Genetic and Genome Analyses of Native Populations of the Honeybee Pathogen *Nosema ceranae*

By

Melissa Peters

Supervisor

Dr. Nicolas Corradi

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
In partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Institute of Biology

Thèse soumise à la
Faculté des études supérieures et Postdoctorales
Université d'Ottawa
En vue de l'obtention de la maîtrise en science à
L'Institut de biologie d'Ottawa-Carleton

Abstract

Microsporidia are a unique phylum of ubiquitous fungal pathogens that are able to infect a wide variety of hosts, including economically and ecologically important organisms. Recently, global declines of the Western honeybee (*Apis mellifera*) have been associated with infections of the microsporidian pathogen *Nosema ceranae*. This species was originally described in the Asiatic honeybee (*A. cerana*), and its identification in global *A. mellifera* hives could result from a recent host transfer. Recent genome studies have found that global populations of this parasite from *A. mellifera* hives are polyploid and that humans may have fueled their global expansion. In this thesis, I investigate the genetic diversity of *N. ceranae* populations from within their native range (Thailand) and among different hosts (*A. mellifera*, *A. cerana*), putting them in context with other previously sequenced global populations. Using both PCR and genome-based methods, my findings reveal that Thai populations of *N. ceranae* exhibit interesting genetic differences from other global pathogen populations but also have some similarities. Thai *N. ceranae* populations share many single nucleotide polymorphisms (SNPs) with other global populations and appear to be clonal. However, in stark contrast with previous studies, these populations carry many SNPs not found in other global populations of this parasite, indicating that these populations have evolved in their current geographic location for some time. This genome analysis also indicates the potential presence of diploidy within Thai populations of *N. ceranae* and possible host-specific loss of heterozygosity. Overall, my findings begin to reveal interesting patterns of genetic diversity in *N. ceranae* populations that bring us one step closer to understanding the biology and genetics of this important honeybee pathogen.

Résumé

Les microsporidies font partie d'un phylum qui inclue des pathogènes fongiques intracellulaires capables d'infecter une grande variété d'hôtes, y compris des organismes économiquement et écologiquement importants. Récemment, des déclins en population mondiale d'abeille mellifère de l'Ouest (*Apis mellifera*) ont été associés à des infections du pathogène microsporidien *Nosema ceranae*. Cette espèce a été décrite à l'origine chez l'abeille asiatique (*A. cerana*), et son identification dans des ruches d'*A. mellifera* pourrait indiquer un transfert d'hôte récent. Des études génomiques sur *N. ceranae* ont démontré que les populations globales de ce parasite isolé de ruches d'*A. mellifera* sont polyploïdes, et que les humains ont probablement contribué à leur expansion globale. Dans cette thèse de Maîtrise, j'ai 'étudié la diversité génétique des populations de *N. ceranae* isolées dans leur aire de répartition native (Thaïlande) et à partir de différents hôtes (*A. mellifera*, *A. cerana*), en les mettant en contexte avec d'autres populations globales de ce parasites séquencés au préalable par d'autres chercheurs. En utilisant des méthodes de PCR et des analyses de leurs génomes complets, mes résultats révèlent que les populations thaïlandaises de *N. ceranae* possèdent des différences génétiques intéressantes par rapport à d'autres populations mondiales de ce pathogène. En particulier, les populations thaïlandaises de *N. ceranae* partagent de nombreux polymorphismes mononucléotidiques (SNP) avec d'autres populations mondiales et semblent être clonales. Cependant, contrairement aux études précédentes, ces populations portent de nombreux SNP qui ne se retrouvent pas dans d'autres populations globales de ce parasite, ce qui indique que ces populations ont évolué dans leur emplacement géographique actuel depuis un certain temps. Nos analyses génomiques indiquent également la présence potentielle de diploïdie dans les populations thaïlandaises de *N. ceranae* et suggèrent une perte possible d'hétérozygotie spécifique à l'hôte. Dans l'ensemble, mes

découvertes commencent à révéler des tendances intéressantes sur la diversité génétique mondiale de ce parasite, ce qui nous rapproche un peu plus à la compréhension de la biologie et de la génétique de ce pathogène important.

Table of Contents

Acknowledgements

First of all, I would like to thank my supervisor Dr. Nicolas Corradi for the continuing support and advice over the past two years. Thank you for always being available to answer my questions and discuss results and for keeping me motivated and encouraged at the most challenging times. I am most grateful for your caring supervision. I would like to extend my appreciation to my committee members Dr. Linda Bonen and Dr. Alex Wong for the valuable feedback and critical advice throughout my project.

I would also like to thank all the members of the lab with a special thanks to those who have had a hand in my project. To Eric Chen and Nathan Liang, thank you for your respective help with the bioinformatic and molecular aspects of my project. To Stephanie Mathieu, I am especially grateful for your incredible friendship. I greatly appreciate all the assistance with troubleshooting in the lab, your advice as I wrote this thesis, and of course the many uplifting and much needed donut runs.

Additionally, I would like to extend my appreciation to my collaborators. Thank you to Dr. Guntima Suwannapong for providing the infected honeybee samples from Thailand and for the advice on the extraction of spores and DNA. I would also like to thank Adrian Pelin for his guidance as I worked through my genomic analyses.

Thank you to Dr. Corradi, the Natural Science and Engineering Research Council (NSERC), the University of Ottawa, and the Department of Biology for the funding they provided throughout this degree.

Finally, I would like to thank my parents and grandparents for their support throughout my graduate studies, and my friends from both Manitoba and Ontario who have been with me along this journey. This accomplishment would not have been possible without you. Thank you!

List of Figures

List of Tables

Chapter 1 – Introduction

Overview of Microsporidia

Microsporidia are a group of unicellular eukaryotic organisms that are obligate intracellular pathogens and have successfully parasitized host species from arthropod, vertebrate, and some protist lineages (Keeling and Fast, 2002; Vávra and Lukeš, 2013; Corradi, 2015). With approximately 1,500 species of Microsporidia currently described (Vávra and Lukeš, 2013), there are many species that impact different aspects of human life. Some microsporidians can cause direct infections in humans such as *Encephalitozoon intestinalis* (Dowd *et al.*, 1998) and *Enterocytozoon bieneusi* (Dowd *et al.*, 1998; Khanduja *et al.*, 2017), but these are mainly associated with immunosuppressed individuals. Other microsporidian species affect humans by shaping the economy and impacting ecosystem stability. Some of these species include the shrimp pathogen *E. hepatopenaei* (Rajendran *et al.*, 2016) and the silkworm pathogen *Nosema bombycis* (Nägeli, 1857; Gupta *et al.*, 2016), where microsporidiosis - *i.e.* the disease associated with microsporidian infections - has had detrimental effects on the farmed shrimp and silk industries respectively. Recently, the honeybee pathogen *N. ceranae* was suggested to be involved in the global decline in honeybee populations which has impacted pollination and honey production worldwide (Fries *et al.*, 1996; Higes *et al.*, 2008; Dussaubat *et al.*, 2012).

There are many ways in which Microsporidia distinguish themselves from other intracellular pathogens. Notably, they harbour a unique specialized structure called the polar tube that functions as the host invasion apparatus (Franzen, 2004; Xu and Weiss, 2005; Vávra and Larsson, 2014). Other unique features of this phylum include the presence of a genome-less mitochondrial remnant called the mitosome that primarily functions in iron-sulfur biochemistry

(Katinka *et al.*, 2001; Williams *et al.*, 2002; Vávra and Larsson, 2014), an unstacked Golgi apparatus (Vávra and Lukeš, 2013; Vávra and Larsson, 2014), and, in some species, a diplokaryon - *i.e.* binucleate - nucleus (Cavalier-Smith, 1998; Lee and Ironside, 2014; Vossbrinck *et al.*, 2014). As an adaptation to intracellular parasitism, most microsporidian species have compacted genomes (Corradi *et al.*, 2010; Keeling *et al.*, 2014). These genomes have reduced gene content and intergenic regions compared to related phyla, resulting in a severe reduction and loss of many metabolic and biochemical pathways that are otherwise conserved in eukaryotes - *e.g.* amino acid and nucleotide *de novo* synthesis (Corradi and Slamovits, 2011; Corradi and Selman, 2013; Keeling *et al.*, 2014; Corradi, 2015). Although microsporidian species may appear primitive, their simplistic cells are the result of reductive evolution from more complex ancestral relatives (Keeling and Fast, 2002; Corradi and Keeling, 2009).

Microsporidia as a Phylum

An Obscure Taxonomic Past

Microsporidia are currently classified in the kingdom Fungi, but the taxonomic history of this group has been plagued with misclassifications and revisions. At its original discovery in 1857, the Microsporidia were included in a group of yeast-like fungi called the Schizomycetes (Nägeli, 1857), which were later found to be a combination of unrelated spore-forming organisms (Corradi and Keeling, 2009; Keeling, 2009). The Microsporidia were promptly reclassified as Protozoa (Sporozoa: Cnidosporidia) (Balbiani, 1882) until a century later when this group was also found to contain unrelated organisms - *i.e.* Sporozoans were similar only in their parasitic nature. From here, Microsporidia were moved into the ancient lineage Archezoa as they seemed to lack conventional mitochondria (Stewart and Mattox, 1980; Cavalier-Smith,

1983). Most recently, the Microsporidia were once again reclassified as Fungi after the discovery of a mitochondrial remnant called the mitosome (Cavalier-Smith, 1998; Williams *et al.*, 2002), publication of the first full microsporidian genome (Katinka *et al.*, 2001), and the addition of molecular and phylogenetic data (Edlund *et al.*, 1996; Keeling and Doolittle, 1996; Fast *et al.*, 1999; Hirt *et al.*, 1999; Vossbrinck and Debrunner-Vossbrinck, 2005; Keeling, 2014).

Evidence linking Microsporidia and Fungi continues to grow, but the intricacies of this relationship are still highly disputed (Lee *et al.*, 2008; Keeling, 2009). It is debated whether Microsporidia form a unique fungal lineage (Lee *et al.*, 2008, 2010) or if they fall as a sister group to Fungi (Tanabe *et al.*, 2002). Despite this disagreement, most well-supported phylogenetic trees produced using protein-coding genes show that Microsporidia are related to Fungi (Cavalier-Smith, 1998; Lee *et al.*, 2008, 2010; Keeling, 2014). Additionally, Microsporidia have recently been shown to be closely related to the phylum Cryptomycota - *i.e.* Rozellida - an elusive group of ubiquitous parasitic organisms found at the base of the Fungal Kingdom (James *et al.*, 2013; Keeling, 2014).

A Unique Life Cycle

Microsporidia differ from other intracellular pathogens in the mechanisms they use to infect their hosts. The life cycle begins when spores are taken up from the environment through oral ingestion and travel to specific host tissues where the microsporidian can undergo its infection cycle (Keeling and McFadden, 1998; Vávra and Lukeš, 2013; Solter, 2014). Spore germination is triggered by unknown stimuli but likely includes shifts in the gut environment such as pH, ionic concentration, or osmotic conditions (Keeling and Fast, 2002; Xu and Weiss, 2005; Vávra and Lukeš, 2013). At germination, the microsporidian spore begins to take water into an organelle called the posterior vacuole through aquaporins which rapidly increases the

osmotic pressure within the spore. This causes the release of the polar tube, a specialized microsporidian invasion apparatus originally tightly coiled inside the spore (Keeling and Fast, 2002; Franzen, 2004; Xu and Weiss, 2005). When ejecting, the polar tube everts as if reversing the finger of a glove, can be several hundred micrometers long (Xu and Weiss, 2005) and can reach speeds of greater than 100 μm/s (Keeling and Fast, 2002). The polar tube can puncture any nearby host gut epithelial cell, and the entire spore contents is then pushed through the polar tube and into the cytoplasm of the host as the posterior vacuole continues to fill with water (Keeling and McFadden, 1998; Keeling and Fast, 2002; Franzen, 2004; Vávra and Lukeš, 2013).

Once inside the host cytoplasm, the microsporidian pathogen can begin the proliferative stage of its life cycle. It first undergoes a process called "merogony" which is a fast asexual reproduction cycle used to rapidly increase the number of pathogens within a single host cell (Keeling and Fast, 2002; Franzen, 2004; Vávra and Lukeš, 2013; Becnel and Andreadis, 2014; Vávra and Larsson, 2014). In this stage, the microsporidian pathogens become surrounded by host organelles, such as endoplasmic reticulum and mitochondria (Keeling and Fast, 2002; Vávra and Lukeš, 2013). This allows Microsporidia to harvest ATP from the host using specialized ADP/ATP transporters (Tsaousis *et al.*, 2008) putting a higher energetic demand on the host and giving the pathogens the energy they need to undergo rapid merogonal divisions (Keeling and Fast, 2002; Vivarès *et al.*, 2002). Without host ATP, these pathogens would not have the energy they need because they have lost many required biochemical and metabolic pathways (Corradi and Slamovits, 2011; Keeling *et al.*, 2014; Corradi, 2015).

Following merogony, the deposition of an electron-dense material to the plasma membrane of the pathogen indicates the initiation of a division process called sporogony (Keeling and Fast, 2002; Vávra and Larsson, 2014). Pathogens undergoing sporogony are called

sporonts which divide to form sporoblasts and mature to produce full microsporidian spores prepared for release from the host cell (Keeling and Fast, 2002; Franzen, 2004; Vávra and Lukeš, 2013; Becnel and Andreadis, 2014; Vávra and Larsson, 2014). Spores are released through host cell lysis and can re-infect the same host individual through autoinfection or be released into the environment through fecal excrement or host death. Once released, these thick-walled spores can survive in the environment until they are taken up by a new host individual where they can then repeat their life cycle (Keeling and Fast, 2002; Becnel and Andreadis, 2014; Vávra and Larsson, 2014).

While Microsporidia have been known to be asexual organisms (Lee and Ironside, 2014), there is some evidence that they may undergo a cryptic sexual reproduction cycle. The discovery of several meiosis genes, such as *Spo11* and *RecA*, in many species indicate that Microsporidia are theoretically capable of undergoing meiosis (Lee *et al.*, 2008, 2010; Cuomo *et al.*, 2012). Some studies have also found evidence for recombination in both rDNA and single-copy genes (Sagastume *et al.*, 2011, 2016; Ironside, 2013). Recombination can occur during mitosis initiated by DNA repair processes (LaFave and Sekelsky, 2009) or can occur during meiosis which would indicate sexual reproduction. Additionally, many microsporidian species harbour a sex-related locus similar to that found in Zygomycete fungi. The gene order in this region is conserved in Microsporidia (Lee *et al.*, 2008, 2010; Cuomo *et al.*, 2012; Lee and Ironside, 2014), but these genes differ from Zygomycetes in that they are convergently transcribed and maintain no synteny with flanking genes (Lee *et al.*, 2008). This suggests that these pathogens may undergo a cryptic sexual reproduction cycle, but there is currently no direct evidence showing sexual reproduction in Microsporidia.

