," soil fertility and food chains in the australian alpine region, and observations concerning the recent population crash of this iconic species. Victorian Naturalist, The 136, 192.

Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770.

Marcel, S., Rodriguez, Y., 2010. Torchvision: The machine-vision package of torch, in: Proceedings of the 18th ACM International Conference on Multimedia. pp. 1485–1488. https://doi.org/10.1145/1873951.1874254

Mardia, K., 1969. On Wheeler and Watson's two-sample test on a circle. Sankhyā: The Indian Journal of Statistics, Series A 177–190.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17, 10–12.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., others, 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20, 1297–1303.

McKinney, W., others, 2010. Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference. Austin, TX, pp. 51–56. https://doi.org/10.25080/Majora-92bf1922-00a

Mendoza, A. de, Poppe, D., Buckberry, S., Pflueger, J., Albertin, C.B., Daish, T., Bertrand, S., Calle-Mustienes, E. de la, Gómez-Skarmeta, J.L., Nery, J.R., others, 2021. The emergence of the brain non-CpG methylation system in vertebrates. Nature ecology & evolution 5, 369–378.

Miklos, G.L.G., Maleszka, R., 2011. Epigenomic communication systems in humans and honey bees: From molecules to behavior. Hormones and Behavior 59, 399–406.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L., Tosatto, S.C., Paladin, L., Raj, S., Richardson, L.J., others, 2021. Pfam: The protein families database in 2021. Nucleic acids research 49, D412–D419.

Mitchell, A.T., Chick, R.R., 2002. A survey of insectivorous bats for the Snowy

Mountains Biodiversity Blitz 12–13 January 2002, in: Green, K. (Ed.), Biodiversity in the Snowy Mountains. Australian Institute of Alpine Studies, Jindabyne, pp. 113–119.

Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., others, 2021. Sustainable data analysis with Snakemake. F1000Research 10.

Monk, N., 2021. Automated light beam surveys: A remote monitoring technique for bogong moths in alpine areas. Falls Creek Resort Management, Falls Creek, Victoria.

Moore, B.R., 1980. A modification of the Rayleigh test for vector data. Biometrika 67, 175–180.

Morin, P., Storey, K.B., 2009. Mammalian hibernation: Differential gene expression and novel application of epigenetic controls. International Journal of Developmental Biology 53, 433–442.

Morton, Y.J., Diggelen, F. van, Spilker Jr, J.J., Parkinson, B.W., Lo, S., Gao, G., 2021. Position, navigation, and timing technologies in the 21st century: Integrated satellite navigation, sensor systems, and civil applications. John Wiley & Sons.

Mouritsen, H., 2018. Long-distance navigation and magnetoreception in migratory animals. Nature 558, 50–59.

Mouritsen, H., Heyers, D., Güntürkün, O., 2016. The neural basis of long-distance navigation in birds. Annual review of physiology 78, 133–154.

Murray, D., Zalucki, M., 1990a. Effect of soil moisture and simulated rainfall on pupal survival and moth emergence of *Helicoverpa punctigera* (Wallengren) and *H. Armigera* (Hübner)(Lepidoptera: Noctuidae). Australian journal of entomology 29, 193–197.

Murray, D., Zalucki, M., 1990b. Survival of *Helicoverpa punctigera* (Wallengren) and *H. Armigera* (Hübner)(Lepidoptera: Noctuidae) pupae submerged in water. Australian journal of entomology 29, 191–192.

Musacchia, X., Barr, R., 1968. Survival of whole-body-irradiated hibernating and active ground squirrels; *Citellus tridecemlineatus*. Radiation research 33, 348–356.

Müssig, C., Schröder, F., Usadel, B., Lisso, J., 2010. Structure and putative function of NFX1-like proteins in plants. Plant Biology 12, 381–394.

Nemmen, R.S., Georganopoulos, M., Guiriec, S., Meyer, E.T., Gehrels, N., Sambruna, R.M., 2012. A universal scaling for the energetics of relativistic jets from black hole systems. Science 338, 1445–1448. https://doi.org/10.1126/science.1227416

O'Donnell, K.J., Meaney, M.J., 2020. Epigenetics, development, and psychopathology. Annual Review of Clinical Psychology 16, 327–350.

Oku, T., 1983. Aestivation and migration in noctuid moths. Diapause and life cycle strategies in insects 219–231.

Pagès, H., Aboyoun, P., Gentleman, R., DebRoy, 2013. Biostrings: Efficient manipulation of biological strings. R package. https://doi.org/10.18129/B9.bioc.Biostrings

Pakhomov, A., Chernetsov, N., 2020. A hierarchy of compass systems in migratory birds.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library, in: Wallach, H., Larochelle, H., Beygelzimer, A., dAlché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035.