The Microsporidian Genome

The field of genomics has recently become a popular way of studying microorganisms, especially pathogens, as genomic data can supply information on pathogen transmission, virulence, adaptability, and provides a snapshot of their general biology. The first microsporidian genome sequenced was that of *E. cuniculi* by Katinka *et al.* in 2001. Since this study, many microsporidian species have been sequenced, including additional *Encephalitozoon* spp. (Corradi *et al.*, 2010; Pombert *et al.*, 2012), *N. ceranae* (Cornman *et al.*, 2009; Pelin *et al.*, 2015), *Nematocida* spp. (Cuomo *et al.*, 2012), *Spraguea lophii* (Campbell *et al.*, 2013; Williams *et al.*, 2016), and *Pseudoloma neurophila* (Ndikumana *et al.*, 2017), among others. As more microsporidian genome data has been obtained, more has been discovered about their evolution and genome structure.

Many microsporidian genomes are highly reduced because of their intracellular way of life, but different species vary greatly in their level of genome compaction. The smallest microsporidian genome known to date is that of *E. intestinalis* at 2.3 Mb (Corradi *et al.*, 2010; Keeling *et al.*, 2014). This is in contrast with the largest microsporidian genome known, that of *Edhazardia aedis*, with a size of 51.3 Mb (Desjardins *et al.*, 2015). Interestingly, species with larger genomes do not typically exhibit a comparable increase in gene content but instead contain more repeated elements, introns, and intergenic regions. This results in smaller genomes having higher gene densities (Keeling, 2009; Corradi and Slamovits, 2011; Corradi and Selman, 2013). Some species such as *Antonospora locustae* (Williams *et al.*, 2005; Corradi *et al.*, 2008) and *E. cuniculi* (Corradi *et al.*, 2008) have genomes that are so compacted that neighbouring genes are transcribed together resulting in a single mRNA molecule containing the genetic material for multiple polypeptides (Keeling, 2009). Interestingly, many genes are unique to Microsporidia

and have no known homologs in organisms outside of the phylum which indicates these genes may be involved in pathogen virulence (Nakjang *et al.*, 2013).

Despite genome size, microsporidians have lost many genes required for a free-living organism resulting in loss of functioning essential biochemical pathways such as the tricarboxylic acids cycle and biosynthesis of nucleic acids (Keeling and Fast, 2002; Corradi and Slamovits, 2011; Corradi, 2015). Loss of these pathways means that microsporidians require host machinery and metabolites to live (Tsaousis *et al.*, 2008; Selman and Corradi, 2011; Cuomo *et al.*, 2012). These pathogens have obtained mechanisms and transporters to steal these requirements, such as ATP and nucleotides, from their host. There is evidence that some of these transporters, as well as other genes, were acquired through horizontal gene transfers from bacteria, other prokaryotes, and in some cases animal hosts (Lee *et al.*, 2009; Selman and Corradi, 2011; Cuomo *et al.*, 2012; Pombert *et al.*, 2012).

Many microsporidians have very diverging genome sequences but otherwise harbour a highly conserved gene order across species and with the Zygomycete fungi (Slamovits *et al.*, 2004; Corradi *et al.*, 2007; Vávra and Lukeš, 2013). Additionally, the level of heterozygosity found within the genome differs greatly between species. Some species such as *N. ceranae* (Pelin *et al.*, 2015; Sagastume *et al.*, 2016) and *N. parisii* (Cuomo *et al.*, 2012) have extensive amounts of heterozygosity, whereas other species such as *E. cuniculi* (Selman *et al.*, 2013) have very reduced levels. Differences in genome heterozygosity could indicate reproduction mode and are also connected to level of ploidy in these organisms. Ploidy has been studied through patterns of heterozygosity in many microsporidian species, including *E. cuniculi* (Selman *et al.*, 2013; Pelin *et al.*, 2016), *Nematocida* spp. (Cuomo *et al.*, 2012), *E. aedis* (Desjardins *et al.*, 2015), *P. neurophila* (Ndikumana *et al.*, 2017), and *Trachipleistophora hominis* (Watson *et al.*, 2015)

which all show clear diploid patterns. In contrast, studies of *N. ceranae* show equally clear patterns of tetraploidy (Pelin *et al.*, 2015).

The Honeybee Pathogen *Nosema ceranae*

Ecological and Economic Importance of Honeybees

Honeybees are important arthropods that have an impact on both the environment and the economy. Both the Asian honeybee, *Apis cerana*, and the Western honeybee, *A. mellifera*, play a vital role in the pollination of native and agricultural plants, helping to maintain natural ecosystems and increase crop yield (Kearns *et al.*, 1998; Suwannapong, Benbow, *et al.*, 2011). Pollinators contribute billions of dollars to economies worldwide through crop pollination (Gallai *et al.*, 2009), and honeybees specifically contribute to the economy through the commercialization of their products. The most popular product is honey, but royal jelly, beeswax, pollen, propolis, and even venom are also harvested for commercial use (Delaplane and Mayer, 2000; Suwannapong, Benbow, *et al.*, 2011). In addition, honeybees are managed and transported around the globe for use in pollination or honey production (Kearns *et al.*, 1998; Delaplane and Mayer, 2000).

Due to the importance of honeybee species, their recent global decline, which has been labelled Colony Collapse Disorder (CCD), has caused a grave concern worldwide. CCD is a threat to ecosystem stability and for beekeeping and agricultural production (Chen *et al.*, 2009; VanEngelsdorp *et al.*, 2009). The cause of these colony losses is not clear and is thought to be due to multiple factors such as changes in climate, habitat destruction, increased pesticide use, as well as diseases and pathogens (VanEngelsdorp *et al.*, 2009; VanEngelsdorp and Meixner, 2010; Pettis *et al.*, 2012). Of these factors, pathogens play a large role in CCD. Some examples include

bee mites such as *Varroa destructor* (Anderson and Trueman, 2000; Rosenkranz *et al.*, 2010; VanEngelsdorp and Meixner, 2010; Suwannapong, Benbow, *et al.*, 2011), viruses such as Kashmir bee virus (de Miranda *et al.*, 2010; VanEngelsdorp and Meixner, 2010; Suwannapong, Benbow, *et al.*, 2011), and bacterial pathogens such as *Paenibacillus larvae* (Nakamura, 1996; VanEngelsdorp and Meixner, 2010; Suwannapong, Benbow, *et al.*, 2011). Additionally, honeybee declines have been associated with infections by the fungal microsporidian pathogen *N. ceranae* (Martín-Hernández *et al.*, 2007; Higes *et al.*, 2008; VanEngelsdorp *et al.*, 2009; Dussaubat *et al.*, 2012) which is now found in hives worldwide (Higes *et al.*, 2006; Chen *et al.*, 2008; Williams *et al.*, 2008; Chen and Huang, 2010; Suwannapong, Yemor, *et al.*, 2011; Medici *et al.*, 2012; Bollan *et al.*, 2013). It has been experimentally shown that *N. ceranae* infection increases the chance of colony depopulation by up to six times (Martín-Hernández *et al.*, 2007; Higes *et al.*, 2008).

The Genus Nosema

Nosema is a genus of Microsporidia that infects a variety of arthropod species (Solter, 2014). Some of the most well-known pathogens within this genus are those that infect honeybee hosts, specifically *N. apis* and *N. ceranae*. *Nosema apis* was first discovered to cause infection in the Western honeybee in 1909 (Zander, 1909) followed by the detection of *N. ceranae* in the Asian honeybee many years later in 1996 (Fries *et al.*, 1996). Spores from both species are highly resistant to external environmental factors (Ptaszyńska *et al.*, 2012) and when taken up by a honeybee, can infect multiple host tissues. The most commonly infected tissue for both *Nosema* species is the epithelial cells of the midgut, also known as the ventriculus (Fries *et al.*, 1996; Suwannapong, Yemor, *et al.*, 2011), but *N. ceranae* can spread to other tissues such as the Malpighian tubules, salivary glands, and fat bodies (Chen *et al.*, 2009). These two species of

Nosema are visually very similar and are difficult to differentiate on morphology alone (Solter, 2014), although some differences do exist. *Nosema ceranae* spores are smaller than spores of *N. apis* and contain fewer and shorter polar tube coils (Higes *et al.*, 2006; Suwannapong, Benbow, *et al.*, 2011).

Surprisingly, honeybees infected with *N. ceranae* do not show as many external disease signs as is usually associated with nosemosis caused by *N. apis*. Unlike *N. apis* infections, honeybees infected with *N. ceranae* do not show signs of dysentery, crawling behaviour, or a milky white gut (Fries *et al.*, 2006; Higes *et al.*, 2008). Instead, *N. ceranae* infections cause digestive disorders, shortened lifespan, decreased population sizes, and ultimately reduced honey production and pollinated crop yield (Fries *et al.*, 2006; Ptaszyńska *et al.*, 2012; Botías *et al.*, 2013). Microsporidian infections also make hives more susceptible to other pathogens which causes heightened population and colony loss (Dussaubat *et al.*, 2012). Many compounds have been tested to combat *N. ceranae* infections, but unfortunately the majority have been found to be ineffective. Only fumagillin is able to reduce levels of *N. ceranae* within a hive, but this treatment does not prevent pathogen reinfection (Higes *et al.*, 2008; Botías *et al.*, 2013).

Nosema ceranae is most pathogenic in experimental infections of honeybee hives (Higes *et al.*, 2007; Paxton *et al.*, 2007; Botías *et al.*, 2013) and is able to infect not only its original host species, *A. cerana*, but also *A. mellifera*, its recently acquired host (Higes *et al.*, 2006; Klee *et al.*, 2007; Paxton *et al.*, 2007; Chen *et al.*, 2008). This hypothesized recent host transfer was made possible because of the host range overlap of *A. cerana* and *A. mellifera* caused by human intervention related to the global transport of the Western honeybee for commercial use (Klee *et al.*, 2007; Paxton *et al.*, 2007; Shafer *et al.*, 2009; Suwannapong, Benbow, *et al.*, 2011). This is not the first time a pathogen has been able to jump hosts in honeybees. A similar situation

occurred with *V. destructor* (previously *jacobsoni*) (Anderson and Trueman, 2000), a parasitic mite of honeybees, where overlaps in host range allowed infection of *A. mellifera* after previously isolated infection of *A. cerana* (Rosenkranz *et al.*, 2010; Beaurepaire *et al.*, 2015). Host transfer to the Western honeybee has subsequently allowed the transport and global expansion of both *V. destructor* and *N. ceranae* which are now found in hives around the world (Klee *et al.*, 2007; Paxton *et al.*, 2007).

Nosema ceranae in the Age of Genomics Research

The first full genome assembly of *N. ceranae* populations infecting an *A. mellifera* hive was produced by Cornman *et al.* in 2009 using pyrosequencing technology. The assembly revealed a genome that is 7.9 Mb in size with 74% AT nucleotides and contains many repetitive elements. Overall, 2614 open reading frames were identified, and out of these genes the *N. ceranae* genome was found to share nearly half with another previously sequenced microsporidian species *E. cuniculi*. Within the *N. ceranae* genome, 11 genes were detected that are specific to Microsporidia, with no homologs found outside of the phylum, possibly representing virulence genes (Cornman *et al.*, 2009). This indicates that *N. ceranae* has been able to accumulate novel genes throughout its evolution despite the extensive reduction found in microsporidian genomes (Cornman *et al.*, 2009; Nakjang *et al.*, 2013).

An additional eight full genome sequences of *N. ceranae* populations infecting *A. mellifera* hives from diverse geographic locations were obtained in a study by Pelin *et al.* in 2015 using Illumina sequencing technology. This study revealed that the actual genome size is slightly smaller at 5.7 Mb and gave insight on both the biology and genetics of *N. ceranae*. Lack of evidence of genetic exchange and high levels of linkage disequilibrium indicate that this pathogen likely follows a clonal mode of propagation, and visualizing allele frequency plots

shows evidence of polyploidy in this species. In addition to these findings, all eight sequenced isolates had a high amount of heterozygosity and genetic diversity (Pelin *et al.*, 2015), which has been detected previously in studies of the small ribosomal subunit (SSU) gene of *N. ceranae* (Fries *et al.*, 1996; Huang *et al.*, 2007, 2008, Sagastume *et al.*, 2011, 2014, 2016; Suwannapong, Yemor, *et al.*, 2011; Medici *et al.*, 2012; Roudel *et al.*, 2013). Surprisingly, Pelin *et al.* (2015) found that 70% of all single nucleotide polymorphisms (SNPs) detected are shared among all populations and very few SNPs are unique to a single isolate. Since diversity within these isolates is widely uniform, it has been hypothesized that these populations may not have had time to evolve separately and that modern human management of honeybees is likely a cause of the global expansion of *N. ceranae* (Pelin *et al.*, 2015).

Research Justification and Goals

The biology and genetics of *N. ceranae* have been highly studied in recent years due to its impact on honeybee populations and its possible association with global CCD. Although many studies have been published, most have focused on pathogen populations isolated from hives of the acquired host, *A. mellifera*. Additionally, these studies have mainly looked at transported hives located far from the suggested original range of *N. ceranae*. It is important to study populations of *N. ceranae* from the original host species, *A. cerana*, within the suggested original range, South-East Asia. Doing so will help us gain a better understanding of the global population dynamics, evolution, and spread of this pathogen.

The aim of this thesis is to obtain new information on the genetic diversity and biology of native *N. ceranae* populations by comparing published data in genetics and genomics from global isolates to new data obtained from *N. ceranae* populations collected within its native

range. To accomplish this, both PCR and genome-based methods were utilized to study the level of diversity found in the SSU gene and the full genome of *N. ceranae* populations isolated from both *A. cerana* and *A. mellifera* hives located throughout the range of Thailand. Determining if differences are present in the genome of pathogen populations from native and other global regions as well as between host species will help illuminate *N. ceranae* population dynamics and provide insight into its hypothesized host transfer and global expansion.