Paterson, A.B."Banjo"., 1890. The man from Snowy River. The Bulletin.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., others, 2011. Scikit-learn: Machine learning in python. the Journal of machine Learning research 12, 2825–2830.

Pepper, J., 1932. Observations on a unidirectional flight of army cutworm moths and their possible bearing on aestivation. The Canadian Entomologist 64, 241–242.

Perales, R., Bentley, D., 2009. "Cotranscriptionality": The transcription elongation complex as a nexus for nuclear transactions. Molecular cell 36, 178–191.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology 33, 290–295.

Pina-Martins, F., Silva, D.N., Fino, J., Paulo, O.S., 2017. Structure_threader: An improved method for automation and parallelization of programs STRUCTURE, fastStructure and MavericK on multicore CPU systems. Molecular ecology resources 17, e268–e274.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., Korobeynikov, A., 2020. Using SPAdes de novo assembler. Current protocols in bioinformatics 70, e102.

Puspitasari, A., Cerri, M., Takahashi, A., Yoshida, Y., Hanamura, K., Tinganelli, W., 2021. Hibernation as a tool for radiation protection in space exploration. Life 11, 54.

Raj, A., Stephens, M., Pritchard, J.K., 2014. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics 197, 573–589.

Ran, F., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nature protocols 8, 2281–2308.

Reid, D., 1979. An algorithm for tracking multiple targets. IEEE transactions on Automatic Control 24, 843–854.

Rennie, R., 2017. Sculpture: *Untitled (Bogong Moth)*. National Museum of Australia.

Rhodes, B., 2019. Skyfield: High precision research-grade positions for planets and earth satellites generator.

Riley, J.R., 1989. Remote sensing in entomology. Annual review of entomology 34, 247–271. https://doi.org/10.1146/annurev.en.34.010189.001335

Rittschof, C.C., Bukhari, S.A., Sloofman, L.G., Troy, J.M., Caetano-Anollés, D., Cash-Ahmed, A., Kent, M., Lu, X., Sanogo, Y.O., Weisner, P.A., others, 2014. Neuromolecular responses to social challenge: Common mechanisms across mouse, stickleback fish, and honey bee. Proceedings of the national Academy of Sciences 111, 17929–17934.

Ruan, J., Li, H., 2020. Fast and accurate long-read assembly with wtdbg2. Nature methods 17, 155–158.

Salewski, V., Bairlein, F., Leisler, B., 2000. Recurrence of some palaearctic migrant passerine species in West Africa.

Sánchez-Bayo, F., Wyckhuys, K.A.G., 2019. Worldwide decline of the entomofauna: A review of its drivers. Biological conservation 232, 8–27. https://doi.org/10.1016/j.biocon.2019.01.020

Satterfield, D.A., Sillett, T.S., Chapman, J.W., Altizer, S., Marra, P.P., 2020. Seasonal insect migrations: Massive, influential, and overlooked. Frontiers in Ecology and the Environment 18, 335–344.

Schmid, M.R., Dziedziech, A., Arefin, B., Kienzle, T., Wang, Z., Akhter, M., Berka, J., Theopold, U., 2019. Insect hemolymph coagulation: Kinetics of classically and non-classically secreted clotting factors. Insect biochemistry and molecular biology 109, 63–71.

Schnute, J.T., Groot, K., 1992. Statistical analysis of animal orientation data. Animal behaviour 43, 15–33.

Schwartz, C., Hampton, M., Andrews, M.T., 2013. Seasonal and regional differences in gene expression in the brain of a hibernating mammal. PloS one 8, e58427.

Scott, A.W., 1873. On the *Agrotis vastator*, a species of moth, now infesting the seaboard of New South Wales. Trans. Entomol. Soc. NSW 2, 40–48.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python, in: Proceedings of the 9th Python in Science Conference. Austin, TX, p. 61. https://doi.org/10.25080/Majora-92bf1922-011

Shamoun-Baranes, J., Bouten, W., Loon, E.E. van, 2010. Integrating meteorology into research on migration. Integrative and Comparative Biology 50, 280–292.

Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proceedings of the National Academy of Sciences 111, E5593–E5601.

Silmon de Monerri, N.C., Kim, K., 2014. Pathogens hijack the epigenome: A new twist on host-pathogen interactions. The American journal of pathology 184, 897–911.

Silvester, S., Tanbakuchi, A., Müller, P., Nunez-Iglesias, J., Harfouche, M., Klein, A., McCormick, M., OrganicIrradiation, Rai, A., Ladegaard, A., Lee, A., Smith, T.D., Vaillant, G.A., jackwalker64, Nises, J., rreilink, Kemenade, H. van, Dusold, C., Kohlgrüber, F., Yang, G., Inggs, G., Singleton, J., Schambach, M., Hirsch, M., Komarčević, M., NiklasRosenstein, Hsieh, P.-C., Zulko, Barnes, C., Elliott, A., 2020. imageio/imageio v2.9.0. Zenodo. https://doi.org/10.5281/zenodo.4972048

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sims, S.R., 2008. Influence of soil type and rainfall on pupal survival and adult emergence of the fall armyworm (Lepidoptera: Noctuidae) in southern Florida. Journal of Entomological Science 43, 373–380.

Smit, A., Hubley, R., 2008-2015. RepeatModeler Open-1.0.

Smit, A., Hubley, R., Green, P., 1996-2010. RepeatMasker Open-3.0.

Smith, A., Broome, L., 1992. The effects of season, sex and habitat on the diet of the mountain pygmy-possum (*Burramys parvus*). Wildlife Research 19, 755–767.

Smith, C.D., Edgar, R.C., Yandell, M.D., Smith, D.R., Celniker, S.E., Myers, E.W., Karpen, G.H., 2007. Improved repeat identification and masking in Dipterans. Gene 389, 1–9.

Sporns, O., Almássy, N., Edelman, G.M., 2000. Plasticity in value systems and its role in adaptive behavior. Adaptive Behavior 8, 129–148.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: *Ab initio* prediction of alternative transcripts. Nucleic acids research 34, W435–W439.

Stavenga, D.G., Wallace, J.R., Warrant, E.J., 2020. Bogong moths are well camouflaged by effectively decolourized wing scales. Frontiers in Physiology 11, 95.

Stephenson, B., David, B., Fresløv, J., Arnold, L.J., Delannoy, J.-J., Petchey, F., Urwin, C., Wong, V.N., Fullagar, R., Green, H., others, 2020. 2000 year-old Bogong moth (*Agrotis infusa*) Aboriginal food remains, Australia. Scientific reports 10, 1–10.

Storey, K.B., Storey, J.M., 2010. Metabolic regulation and gene expression during aestivation. Aestivation 25–45.

Strydom, R., Denuelle, A., Srinivasan, M.V., 2016. Bio-inspired principles applied to the guidance, navigation and control of UAS. Aerospace 3, 21.

Su, C.-H., Tarn, W.-Y., others, 2018. Alternative splicing in neurogenesis and brain development. Frontiers in molecular biosciences 5, 12.

Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S., Narukawa, J., Ajimura, M., Jouraku, A., Namiki, N., Shimomura, M., others, 2013. Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect, *Bombyx mori*. G3: Genes, Genomes, Genetics 3, 1481–1492.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Talla, V., Pierce, A.A., Adams, K.L., Man, T.J. de, Nallu, S., Villablanca, F.X., Kronforst, M.R., Roode, J.C. de, 2020. Genomic evidence for gene flow between monarchs with divergent migratory phenotypes and flight performance. Molecular ecology 29, 2567–2582.

Tenger-Trolander, A., Lu, W., Noyes, M., Kronforst, M.R., 2019. Contemporary loss of migration in monarch butterflies. Proceedings of the National Academy of Sciences 116, 14671–14676.

Thieding, T., ArgiesDario, Saunders, J., RKrahl, chbndrhnns, Monzel, R., 2021. exif. PyPi.

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Briefings in bioinformatics 14, 178–192.

Torney, C.J., Hopcraft, J.G.C., Morrison, T.A., Couzin, I.D., Levin, S.A., 2018. From single steps to mass migration: The problem of scale in the movement ecology of the Serengeti wildebeest. Philosophical Transactions of the Royal Society B: Biological Sciences 373, 20170012.

Tsuri, H., 2017. Mural: *Big Fella*. Falls Creek.

Urquhart, F.A., 1987. The monarch butterfly: International traveler. Burnham Incorporated Pub.

Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., 2014. Scikit-image: Image processing in python. PeerJ 2, e453. https://doi.org/10.7717/peerj.453

Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome research 27, 737–746.

Vasimuddin, M., Misra, S., Li, H., Aluru, S., 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems, in: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE, pp. 314–324.

Villanueva, P., Nudel, R., Hoischen, A., Fernández, M.A., Simpson, N.H., Gilissen, C., Reader, R.H., Jara, L., Echeverry, M.M., Francks, C., others, 2015. Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. PLoS genetics 11, e1004925.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der

Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, lhan, Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 10 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in python. Nature methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vries, L. de, Pfeiffer, K., Trebels, B., Adden, A.K., Green, K., Warrant, E., Heinze, S., 2017. Comparison of navigation-related brain regions in migratory versus non-migratory noctuid moths. Frontiers in behavioral neuroscience 11, 158.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., others, 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one 9.