Chapter 2 – Genetic and genome analyses of *Nosema ceranae* populations isolated from Thailand

Melissa Peters[1], Guntima Suwannapong[2], Adrian Pelin[3,4], Nicolas Corradi[1]*

[1] *Department of Biology, University of Ottawa, Ottawa, ON, Canada*

[2] *Department of Biology, Burapha University, Chon Buri 20131, Thailand*

[3] *Department of Biochemistry, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada*

[4] *Ottawa Hospital Research Institute, Center for Cancer Therapeutics, Ottawa, ON, Canada*

* To whom correspondence should be addressed: ncorradi@uottawa.ca

Data deposition: Nucleotide sequence data are available in the Genbank database under the accession numbers MH349761-MH349772, MH349776-MH350028, MH350034-MH350091, MH350093-MH350186, MH350257-MH350385, MH350396-MH350400 and MH356491-MH356536. Next-Generation sequence reads are available in the National Center for Biotechnology Information, USA, sequence read archive (SRA) database under the accession number SRP148138 as part of the BioProject PRJNA471856.

Comments:
A version of this work has been submitted for publication in *Microbial Ecology*. My contributions to this work are all aspects of this study. Contributions from other authors:

- GS provided infected honeybee samples and provided extraction protocols.
- AP helped with the genomic analysis pipeline.

Abstract

The recent global decline in Western honeybee (*Apis mellifera)* populations is of great concern for pollination and honey production worldwide. Declining honeybee populations are frequently infected by the microsporidian pathogen *Nosema ceranae*. This species was originally described in the Asiatic honeybee (*A. cerana*), and its identification in global *A. mellifera* hives could result from a recent host transfer. Recent genome studies have found that global populations of this parasite are polyploid and that humans may have fueled their global expansion. To better understand *N. ceranae* biology, we investigated its genetic diversity within part of their native range (Thailand) and among different hosts (*A. mellifera, A. cerana*) using both PCR and genome-based methods. We find that Thai *N. ceranae* populations share many SNPs with other global populations and appear to be clonal. However, in stark contrast with previous studies, we found that these populations carry many SNPs not found in global populations of this parasite, indicating that these populations have evolved in their current geographic location for some time. Our genome analyses also indicate the potential presence of diploidy within Thai populations of *N. ceranae*.

Keywords

Microsporidia, *Nosema ceranae*, honeybees, small ribosomal subunit RNA gene, Next-Generation sequencing, genome diversity

Introduction

Since the classification of the first microsporidian species (Nägeli, 1857), this group of organisms have been a source of mystery and intrigue to researchers. The phylum Microsporidia encompasses unicellular eukaryotic fungi that are obligate intracellular pathogens frequently observed in both vertebrate and invertebrate hosts (Keeling and Fast, 2002; Vávra and Lukeš, 2013; Corradi, 2015). Microsporidia have several common defining features, most notably the presence of genome-less mitochondrial remnants called mitosomes (Williams *et al.*, 2002; Vávra and Larsson, 2014), reduced cellular and metabolic pathways resulting from their obligate intracellular life cycle (Corradi and Slamovits, 2011), and a specialized invasion apparatus called the polar tube used to pierce and invade the host cells (Franzen, 2004; Xu and Weiss, 2005; Vávra and Larsson, 2014). The phylum Microsporidia has undergone extensive reclassifications since its discovery, but most recent genome analyses have shown that Microsporidia are closely related to Rozellids, a lineage that sits at the base of the Fungal Kingdom (James *et al.*, 2013; Quandt *et al.*, 2017).

Microsporidian species impact human epidemiology, the economy, and ecosystem stability as they infect host organisms important to each of these facets. For example, the human pathogen *Encephalitozoon intestinalis* is mainly associated with immunosuppressed individuals (Dowd *et al.*, 1998; Corradi *et al.*, 2010), while the silkworm pathogen *Nosema bombycis* had detrimental effects on the silk industry (Nägeli, 1857; Gupta *et al.*, 2016). More recently, the honeybee pathogen *N. ceranae* has been associated with the global decline in honeybee populations (Fries *et al.*, 1996; Higes *et al.*, 2008, 2013; Dussaubat *et al.*, 2012).

Honeybees are environmentally and economically important arthropods that play a key role in the pollination of wild and agricultural plants (Kearns *et al.*, 1998; Suwannapong,

Benbow, *et al.*, 2011), as well as in the production of goods sold commercially around the world (Suwannapong, Benbow, *et al.*, 2011; Higes *et al.*, 2013). The recent global decline in their populations has been labelled Colony Collapse Disorder (CCD) with the cause of these losses including habitat destruction, human management, increased pesticide use, as well as diseases and pathogens (VanEngelsdorp *et al.*, 2009; VanEngelsdorp and Meixner, 2010; Suwannapong, Benbow, *et al.*, 2011). *Nosema ceranae* is a microsporidian pathogen of honeybees that has been proposed to be associated with CCD (Higes *et al.*, 2008, 2013), but the extent of its role is controversial (Dussaubat *et al.*, 2012). *Nosema ceranae* infections were first detected in hives of the Asiatic honeybee, *Apis cerana* (Fries *et al.*, 1996) but have more recently been found in hives of the Western honeybee, *A. mellifera*, from around the world. The global presence of *N. ceranae* in *A. mellifera* hives was proposed to result from a host transfer, allowing the expansion of this pathogen due to the widespread commerce of the Western honeybee and its products (Higes *et al.*, 2006; Huang *et al.*, 2007; Klee *et al.*, 2007; Martín-Hernández *et al.*, 2007; Williams *et al.*, 2008; Chen and Huang, 2010; Suwannapong, Benbow, *et al.*, 2011; Botías *et al.*, 2012; Gómez-Moracho *et al.*, 2015). Honeybees infected with *N. ceranae* exhibit few outward symptoms, but these infections can cause digestive disorders, shortened lifespan, smaller colony sizes, along with reduced honey production and pollinated crop yield (Higes *et al.*, 2008; Suwannapong, Yemor, *et al.*, 2011; Dussaubat *et al.*, 2012).

Previous genetic studies of *N. ceranae* based on small ribosomal subunit (SSU) gene sequences have found an elevated amount of parasite nucleotide diversity in infected *A. mellifera* hives located in Europe (Huang *et al.*, 2008; Sagastume *et al.*, 2011, 2014, 2016; Roudel *et al.*, 2013), Australia (Sagastume *et al.*, 2016), South America (Huang *et al.*, 2008; Medici *et al.*, 2012), and Asia (Fries *et al.*, 1996; Huang *et al.*, 2007, 2008; Sagastume *et al.*, 2011;

Suwannapong, Yemor, *et al.*, 2011; Roudel *et al.*, 2013). *Nosema ceranae* populations within a single host individual will often show as much genetic variability as is seen between hives (Higes *et al.*, 2013). This high global genetic diversity, combined with evidence of recombination along the SSU fragments, was proposed to result from the presence of multiple *N. ceranae* strains within hives and the presence of cryptic sexual reproduction (Sagastume *et al.*, 2011, 2014, 2016; Higes *et al.*, 2013; Roudel *et al.*, 2013; Gómez-Moracho *et al.*, 2015).

The full genome of *N. ceranae* was first sequenced in 2009 and revealed an 7.9 Mb AT-rich, fairly repetitive genome (Cornman *et al.*, 2009). A more recent full genome analysis of several pathogen populations from global *A. mellifera* hives provided further insight on this parasite's biology by showing that populations of *N. ceranae* around the globe harbour an extensive amount of nucleotide diversity but are not genetically distinct from each other. Furthermore, these genome analyses revealed that *N. ceranae* is both polyploid and likely clonal (Pelin *et al.*, 2015).

Surprisingly, despite the high volume of studies published on *N. ceranae*, very few have focused on populations infecting its original host, *A. cerana*, and on those situated in its suggested original host range, South-East Asia. Comparing the genetic diversity of native populations to that of other global populations of this parasite could provide important insight on the evolution and genetics of this pathogen, and its recent global spread. In the present study, we aim to better understand the biology and genetics of *N. ceranae* by investigating both SSU gene sequences and genome data from *N. ceranae* populations infecting both *A. cerana* and *A. mellifera* hives located across the range of Thailand. The specific aims of this study are to 1) reveal new information on the genetic and genomic diversity of native *N. ceranae* populations

through comparisons with similar published data from other parts of the world and 2) determine if pathogen genetic diversity differs among host honeybee species.

Materials and Methods

Spore Collection and DNA Extraction

Nosema ceranae samples were collected from naturally infected *A. cerana* hives within two provinces in Thailand (15 from Chumphon, 17 from Samut Songkhram) and from naturally infected *A. mellifera* hives within four provinces in Thailand (23 from Chiang Mai, 15 from Lamphun, 27 from Surat Thani, 7 from Chumphon, Fig. 1). Each sample represents spores collected from the midgut of five individual honeybees originating from a single colony, which are hereinafter referred to as populations or isolates. Spores were shipped and stored in absolute ethanol at -20°C. They were purified by centrifugation at 6,000 g for 10 minutes. The top white layer of sediment containing spores was collected and DNA was extracted using a MasterPure Complete DNA and RNA purification kit from Epicentre Biotechnologies (Madison, WI, USA). DNA was suspended in 35 µL of distilled water and stored at -20°C until required for analysis.

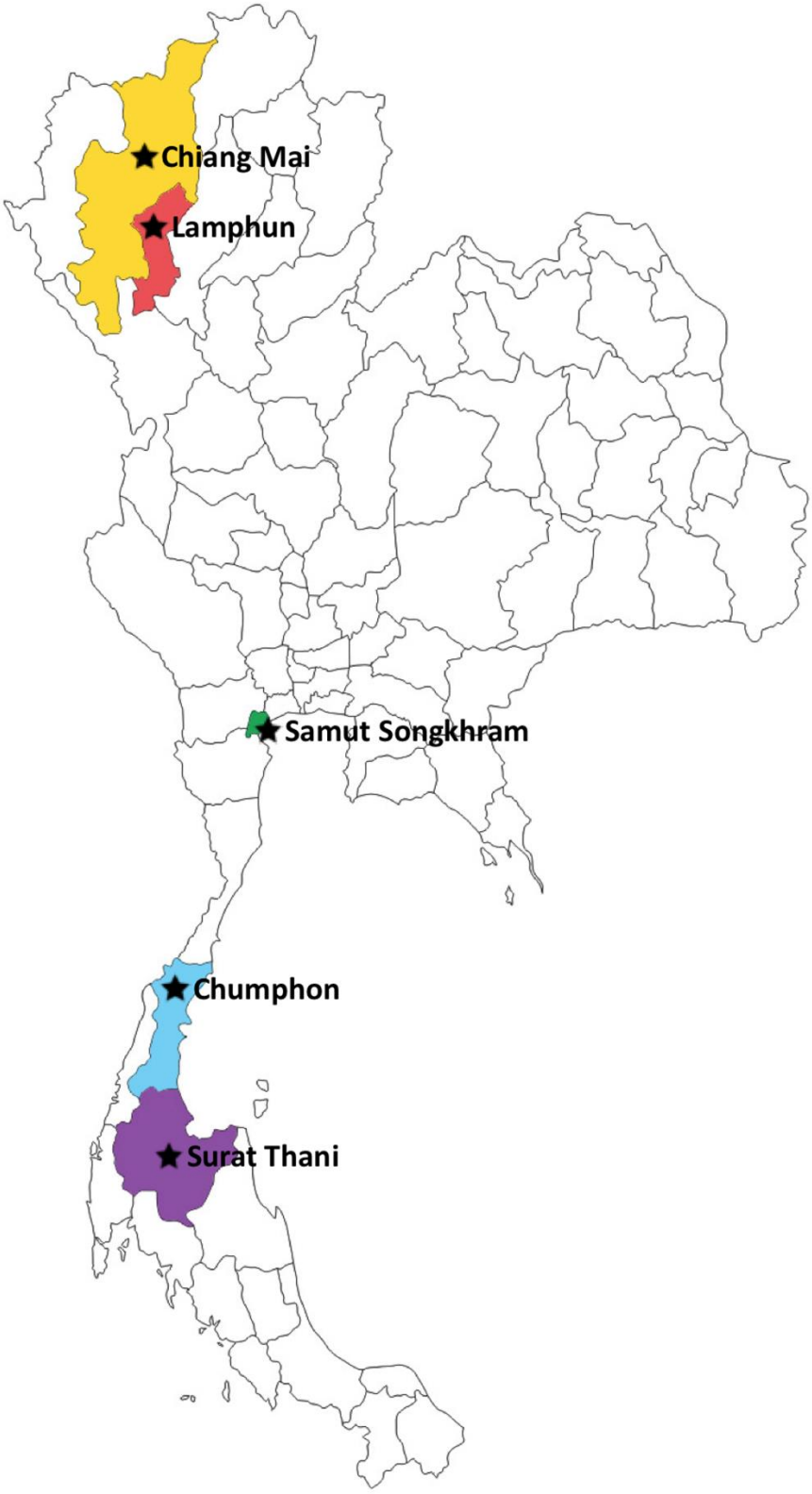


Fig. 1. Sampling locations of *N. ceranae* populations within Thailand.

SSU Gene Sequencing

Polymerase chain reaction (PCR) products for a region of the SSU gene were obtained for all *N. ceranae* populations from both *A. cerana* and *A. mellifera* hives in Northern, Central, and Southern Thailand using the forward primer SSU-F (5' GTA TAT GAG GTG ATT AAT TGG AGG GCA 3') and the reverse primer SSU-R (5' TAT ATC TAT TGT ATT GCG CGT GCA GC 3') forming a 673 bp product. PCR reactions were completed using either a low fidelity EconoTaq PLUS GREEN 2X DNA polymerase Master Mix (hereinafter called *Taq* polymerase) from Lucigen (Middleton, WI, USA) or a high fidelity PfuUltra II fusion Hot Start DNA polymerase (hereinafter called *Pfu* polymerase) from Agilent Technologies (Santa Clara, CA, USA). Sequences were amplified using a Biometra TProfessional Basic Gradient thermocycler or a Biometra T3000 thermocycler (Göttingen, Germany) with the following PCR conditions: initial denaturation at 95°C for 3 minutes, followed by 40 cycles of 95°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute, with a final extension of 72°C for 10 minutes and a 4°C hold. PCR products were analyzed on a 1.5% agarose gel containing ethidium bromide for DNA visualization and extracted using a GeneJET Gel Extraction kit from Thermo Fisher Scientific (Waltham, MA, USA). PCR products amplified with *Taq* polymerase of 24 populations from *A. mellifera* hives and 8 populations from *A. cerana* hives were cloned into competent bacterial cells using the StrataClone PCR Cloning kit (Santa Clara, CA, USA). PCR products amplified with *Pfu* polymerase of 3 populations from *A. mellifera* hives and 3 populations from *A. cerana* hives were similarly cloned using the StrataClone Blunt PCR Cloning kit (Santa Clara, CA, USA). Confirmation of the correct insertion into the bacterial plasmids was completed through PCR amplification using the plasmid-specific primers M13-F and M13-R and the following PCR reaction conditions: initial denaturation at 96°C for 3

minutes, followed by 45 cycles of 96°C for 30 seconds, 54°C for 30 seconds, and 60°C for 1.5 minutes, with a final extension of 60°C for 10 minutes and a 4°C hold. Plasmids containing inserts originally amplified with *Taq* polymerase were directly sequenced after amplification using the M13 primers, and plasmids containing inserts originally amplified with *Pfu* polymerase were first purified using the GeneJET Plasmid Miniprep kit from Thermo Fisher Scientific (Waltham, MA, USA) and then sequenced using the M13 primers. Twenty clones for each *N. ceranae* population were sent for sequencing resulting in between 7 and 20 viable sequences. Sanger dideoxy sequencing was completed at Génome Québec (Montréal, QC, Canada) using Applied Biosystems 3730xl DNA Analyzer technology (Foster City, CA, USA) and the sequences were submitted to Genbank (Accession numbers MH349761-MH349772, MH349776-MH350028, MH350034-MH350091, MH350093-MH350186, MH350257-MH350285, MH350296-MH350358, and MH356491-MH356536.).

Reference sequences for the region of the SSU gene amplified in this study were obtained from Genbank (Accession numbers DQ078785, DQ329034, DQ374656, DQ486027-DQ486028, DQ673615, EU025027, EU045844, FJ481912, FJ789785-FJ789786, FJ789788-FJ789789, FJ789791, FJ789795, FJ789797, FJ789799, JX205129, JX205149-JX205152, KC680618-KC680656, KU937099-KU937105, and U26533). Sequence alignment was completed in MEGA v7.0.20 (Kumar *et al.*, 2016) using the MUSCLE algorithm (Edgar, 2004). Nucleotide diversity was calculated using the "nuc.div" function (Nei, 1987) from the R package "pegas v0.10" (Paradis, 2010) and evidence for recombination was determined using DnaSP v5.101 (Librado and Rozas, 2009). A one-way ANOVA and Tukey Honest Significant Difference Post-hoc test was completed to determine if significant differences exist in the nucleotide diversity detected

between SSU sequences amplified with the *Taq* polymerase, *Pfu* polymerase, and sequences amplified in previous studies.

Detection and Analysis of a Novel Genotype

A distinct genotype showing similarity to both *N. ceranae* and its European sister species *N. apis* detected from original amplifications was specifically targeted with the same forward primer SSU-F (as described above) and the reverse primer SSUnovel-R (5' CCT CAG ATC ATA TCC TCG CCG AAC 3') using *Pfu* polymerase and the following PCR conditions: initial denaturation at 95 °C for 3 minutes, followed by 40 cycles of 95 °C for 30 seconds, 58.8 °C for 30 seconds, and 72 °C for 1 minute, with a final extension of 72 °C for 10 minutes and a 4 °C hold. Subsequent steps were completed as described above and the sequences were submitted to Genbank (Accession numbers MH350286-MH350295, MH350359-MH350385, and MH350396-MH350400).

Full Genome Sequencing and Mapping

Five *N. ceranae* populations were sent for Next-Generation full genome sequencing (Table S1). Libraries were created, multiplexed, and sequenced using Illumina HiSeq 4000 technology by Fasteris S.A. (Geneva, Switzerland). Sequencing produced paired end reads 150 bp in length that were subsequently demultiplexed. Adaptors were trimmed using the PERL script trim_fastq.pl from the PoPoolation v1.2.2 toolkit (Kofler, Orozco-terWengel, *et al.*, 2011) and overlapping paired reads were merged using SeqPrep (github.com/jstjohn/SeqPrep). Reads were then mapped to a previously assembled *N. ceranae* reference genome retrieved from Genbank (Accession number JPQZ00000000) (Pelin *et al.*, 2015) using BWA-aln v0.7.10-r789 (Li and Durbin, 2009) with the parameter '-n 0.1'. Resulting Sequence Alignment/MAP (SAM) files were converted to binary alignment (BAM) format, sorted, and indexed, followed by the

calculation of the average genome sequencing coverage (Table S1) using samtools v0.1.19-44428cd (Li *et al.*, 2009).

Polymorphism Comparison

Variants were called using the same parameters and software as Pelin *et al.* (2015) in order to facilitate comparisons between studies. For each isolate, heterozygous sites were identified using the PERL script snp-frequency-diff.pl from the PoPoolation2 v1.201 toolkit (Kofler, Pandey, *et al.*, 2011) with the parameters '--min-count 2 --min-coverage 2' and maximum coverage 25% higher than the average coverage of each isolate. Single nucleotide polymorphisms (SNP) were further filtered for only one alternate allele and remaining SNPs were used for further analyses.

To determine the level of intra- and inter-isolate diversity as well as the presence of distinct genetic populations, individual SNPs were compared between each of the Thailand isolates sequenced in this study along with eight other previously sequenced global isolates (Pelin *et al.*, 2015). SNP variation between isolates and study groups were compared to detect shared and unique genomic diversity. To measure the extent of genetic differentiation between populations, the fixation index (F_{ST}) was calculated for each pair of pathogen populations using the PERL script fst-sliding.pl from the PoPoolation2 v1.201 toolkit (Kofler, Pandey, *et al.*, 2011) with the parameters '--min-count 2 --min-coverage 2 --max-coverage 185' and a sliding window of 1 kb. To test the effect of lower sequencing coverage on genetic differentiation between populations, previously sequenced high coverage isolates were made to have artificially lower coverage. Ten replicates of a randomized subset of reads were taken using seqtk v1.2 (seqtk: Toolkit for processing sequences in FASTA/Q formats) which were mapped using the same pipeline as described above. Tests for genetic differentiation were made between isolates of

comparable coverage to show that variation in coverage does not greatly impact these results. The same heterozygous sites were used to calculate allele frequency plots for an analysis of ploidy. Both results were plotted in R using the 'ggplot2' package (Wickam, 2009).

Polymorphism Density

For the detection of regions showing decreased heterozygosity, SNPs were called using FreeBayes v0.9.21-5-g018c661 (Garrison and Marth, 2012) with the parameters '--ploidy 4 --pooled-continuous -F 0.10' with minimum coverage of two and maximum coverage 25% higher than the average coverage of each isolate. Density of heterozygous sites were calculated using the SNPdensity function in VCFtools v0.1011 (Danecek *et al.*, 2011) with a sliding window of 1 kb. Regions exhibiting loss of heterozygosity (LOH) were detected as adjacent windows containing equal or less than 2 SNPs/kb. Regions with drastic changes in sequencing coverage compared to the whole scaffold as determined in samtools v0.1.19-44428cd (Li *et al.*, 2009) were excluded and the remaining scaffolds containing LOH regions were plotted in R using the 'ggplot2' package (Wickam, 2009). Coding regions located in these areas were subsequently run through Blast2GO (Conesa *et al.*, 2005) to determine potential gene function.

Gene Diversity Analysis

Coding region nucleotide diversity was estimated with Watterson's θ values for every open reading frame (ORF) in each isolate. Watterson's θ is a measure of the population mutation rate that takes into account the number of SNPs found in each gene and corrects for the number of sequences (Watterson, 1975), where a higher θ value indicates more diversity. Values were estimated using the PERL script Variance-at-position.pl from the PoPoolation v1.2.2 toolkit (Kofler, Orozco-terWengel, *et al.*, 2011). The top 50 coding regions were subsequently run through Blast2GO (Conesa *et al.*, 2005) to determine potential gene function. Genes with

sequencing coverage too low to accurately determine θ and genes with coverage more than 25% higher than the average coverage for at least one of the four isolates were excluded from this analysis.

Recombination Analysis

Linkage disequilibrium is a measure of how often alleles are associated at two different loci and can be used to estimate recombination events within a genome, where drops in linkage disequilibrium indicate possible recombination events. Linkage disequilibrium (r^2) was estimated using LDx (Feder *et al.*, 2012), which is designed specifically for high-throughput sequencing data. Estimated r^2 values were plotted along the three largest contigs in R using the 'ggplot2' package (Wickam, 2009). LD decay was shown by plotting the mean r^2 by the distance between SNP pairs using the 'ggplot2' package (Wickam, 2009) in R.

De novo Assembly

Previously trimmed reads from the two isolates with the highest genomic coverage (SuratThani_Am, SamutSongkhram_Ac_2) were used to assemble a genome draft using SPAdes v3.11.0 (Nurk *et al.*, 2013) with kmers ranging from 21 to 127 (21, 33, 55, 77, 99, 127). Due to the intracellular nature of *N. ceranae*, DNA used for sequencing was highly contaminated with bacteria, host *Apis* sp., pollen, and other sources. Each assembly was separately screened for contamination using a combination of filtering based on GC content and BLAST (Altschul *et al.*, 1990). All contigs with GC content between 20-30% and obtaining a hit in the phylum Microsporidia or its sister phylum Cryptomycota using BLAST were retained, and these reads were reassembled with SPAdes v3.11.0 (Nurk *et al.*, 2013). Scaffolding was completed using AlignGraph (Bao *et al.*, 2014) which first maps previously assembled contigs onto a reference *N. ceranae* genome (Pelin *et al.*, 2015) and then uses reads to extend the contigs and ideally fills

gaps. Scaffolds less than 300 bp in length were discarded and BLAST was subsequently used to confirm the identity of the remaining scaffolds as microsporidian in origin. To determine if there were larger genomic rearrangements between Thai *N. ceranae* isolates and previously sequenced isolates, the best partial *de novo* assembly obtained here was compared to a previously assembled genome (Pelin *et al.*, 2015) using MUMmer (Delcher *et al.*, 2002).

Results

Nosema ceranae SSU Gene Diversity in Thailand

The SSU gene is often used for molecular characterization (Dong *et al.*, 2010; Shigano *et al.*, 2015) and diversity studies of microsporidian species such as *Pseudoloma neurophila* (Whipps and Kent, 2006), *Enterocytozoon bieneusi* (Zhang *et al.*, 2017), and *N. ceranae* (Huang *et al.*, 2008). In this study, presence of *N. ceranae* was detected in 104 sampled honeybee colonies through amplification of the SSU gene. This gene was sequenced to investigate nucleotide diversity present in 32 of these pathogen populations infecting Thai honeybees, resulting in 555 SSU sequences. We found that when using a conventional *Taq* DNA polymerase, the amount of nucleotide diversity (π) detected in the SSU gene (0.00732) is not significantly different to that of previously published sequences containing the same SSU region at 0.00712 ($p = 0.9284$). However, amplification of the same Thailand samples with a high fidelity *Pfu* DNA polymerase, which contains the ability to better proof-read sequences during PCR amplification, results in a much lower nucleotide diversity at 0.00285. The *Pfu*-driven sequence diversity is significantly lower than that of the other two groups ($p = 0.0000$, Fig. 2), and is similar to the diversity observed in this SSU region of related microsporidian species such as *N. bombycis*, *N. granulosis*, and *Vairimorpha cheracis* (<0.002) (Ironside, 2013).

Interestingly, using a high fidelity polymerase also eliminated any evidence for recombination (Fig. 2). These findings indicate that the amount of parasite nucleotide diversity, as well as evidence of recombination, is strongly linked with polymerase quality. Additionally, the observed nucleotide diversity shows no correlation to host honeybee species or location of the hive within Thailand – *i.e.* different honeybee hosts harbour similar levels of parasite genetic diversity.

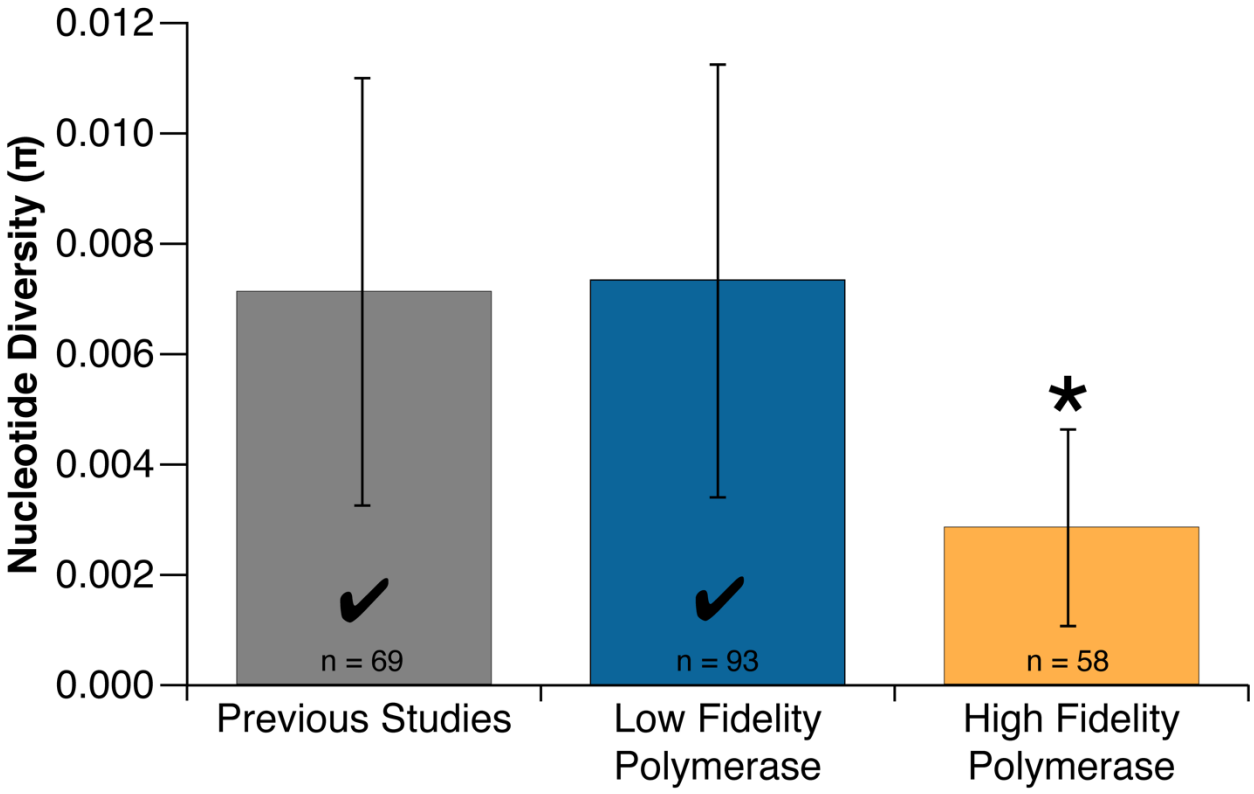


Fig. 2. Nucleotide diversity (π) and evidence for recombination found in the SSU gene of *N. ceranae* populations amplified in previous studies, in Thailand populations amplified using a low fidelity *Taq* DNA polymerase, and in Thailand populations amplified using a high fidelity *Pfu* DNA polymerase. Check marks (✓) indicate that evidence of recombination was detected. Error bars show the standard deviation calculated from the variance for each group.