Wallace, J.R.A., 2021a. J-wall/camfi: Version 2.1.3. Zenodo. https://doi.org/10.5281/zenodo.5194496

Wallace, J.R.A., 2021b. A dataset for tracking the activity, abundance, and wingbeat frequency of Bogong moths near Cabramurra, NSW in November 2019 [WWW Document]. https://doi.org/10.5281/zenodo.4950570

Wallace, J.R.A., Reber, T., Beaton, B., Dreyer, D., Warrant, E.J., 2021. Inexpensive monitoring of flying insect activity and abundance using wildlife cameras. bioRxiv.

Wang, B., Jie, Z., Joo, D., Ordureau, A., Liu, P., Gan, W., Guo, J., Zhang, J., North, B.J., Dai, X., others, 2017. TRAF2 and OTUD7B govern a ubiquitin-dependent switch that regulates mTORC2 signalling. Nature 545, 365–369.

Warrant, E., Frost, B., Green, K., Mouritsen, H., Dreyer, D., Adden, A., Brauburger, K., Heinze, S., 2016. The Australian Bogong moth *Agrotis infusa*: A long-distance nocturnal navigator. Frontiers in behavioral neuroscience 10. https://doi.org/10.3389/fnbeh.2016.00077

Warrant, E.J., Whitehouse, M.E.A., Green, K., Wallace, J.R.A., Caley, P., Tomlinson, S., Umbers, K., 2021. *Agrotis infusa*. The IUCN Red List of Threatened Species 2021. https://doi.org/10.2305/IUCN.UK.2021-3.RLTS.T190513532A196183274.en

Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J., Birol, I., 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. GigaScience 4, s13742–015.

Watson, M., Warr, A., 2019. Errors in long-read assemblies can critically affect protein prediction. Nature biotechnology 37, 124–126.

Wedd, L., Kucharski, R., Maleszka, R., 2016. Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honeybee *Apis mellifera*. Epigenetics 11, 1–10.

Wedd, L., Maleszka, R., 2016. DNA methylation and gene regulation in honeybees: From genome-wide analyses to obligatory epialleles. DNA Methyltransferases-

Role and Function 193–211.

Welch, H., 1963. *Amphimermis bogongae* sp. Nov. And *Hexamermis cavicola* sp. Nov. From the Australian bogong moth, *Agrotis infusa* (Boisd.), With a review of the genus *Amphimermis* Kaburaki & Imamura, 1932 (Nematoda: Mermithidae). Parasitology 53, 55–62.

Welsh, L., Maleszka, R., Foret, S., 2017. Detecting rare asymmetrically methylated cytosines and decoding methylation patterns in the honeybee genome. Royal Society open science 4, 170248.

Welzl, E., 1991. Smallest enclosing disks (balls and ellipsoids), in: New Results and New Trends in Computer Science. Springer, pp. 359–370. https://doi.org/10.1007/BFb0038202

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Group, N.G.W., Beukeboom, L.W., Desplan, C., Elsik, C.G., others, 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327, 343–348.

Williams, J., 2003. Painting: *Kicked out of Parliament.* National Museum of Australia.

Williams, J., Harding, M., 2001. Sculpture: *Bogong Moths.* National Capital Authority.

Wiltschko, W., 1983. Compasses used by birds. Comparative Biochemistry and Physiology Part A: Physiology 76, 709–717.

Wintle, B., Tunney, O., Cadenhead, N., Visintin, C., Mackay, E., Crossing, I., Hutton, C., Monk, N., Naccarella, A., Southwell, D., others, 2021. The bogong moth, agrotis infusa: Cultural context, knowledge gaps, conservation and monitoring options. Interim report. NESP Threatened Species Recovery Hub Project 4.2.1.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome biology 20, 1–13.

Xu, J., Jarocha, L.E., Zollitsch, T., Konowalczyk, M., Henbest, K.B., Richert, S., Golesworthy, M.J., Schmidt, J., Déjean, V., Sowood, D.J., others, 2021. Magnetic sensitivity of cryptochrome 4 from a migratory songbird. Nature 594, 535–540.

You, M., Ke, F., You, S., Wu, Z., Liu, Q., He, W., Baxter, S.W., Yuchi, Z., Vasseur, L., Gurr, G.M., others, 2020. Variation among 532 genomes unveils the origin and evolutionary history of a global insect herbivore. Nature communications 11, 1–8.

Zeil, J., 2012. Visual homing: An insect perspective. Current opinion in neurobiology 22, 285–293.

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nature genetics 44, 821–824.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A., 2013.

The MaSuRCA genome assembler. Bioinformatics 29, 2669–2677.