To date, SSU analyses of global *N. ceranae* populations have failed to identify genotypes that are unique to specific regions of the world. Instead, most SSU variability is seemingly spread across the globe (Huang *et al.*, 2008; Sagastume *et al.*, 2011, 2014, 2016; Roudel *et al.*, 2013). In contrast to these findings, our analyses identified a genotype that is clearly distinct from those reported in global *N. ceranae* populations to date. This novel sequence contains several unique polymorphisms and shows some similarity to both species (Fig. 3). Targeting this sequence using genotype-specific primers for a total of 43 SSU sequences shows that it can be detected in many, but not all locations investigated across Thailand and in both host species.

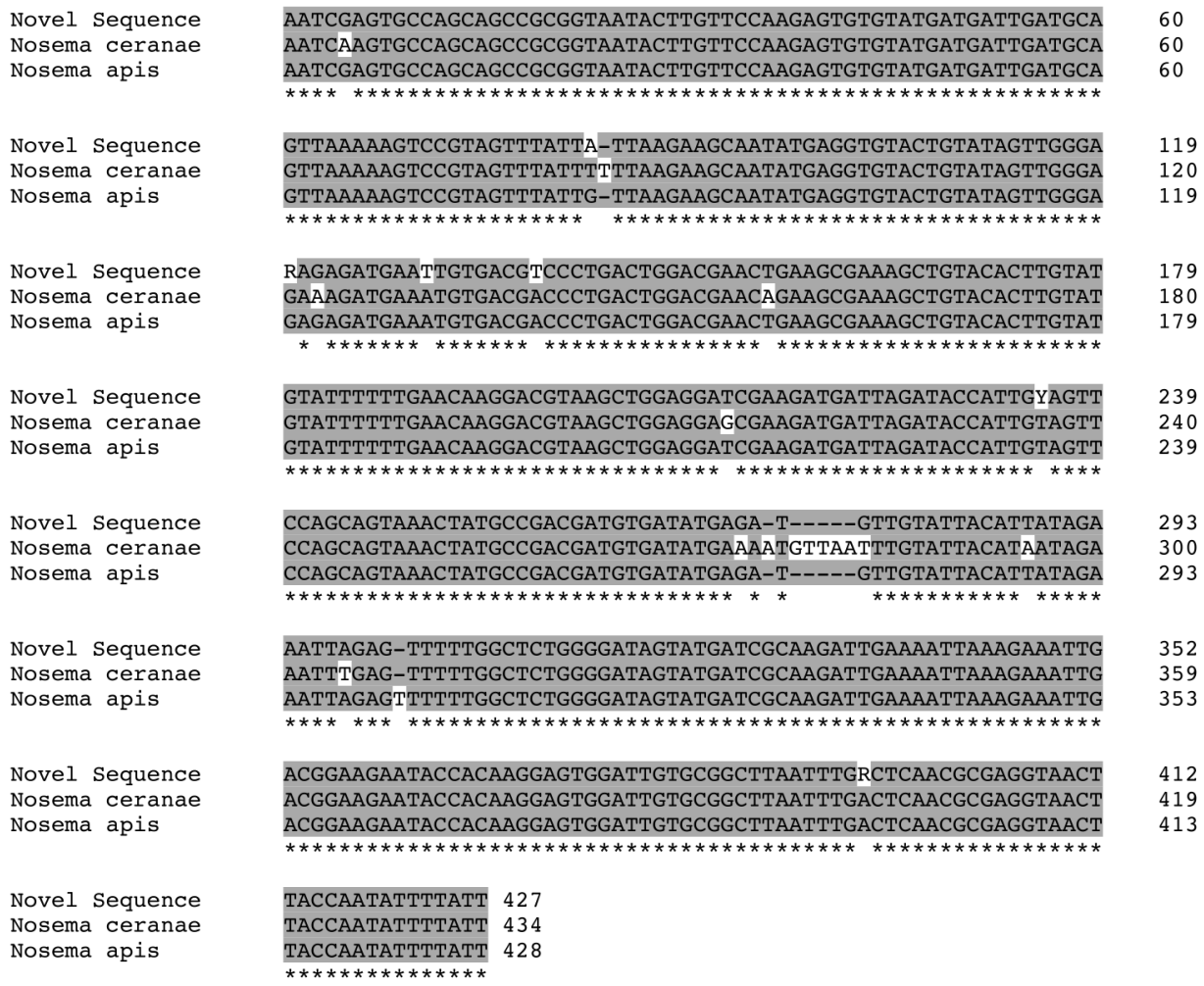


Fig. 3. Multiple sequence alignment of a region of the SSU gene. Includes a consensus sequence of the novel *Nosema* sequence detected in Thai honeybee populations (top row), *N. ceranae* (middle row), and *N. apis* (bottom row). Asterisks (*) indicate sites without polymorphisms and SNPs are highlighted in white.

***Nosema ceranae* Allelic Diversity Across the Genome in Thai Isolates**

To gain additional insight into the genetics of *N. ceranae* populations from Thailand, Illumina genome sequencing was performed on five *N. ceranae* samples. These samples were chosen because they produced robust SSU bands after PCR amplifications and represent parasite populations isolated from both host species and from hives located across the range of Thailand. Illumina sequencing resulted in the acquisition of 100,595,258 to 178,673,790 paired end reads depending on the isolate, with each library having over 86% Q30 following demultiplexing. Read mapping was performed using the *N. ceranae* reference genome (Accession number JPQZ00000000) (Pelin *et al.*, 2015), resulting in coverage ranging from a mere 1.6x to a maximum of 20.5x (Table S1). Data from the isolate Chumphon_Ac was not used for downstream analyses because of extremely low coverage.

In contrast to recent findings based on genome analyses of *N. ceranae* populations isolated from other regions of the globe (Pelin *et al.*, 2015), we find that Thai populations share few common polymorphic sites - *i.e.* 3,745 SNPs or between 8% and 18% of SNPs from one Thai isolate are shared among all isolates. However, many isolate-specific SNPs are also found in Thailand populations - *i.e.* between 9% and 20% of SNPs are isolate-specific (see solid bars in Fig. 4). This data is in contrast to the study by Pelin *et al.* (2015) which found that global *N. ceranae* isolates share up to 73% common sites, and only 2% isolate-specific SNPs. A large proportion of polymorphic sites are shared between both studies, as we found that between 56% and 72% of SNPs are found in at least one isolate from both studies (see spotted bars in Fig. 4), but there is also a large number of SNPs found only within one study group - *i.e.* between 28% and 44% of SNPs are found in more than one isolate from only one study group (see striped bars in Fig. 4). The high percentage of SNPs specific to each study indicates that the global isolates

analyzed by Pelin *et al.* (2015) and those analyzed in this study represent distinct populations. This is consistent with F_{ST} values for each pair of *N. ceranae* populations, which show that previously sequenced global isolates (Pelin *et al.*, 2015) are highly similar ($F_{ST} < 0.042$), whereas Thai populations show more genetic differentiation ($F_{ST} > 0.145$, Fig. 5).

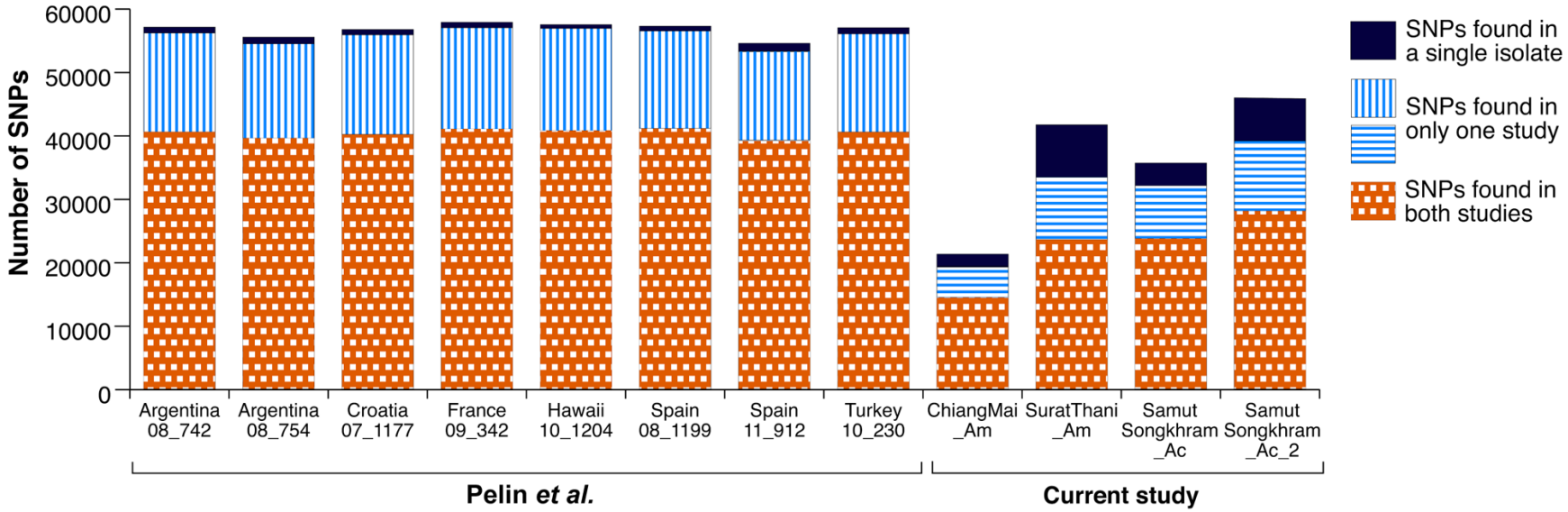
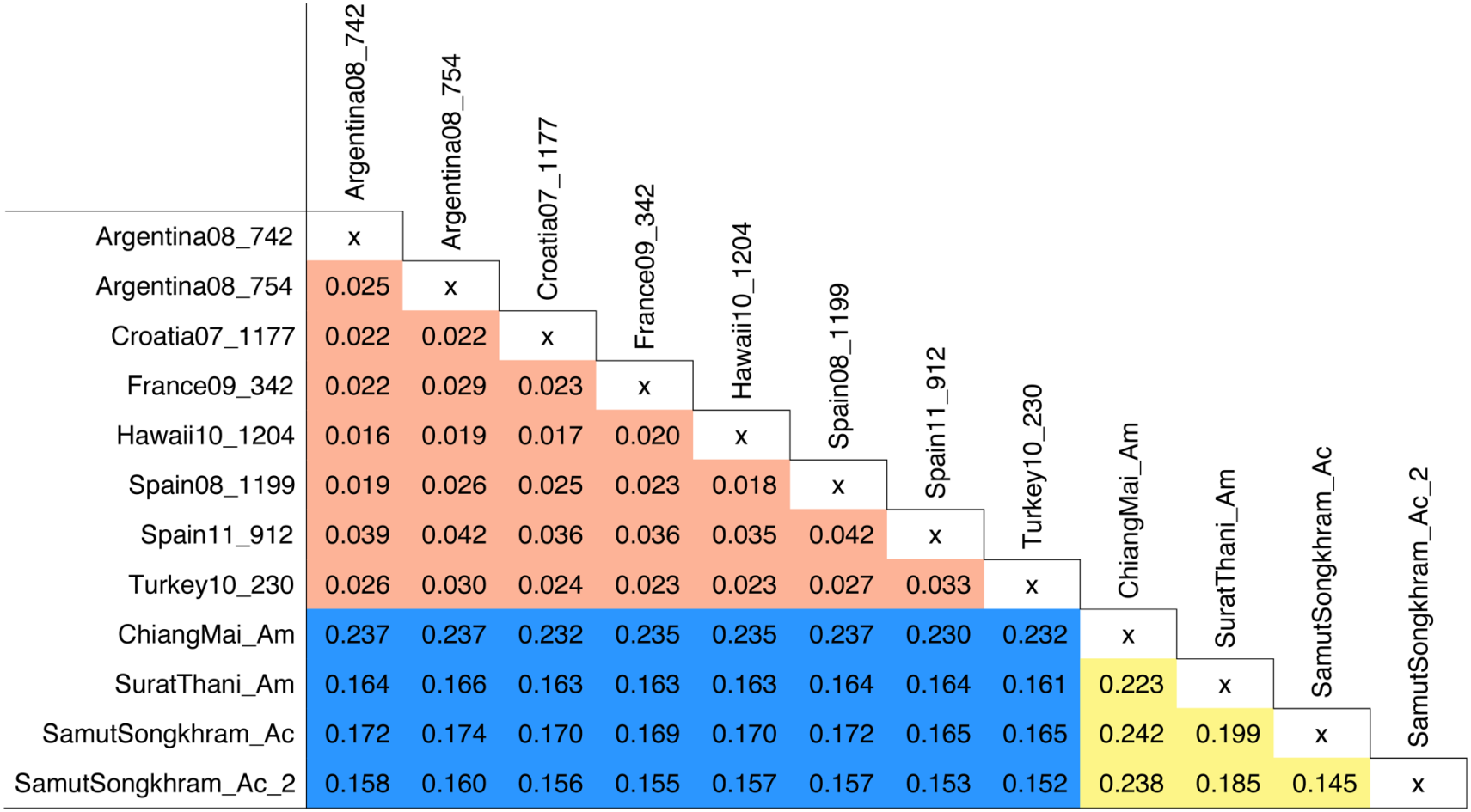


Fig. 4. Comparison of the polymorphic sites present in previously sequenced global *N. ceranae* populations (Pelin *et al.*, 2015) and newly sequenced populations from Thailand. SNPs found in isolates from both study groups are indicated by the spotted bars (bottom), SNPs found in isolates from only one study group are indicated by the striped bars (middle), and SNPs found in a single isolate are indicated by the solid bars (top).



	Argentina08_742	Argentina08_754	Croatia07_1177	France09_342	Hawaii10_1204	Spain08_1199	Spain11_912	Turkey10_230	ChiangMai_Am	SuratThani_Am	SamutSongkhram_Ac	SamutSongkhram_Ac_2
Argentina08_742	x											
Argentina08_754	0.025	x										
Croatia07_1177	0.022	0.022	x									
France09_342	0.022	0.029	0.023	x								
Hawaii10_1204	0.016	0.019	0.017	0.020	x							
Spain08_1199	0.019	0.026	0.025	0.023	0.018	x						
Spain11_912	0.039	0.042	0.036	0.036	0.035	0.042	x					
Turkey10_230	0.026	0.030	0.024	0.023	0.023	0.027	0.033	x				
ChiangMai_Am	0.237	0.237	0.232	0.235	0.235	0.237	0.230	0.232	x			
SuratThani_Am	0.164	0.166	0.163	0.163	0.163	0.164	0.164	0.161	0.223	x		
SamutSongkhram_Ac	0.172	0.174	0.170	0.169	0.170	0.172	0.165	0.165	0.242	0.199	x	
SamutSongkhram_Ac_2	0.158	0.160	0.156	0.155	0.157	0.157	0.153	0.152	0.238	0.185	0.145	x

Fig. 5. Fixation indices (F_{ST}) calculated for each pair of *N. ceranae* populations. Values are highlighted to emphasize comparisons between previously sequenced global populations (Pelin *et al.*, 2015) (upper left), comparisons between newly sequenced populations (lower right), and comparisons across studies (lower left).

It is noteworthy that the total number of SNPs found in Thai isolates is also much lower than that found by Pelin *et al.* (2015). This reduction cannot be solely attributed to the lower sequencing coverage of Thai isolates because coverage does not directly correlate with the number of SNPs found. Specifically, previous analyses of samples with only slightly higher coverage (*e.g.* 31x, Spain isolate 11_191 in Pelin *et al.* (2015)) were found to have much higher SNP levels. Furthermore, artificially reducing the coverage of isolates from Pelin *et al.* (2015) to levels comparable to those we obtained from Thai samples yielded the same results – *i.e.* Thai isolates always have a lower number of SNPs compared to other global populations of *N. ceranae*.

Recent genome data has revealed that *N. ceranae* is polyploid, harbouring more than two copies of each chromosome within its cells (Pelin *et al.*, 2015). This is in contrast with other microsporidian species with sequenced genomes, including *Edhazardia aedis* (Desjardins *et al.*, 2015), *Nematocida parisii* (Cuomo *et al.*, 2012), *P. neurophila* (Ndikumana *et al.*, 2017), and *E. cuniculi* (Selman *et al.*, 2013; Pelin *et al.*, 2016), which all show a clear diploid pattern following allele frequency analyses. Interestingly, allele frequency plots from Thai *N. ceranae* populations result in a complex allele frequency spectrum (Fig. S1), with a high allele frequency peak emerging at 0.5 and multiple lower peaks appearing along the allele frequency plot in all samples investigated. These scattered plots are difficult to interpret as they could represent either diploid (*i.e.* very high peak at 0.5) or slightly polyploid (*i.e.* scattered allele frequencies) organisms.

Nosema ceranae populations from Thailand show an average genome-wide SNP density of 5 SNPs/kb. This is 2.5-fold lower than previously sequenced isolates of *N. ceranae* (13 SNPs/kb) (Pelin *et al.*, 2015) and more closely resemble data from other microsporidian species such as *E. cuniculi* (4 SNPs/kb) (Pombert *et al.*, 2013) and *P. neurophila* (2 SNPs/kb)

(Ndikumana *et al.*, 2017). Regions showing loss of heterozygosity (LOH) have been suggested to affect adaptability and plasticity in other pathogens such as *Candida albicans* (Diogo *et al.*, 2009) and *Phytophthora capsici* (Lamour *et al.*, 2012). LOH is common in Microsporidia and has been previously detected in *P. neurophila* (Ndikumana *et al.*, 2017), *Nematocida* sp. (Cuomo *et al.*, 2012), and *N. ceranae* (Pelin *et al.*, 2015). Regions exhibiting LOH in Thai *N. ceranae* isolates (Fig. S2) range in length from 16 to 39 kb and harbour SNP densities from 0.080 to 1.500 SNPs/kb. Thai isolates have LOH regions found in other global isolates (scaffold JPQZ01000026.1) (Pelin *et al.*, 2015), but also some that appear unique to Thailand populations (scaffolds JPQZ01000014.1, JPQZ01000016.1). Only one LOH region is found across all the sequenced Thailand isolates (scaffold JPQZ01000014.1), whereas the other regions are present in some isolates but absent in others (scaffolds JPQZ01000016.1, JPQZ01000026.1). Interestingly, LOH seems to differ based on host species, as *N. ceranae* populations isolated from *A. mellifera* (ChiangMai_Am, SuratThani_Am) hives have larger (scaffold JPQZ01000026.1) and more frequent (scaffold JPQZ01000016.1) LOH regions compared to those infecting *A. cerana* (SamutSongkhram_Ac, SamutSongkhram_Ac_2) hives (Fig. S2). Genes located in these regions have a variety of functions and include ATPases, reverse transcriptases, proteins harbouring secretion signals, transport proteins, as well as other membrane components and enzymes (Table S2).

The genomes of our isolates were also inspected to detect genes undergoing rapid sequence change. In this case, Watterson's θ, which uses the number of SNPs within a coding region, was used to estimate nucleotide diversity, with larger θ values indicating more diversity. The average exome-wide θ is 0.0045, with the top 1% of genes having $\theta > 0.0249$ (13 genes, Fig. S3, Table S3), which is similar to data from previously sequenced global isolates (average θ =

0.0039) (Pelin *et al.*, 2015). It is evident that some genes have similar θ values (*e.g.* locus AAJ76_5300030346, θ = 0.0398 and 0.0387 respectively) and others have differing θ values (*e.g.* locus AAJ76_5300024082, θ = 0.0313 and 0.0153 respectively) between the two studies. The same pattern is found between Thai isolates from different hosts and diverse hive locations. Most of the coding regions identified are hypothetical proteins with no known functions, but genes with known functions are also found. These include DNA binding proteins, transposases, transmembrane transport proteins, RNA polymerases, proteins harbouring secretion signals, and many others (Table S3). Many of these genes, especially secreted proteins, can be involved in pathogen interactions with the host, where rapid sequence change could contribute to increased pathogenicity.

Linkage disequilibrium (LD) is a measure of the non-random association of alleles and can be used to estimate recombination events within a genome, where drops in LD values indicate possible recombination events. Levels of LD are also affected by mutation, genetic drift, and other gene-based or population-wide changes that alter allele frequencies and mix up the association of different loci. Levels of LD, reported as r^2, were determined across the three largest scaffolds in the Thai *N. ceranae* genomes (Fig. S4). Average LD levels across these three scaffolds are relatively high at 0.616 ± 0.117. This value is nearly identical to LD levels found in eight previously sequenced global isolates (r^2 = 0.6087 ± 0.3953) (Pelin *et al.*, 2015) and to data from distant clonal pathogens such as *C. albicans* (r^2 = 0.51) (Nebavi *et al.*, 2006) and *Leishmania infantum* (r^2 = 0.63 ± 0.41) (Rogers *et al.*, 2014). Similar to previous genome analyses (Pelin *et al.*, 2015), much of the variation in r^2 along these scaffolds seems to be conserved, indicating that these isolates may not be currently undergoing recombination. Slight differences in LD levels across the isolates could be due to variations in sequencing coverage. In

addition, LD appears to be quite stable along *N. ceranae* reads with little recombination breaking up haplotypes between SNP distances of up to 100 bp, but LD decay comes into play at larger distances (Fig. S5).

Partial *de novo* assemblies were produced to detect structural variations in the genome of Thai *N. ceranae* populations. In all cases, large scaffolds aligned to a previously assembled *N. ceranae* reference genome (Accession number JPQZ00000000) (Pelin *et al.*, 2015) were >99.8% identical in sequence and did not reveal any major inversions, insertions, or deletions.

Discussion

The *N. ceranae* isolates from Thailand we analyzed show some genetic similarities to those of global populations previously analyzed by others, but also exhibit many distinct features. In particular, our data confirms that populations of this pathogen share many SNPs with other global populations, and harbour similar levels of LD. The latter may indicate *N. ceranae* populations from Thailand are also clonal, as was previously hypothesized for their global counterparts (Pelin *et al.*, 2015). Clonal reproduction is also supported by a lack of recombining sequences among the cloned PCR products obtained using a high-fidelity polymerase.

Major differences with previous genetic and genome studies includes an overall lower within-population genetic diversity, as evidenced by both Sanger sequencing and SNP analysis. Interestingly, both analyses showed that Thai populations of *N. ceranae* harbour a genetic diversity that is similar to that found in other microsporidian species, including many that are diploid. Our allele frequency analyses are difficult to interpret because they show a scattered frequency pattern, but they all harbour a strong peak around the allele frequency of 0.5 suggesting that most variation in these isolates is in a diploid state (Fig. S1). This genomic

situation is drastically different from that previously found in other *N. ceranae* populations, where a similar number of SNP frequencies found at 0.5 are also found at 0.25 and 0.75 (Pelin *et al.*, 2015). Future analyses of native pathogen populations should prioritize acquiring better genome coverage from more isolates to understand the ploidy state of these populations. Such data will be essential to determine if polyploidization in *N. ceranae* is ancestral and conserved among all members of the species, or if its emergence coincided with the recent global spread of *N. ceranae*. If *N. ceranae* did undergo a polyploidization event, this could explain the higher genetic diversity found in global populations compared to Thai populations.

We also show that Thailand populations of *N. ceranae* have distinct genetic features not found in other global populations investigated to date, in contrast to previous findings showing that global populations are extremely similar from a genetic perspective. In particular, Thai populations show reduced intra-hive genetic variation compared to geographically distant populations and harbour distinct genotypes. A novel SSU genotype detected within Thailand populations of *Nosema* has not been previously reported in studies of *N. ceranae* (Fries *et al.*, 1996; Huang *et al.*, 2007, 2008; Medici *et al.*, 2012; Roudel *et al.*, 2013; Sagastume *et al.*, 2014) and its European sister species *N. apis* (Gatehouse and Malone, 1998). Its similarity to both *Nosema* species (Fig. 3) suggests that the sequence could be representative of a lineage that is intermediate between *N. ceranae* and *N. apis*. A clear genetic population distinction in *N. ceranae* based on geographic location is also evident through genome-wide comparisons of polymorphic sites among global isolates, as these revealed a large number of SNPs found exclusively in Thailand isolates. Overall, the reduced intra-hive diversity and the presence of location-specific genetic variation supports the notion that *N. ceranae* populations from Thailand have evolved independently of other global populations for some time.

An additional difference is found in two LOH regions exclusive to Thai *N. ceranae* genomes (Fig. S2). The sequencing coverage in the LOH regions show no signs of aneuploidy indicating LOH does not result from partial chromosome loss. Interestingly, inter-isolate variation in LOH appears to correlate with host honeybee species – *i.e. Nosema ceranae* populations isolated from *A. mellifera* hives show more LOH regions compared to those isolated from *A. cerana* hives - and geographic location - *i.e.* Thai isolates harbour specific LOH events. LOH events have been hypothesized to play a role in the ability of a pathogen to adapt to new environments (Diogo *et al.*, 2009; Rosenberg, 2011; Lamour *et al.*, 2012), and it will be important to see if this inter-isolate, inter-host, and inter-geographic pattern of variation in LOH holds as additional genomes are sequenced.

In conclusion, our study reveals that 1) populations of *N. ceranae* from Thailand show evidence of clonal evolution, similarly to what is found in other global populations, and 2) Thai populations contain distinct genotypes compared to other populations analyzed to date, have reduced intra-hive genetic variation, and harbour specific genetic diversity. Our data also indicates that LOH varies across global populations and, possibly, across hosts. From this study, it is clear that there is still novel diversity left to discover within populations of this honeybee-infecting pathogen. In the future, sequencing of additional isolates of *N. ceranae* from native geographic locations will further illuminate the global population dynamics of this pathogen. Particular attention should be directed at acquiring greater coverage and high-quality genome data, so that conclusions can be confidently drawn about the biology and genetics of this pathogen. Ultimately, despite its many challenges, using single-celled genomics and flow-cytometry analyses on *N. ceranae* individuals may be necessary to fully comprehend how genome polymorphism evolves within populations of these parasites.

Acknowledgements

We are grateful to Eric Chen, Nathan Liang, and Stephanie Mathieu for assistance in the laboratory. The authors declare no conflicts of interests. NC's work is supported by the Discovery program from the Natural Sciences and Engineering Research Council of Canada (NSERC-Discovery) and an Early Researcher Award from the Ontario Ministry of Research and Innovation (ER13-09-190).

Supplementary Data

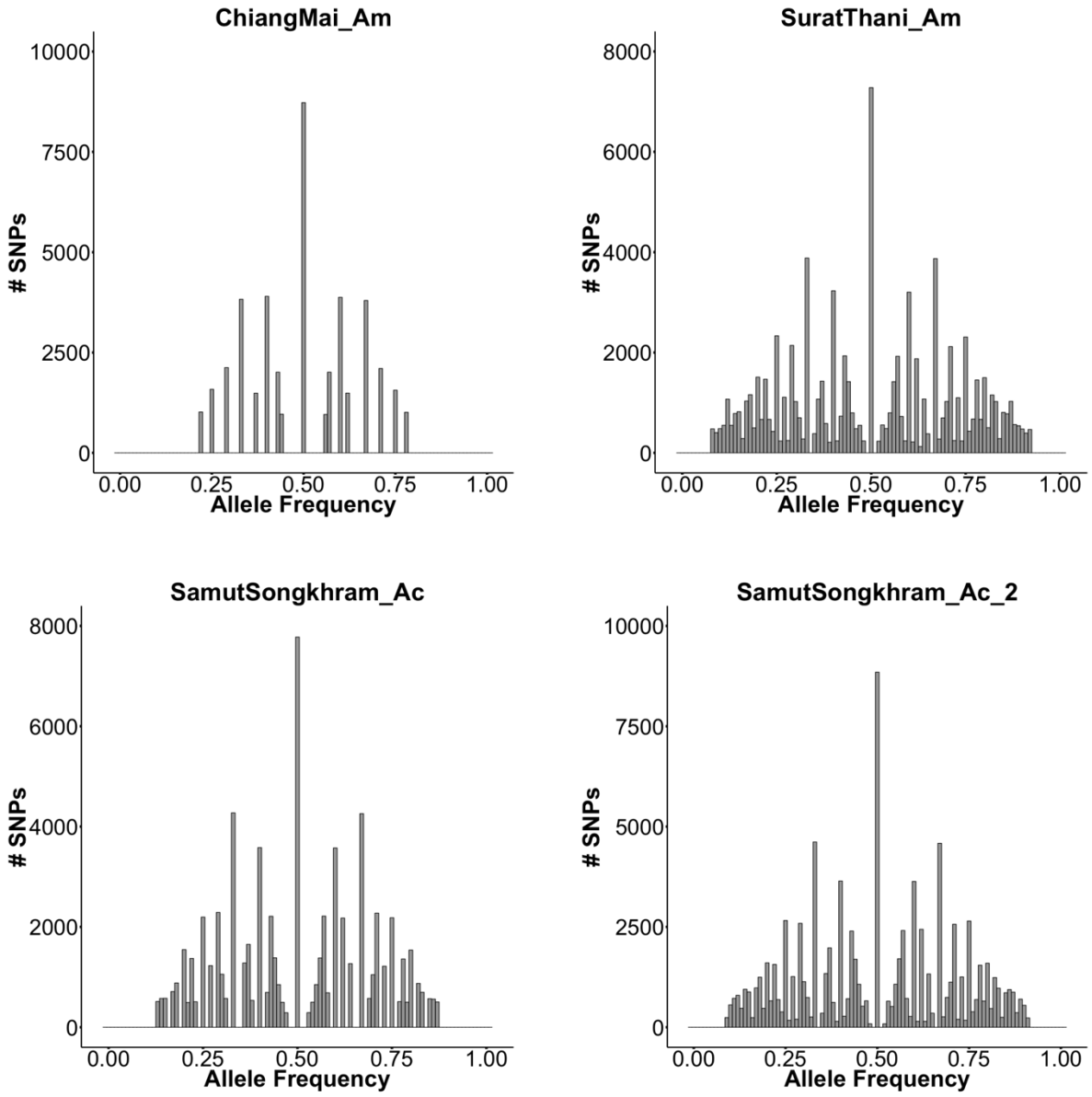


Fig. S1. SNP allele frequency distributions within the genome of Thailand isolates of *N. ceranae*. Frequencies are based on read counts of bi-allelic SNPs within each genome.

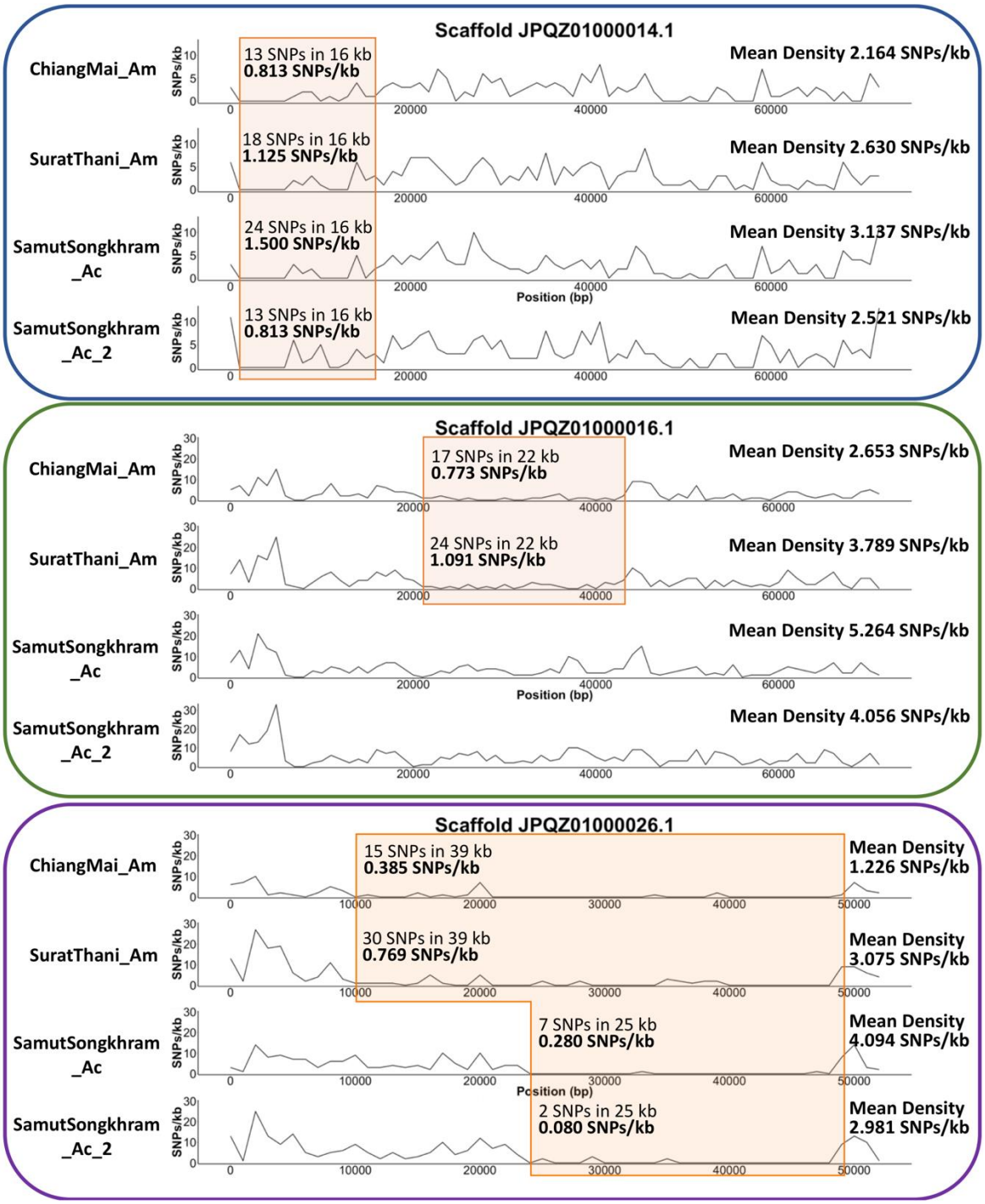


Fig. S2. SNP density within scaffolds containing evidence for loss of heterozygosity (LOH) in the genome of *N. ceranae* isolates from Thailand. SNPs were plotted in a 1 kb window and the regions exhibiting LOH are indicated. LOH present in scaffolds JPQZ01000016.1 and JPQZ01000026.1 are specific to Thailand populations.

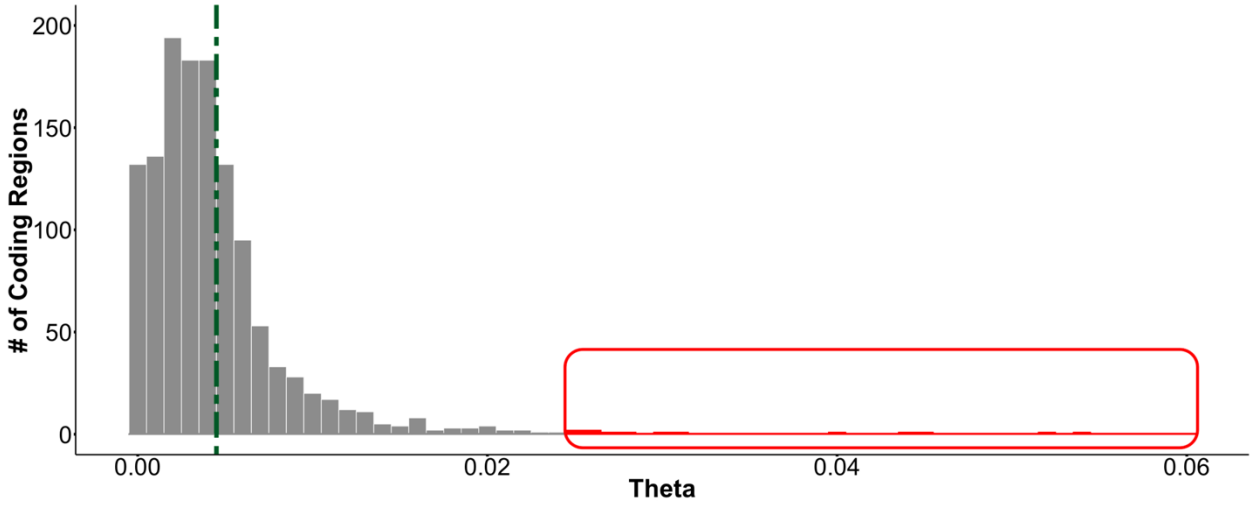


Fig. S3. Distribution of average Watterson's θ values for coding regions within *N. ceranae* populations from Thailand. The top 1% of genes are indicated in the boxed region and the average exome-wide θ is indicated by the dotted line.

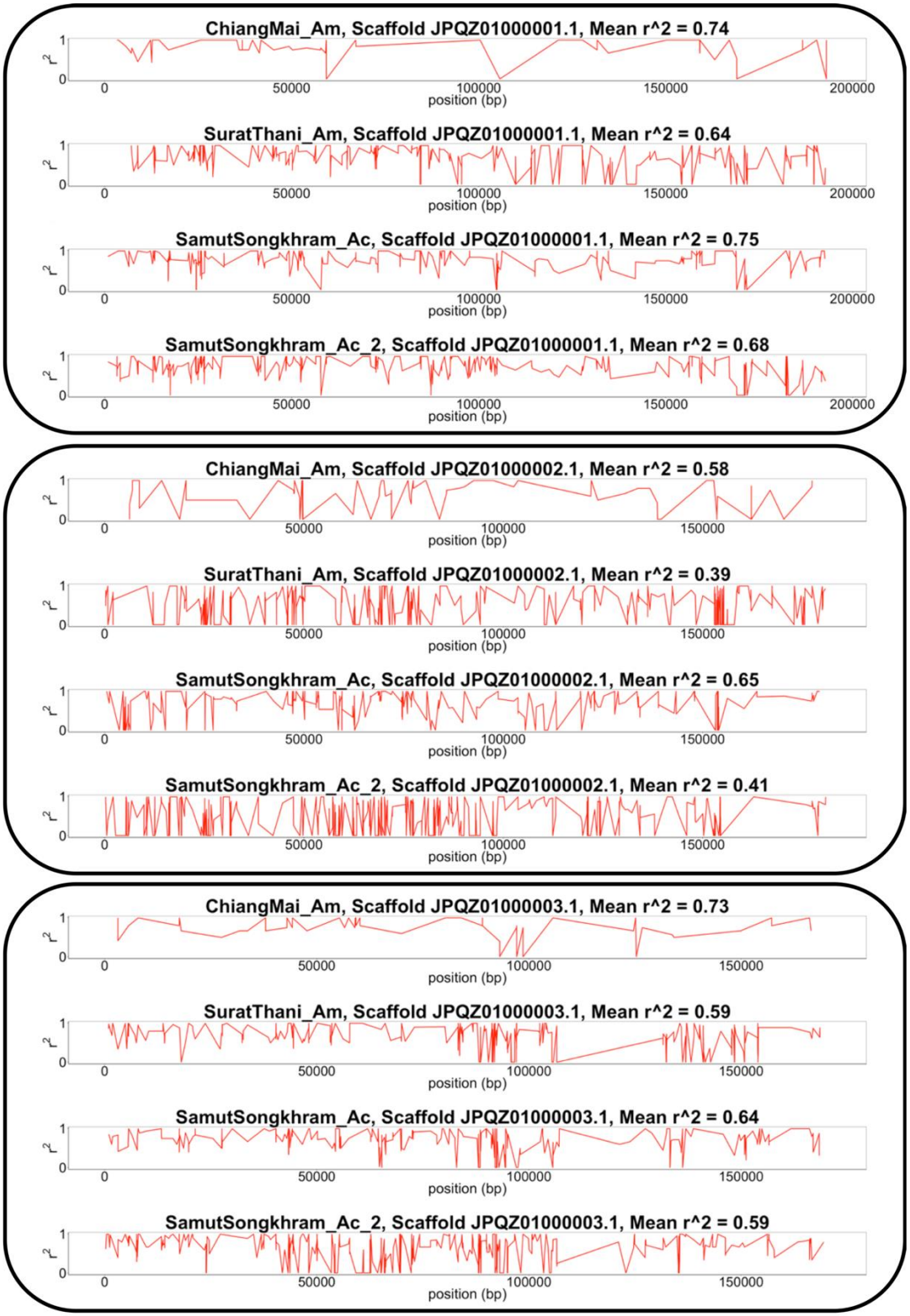


Fig. S4. Linkage disequilibrium (r^2) plotted along the 3 largest scaffolds of Thailand populations of *N. ceranae*.

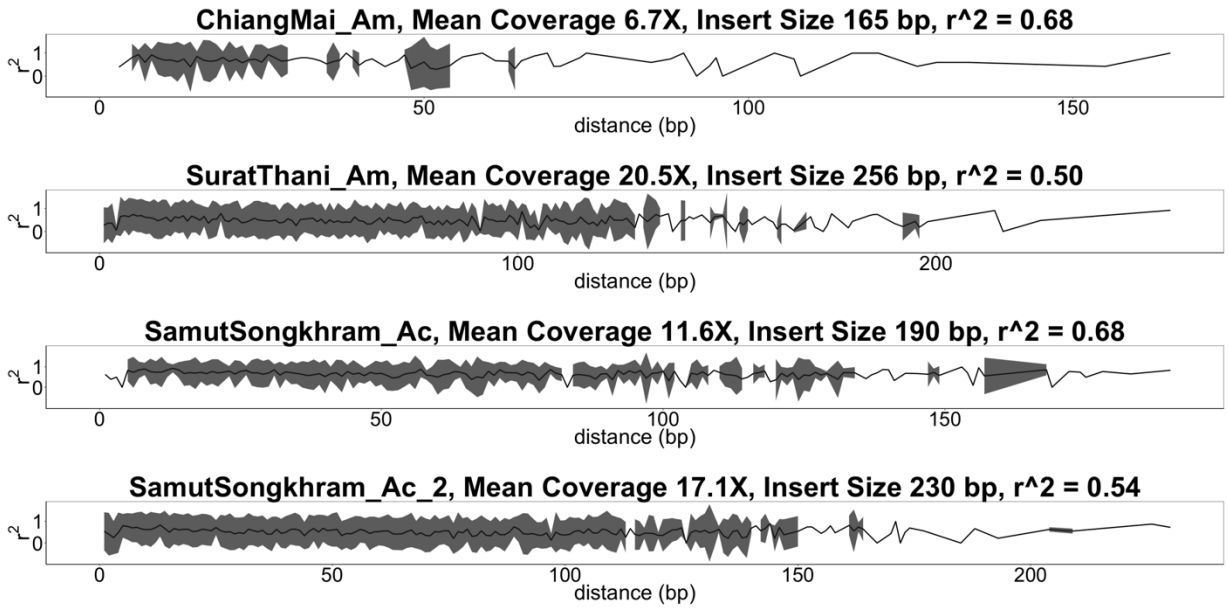


Fig. S5. Linkage disequilibrium decay plotted as mean r^2 as a function of distance between SNPs on the 3 largest scaffolds of *N. ceranae*. Grey area surrounding the curve indicates standard deviation from the mean at each distance.

Table S1. Sample and sequencing coverage information for populations of *N. ceranae* from Thailand sent for Illumina sequencing.

Sample ID	Thailand Province	Host Species	Coverage
ChiangMai_Am	Chiang Mai	*Apis mellifera*	6.74
SuratThani_Am	Surat Thani	*Apis mellifera*	20.5
SamutSongkhram_Ac	Samut Songkhram	*Apis cerana*	11.6
SamutSongkhram_Ac_2	Samut Songkhram	*Apis cerana*	17.1
Chumphon_Ac	Chumphon	*Apis cerana*	1.71

Table S2. Predicted functions for genes within regions exhibiting loss of heterozygosity (LOH).

Scaffold	Gene locus	Blast2GO results[a]
JPQZ01000014.1	AAJ76_140004215	membrane component
	AAJ76_140004770	-
	AAJ76_140005015	-
	AAJ76_140008469	**RNA polymerase II transcription initiation**; nucleotide excision repair; phosphorylation
	AAJ76_140008579	**tRNA-Ser**
	AAJ76_140009109	**AN1-type zinc finger 6**
	AAJ76_140009564	**heat shock transcription factor**; DNA binding
	AAJ76_1400011551	**ribosomal l14e 16e 127e**; transmembrane transport component
	AAJ76_1400012363	membrane component
	AAJ76_1400013838	membrane component
	AAJ76_1400018261	**DNA topoisomerase II**; DNA binding; ATP binding
JPQZ01000016.1	AAJ76_1600021854	**60S ribosomal l6**
	AAJ76_1600023079	**cyclin-dependent kinase**; ATP binding; phosphorylation;
	AAJ76_1600023921	signal peptide
	AAJ76_1600024122	**pp-loop ATPase**; ATP binding
	AAJ76_1600025098	**-**
	AAJ76_1600025958	**non-structural maintenance of chromosome element 4-like**; DNA repair
	AAJ76_1600026917	**glucose-6-phosphate 1-dehydrogenase**; NADP binding; oxidation-reduction reaction
	AAJ76_1600028830	**methyltransferase 5**; nucleic acid binding; methylation
	AAJ76_1600028874	**proteasome subunit beta type-5**; peptidase activity
	AAJ76_1600030566	HTH-like domain containing protein
	AAJ76_1600033726	membrane component
	AAJ76_1600034241	**willebrand factor type a domain-containing aaa ATPase**; ATP binding
	AAJ76_1600043340	**Sar1 GTPase**
JPQZ01000026.1	AAJ76_2600011763	membrane component
	AAJ76_2600014105	membrane component; signal peptide
	AAJ76_2600018150	membrane component

AAJ76_2600021468	membrane component
AAJ76_2600023707	membrane component; signal peptide
AAJ76_2600024453	**ubiquitin conjugating enzyme e2**
AAJ76_2600025454	**E3 ubiquitin-ligase RLIM-like**; zinc ion binding
AAJ76_2600026899	**homeobox hd-10**; DNA binding
AAJ76_2600030604	-
AAJ76_2600031547	-
AAJ76_2600031558	**reverse transcriptase domain containing**; RNA-directed DNA polymerase activity
AAJ76_2600031767	**reverse transcriptase domain containing**; RNA-directed DNA polymerase activity
AAJ76_2600035285	**abc transporter**; ATPase activity; membrane component
AAJ76_2600040696	**ATP-dependent DNA helicase**; DNA double-stranded break repair; nucleic acid binding
AAJ76_2600041420	-
AAJ76_2600043692	**S-adenosylmethionine synthase**; ATP binding
AAJ76_2600044091	-
AAJ76_2600045906	**glucose transporter type 3**; membrane component
AAJ76_2600047294	membrane component
AAJ76_2600048629	-

[a] *Functions displayed in bold are known genes.*

Table S3. The top 50 coding regions with the highest Watterson's θ values for each population of *N. ceranae* from Thailand and their predicted gene functions.

Locus	Average[a]	ChiangMai_Am	SuratThani_Am	SamutSongkhram_Ac	SamutSongkhram_Ac_2	Blast2GO results[b]
AAJ76_440004024	0.0544	0.0861	0.0534	0.0146	0.0636	membrane component
AAJ76_7400018155	0.0524	0.0224	0.0410	0.0658	0.0803	membrane component; mule transposase domain
AAJ76_2080003907	0.0447	0.0523	0.0525	0.0509	0.0233	DNA binding; tc3 transposase domain; homeobox-like domain
AAJ76_1900002416	0.0440	0.0500	0.0418	0.0418	0.0425	DNA binding
AAJ76_5300030346	0.0398	0.0433	0.0360	0.0364	0.0434	**mule transposase domain**; membrane component
AAJ76_5300024082	0.0313	0.0461	0.0302	0.0175	0.0314	-
AAJ76_450000581	0.0298	0.0381	0.0217	0.0223	0.0369	zinc finger
AAJ76_890001	0.0281	0.0293	0.0216	0.0281	0.0335	**mule transposase domain**; membrane component
AAJ76_2530002	0.0272	0.0636	0.0000	0.0000	0.0452	-
AAJ76_840007392	0.0262	0.0296	0.0352	0.0218	0.0181	membrane component; transmembrane domain; signal peptide
AAJ76_3400042604	0.0257	0.0134	0.0236	0.0279	0.0376	membrane component; transporter activity
AAJ76_2300002522	0.0250	0.0370	0.0209	0.0233	0.0189	-
AAJ76_610001	0.0250	0.0196	0.0253	0.0349	0.0201	-

AAJ76_400015576	0.0247	0.0336	0.0275	0.0203	0.0175	membrane component; transmembrane domain
AAJ76_469000831	0.0237	0.0285	0.0318	0.0035	0.0309	-
AAJ76_810006970	0.0225	0.0270	0.0191	0.0222	0.0219	-
AAJ76_396000978	0.0219	0.0389	0.0188	0.0053	0.0244	-
AAJ76_1270007257	0.0219	0.0338	0.0076	0.0289	0.0172	-
AAJ76_400076563	0.0213	0.0171	0.0352	0.0166	0.0162	signal peptide
AAJ76_4100015337	0.0206	0.0220	0.0208	0.0174	0.0220	-
AAJ76_258000249	0.0203	0.0256	0.0161	0.0244	0.0149	-
AAJ76_900086231	0.0200	0.0361	0.0139	0.0144	0.0157	-
AAJ76_187000383	0.0197	0.0238	0.0178	0.0180	0.0193	-
AAJ76_142000996	0.0196	0.0062	0.0208	0.0112	0.0399	signal peptide
AAJ76_6400012338	0.0193	0.0062	0.0259	0.0062	0.0390	membrane component; transmembrane domain
AAJ76_1990003537	0.0190	0.0099	0.0209	0.0245	0.0208	-
AAJ76_2400028038	0.0186	0.0143	0.0206	0.0231	0.0166	signal peptide
AAJ76_1160009299	0.0185	0.0000	0.0218	0.0294	0.0227	-
AAJ76_132000383	0.0184	0.0169	0.0155	0.0204	0.0208	-

AAJ76_230005011	0.0177	0.0000	0.0224	0.0484	0.0000	**ATP synthase subunit H**; membrane component; hydrogen ion transport; proton transport; signal peptide
AAJ76_4100030425	0.0171	0.0205	0.0172	0.0183	0.0125	signal peptide
AAJ76_1260003553	0.0167	0.0095	0.0122	0.0181	0.0270	nucleotide binding
AAJ76_3200011367	0.0165	0.0367	0.0000	0.0159	0.0133	-
AAJ76_1720002589	0.0164	0.0389	0.0000	0.0268	0.0000	signal peptide
AAJ76_2170003784	0.0161	0.0103	0.0186	0.0174	0.0182	-
AAJ76_2530002800	0.0160	0.0374	0.0126	0.0088	0.0054	-
AAJ76_4000157161	0.0160	0.0111	0.0239	0.0128	0.0162	signal peptide
AAJ76_3600011248	0.0158	0.0221	0.0122	0.0093	0.0197	-
AAJ76_910008052	0.0158	0.0305	0.0178	0.0051	0.0095	mule transposase domain
AAJ76_1930002	0.0156	0.0296	0.0049	0.0094	0.0184	transmembrane domain
AAJ76_70001144	0.0154	0.0219	0.0224	0.0000	0.0173	-
AAJ76_3430002	0.0149	0.0171	0.0110	0.0117	0.0199	-
AAJ76_8100013039	0.0147	0.0222	0.0122	0.0000	0.0246	-
AAJ76_1330007479	0.0146	0.0000	0.0121	0.0131	0.0334	membrane component; transmembrane domain
AAJ76_	0.0144	0.0086	0.0117	0.0261	0.0111	-

3000132534						
AAJ76_3700041073	0.0143	0.0130	0.0101	0.0175	0.0165	**RNA polymerase III transcription factor IIIC subunit**; transcription factor
AAJ76_1340003383	0.0141	0.0033	0.0164	0.0163	0.0203	-
AAJ76_200072272	0.0139	0.0120	0.0147	0.0144	0.0145	transmembrane domain
AAJ76_1510006055	0.0138	0.0081	<u>0.0258</u>	0.0139	0.0076	signal peptide
AAJ76_6800012709	0.0134	0.0201	0.0119	0.0088	0.0129	-

[a] *Watterson's θ values within the top 1% of genes (θ > 0.0249) are underlined.*

[b] *Functions displayed in bold are known genes.*

Chapter 3 – Discussion, Future Directions, and Concluding Remarks

Importance of Research

There has been knowledge of microsporidian infections in honeybees since *Nosema apis* was first discovered in hives of the Western honeybee, *Apis mellifera*, in 1909 (Zander, 1909). *Nosema ceranae* was only identified in 1996 (Fries *et al.*, 1996) as a pathogen exclusive to the Asian honeybee, *A. cerana*. Not surprisingly, scientific interest in *N. ceranae* did not develop until 10 years later when it was first found to also infect the more economically and ecologically important honeybee, *A. mellifera* (Higes *et al.*, 2006). From then on, scientific and public interest in this pathogen has been increasing due to the importance of the Western honeybee to honey production and pollination globally (Kearns *et al.*, 1998; Delaplane and Mayer, 2000; Suwannapong, Benbow, *et al.*, 2011). *Nosema ceranae* is of specific concern because it is highly pathogenic to the Western honeybee (Higes *et al.*, 2007; Paxton *et al.*, 2007; Botías *et al.*, 2013), and some research has found an association between the presence of this pathogen and Colony Collapse Disorder (CCD) (Martín-Hernández *et al.*, 2007; Higes *et al.*, 2008; VanEngelsdorp *et al.*, 2009; Dussaubat *et al.*, 2012). There have been many studies focusing on *N. ceranae* infections with precisely 177 publications found on PubMed from the last five years. Genetic and genomic studies have shown that pathogen populations harbour a large amount of diversity (Fries *et al.*, 1996; Huang *et al.*, 2007, 2008, Sagastume *et al.*, 2011, 2014, 2016; Suwannapong, Yemor, *et al.*, 2011; Medici *et al.*, 2012; Roudel *et al.*, 2013) but have also found that geographically separated populations are not genetically distinct (Pelin *et al.*, 2015). Interestingly, only five studies from the last five years have looked at native pathogen populations from *A. cerana* hives located within South-East Asia. This is surprising because

understanding how *N. ceranae* populations evolve in their native habitat can help us in understanding the global dynamics of this pathogen, and provide important insights into the evolution, genetics, and recent expansions of *N. ceranae*.

Summary of Novel Findings

This thesis aimed to gather information on native populations of *N. ceranae* and compare their genetic and genome diversity to previously sequenced populations from other regions of the world. The biology and genetics of *N. ceranae* were studied here using both PCR and genome-based methods, revealing several similarities to global populations as well as multiple features that appear to be unique to native pathogen populations. Interestingly, all *N. ceranae* populations from Thailand show the same patterns in terms of genetics and genome evolution, with the exception of one analysis that appears to show the presence of host species specificity.

Clonal Reproduction

Reproduction in Microsporidia has not been well-studied but can reveal important information on the basic biology of a pathogen and can help explain patterns of genetic diversity. A clonal mode of reproduction can be indicated by the lack of recombining sequences. Recombination is easily detected in single gene sequences and can also be identified through low linkage disequilibrium (LD), which occurs when recombination or other cellular events break up the association of alleles. Previous studies have detected recombination events within single gene sequences of *N. ceranae* (Sagastume *et al.*, 2011, 2014, 2016; Roudel *et al.*, 2013; Gómez-Moracho *et al.*, 2015), but the results presented here corroborate with a previous whole genome study by Pelin *et al.* (2015) as there is evidence that *N. ceranae* likely undergoes a clonal reproduction cycle. The lack of recombining sequences found in SSU fragments and high LD

levels in the genome indicate native populations of *N. ceranae* are also consistent with a clonal mode of reproduction (Pelin *et al.*, 2015).

Reduced Genetic Diversity

High levels of nucleotide diversity have been commonly found in single gene sequences of *N. ceranae*, especially within the SSU gene (Fries *et al.*, 1996; Huang *et al.*, 2007, 2008, Sagastume *et al.*, 2011, 2014, 2016; Suwannapong, Yemor, *et al.*, 2011; Medici *et al.*, 2012; Roudel *et al.*, 2013). Unlike previous studies of *N. ceranae*, the diversity reported here in native pathogen populations is much lower. This reduced genetic diversity is evident in both the SSU gene and across the whole genome, where the overall number of SNPs is much lower than has been previously detected in the genomes of other global populations. This unique feature of native *N. ceranae* populations indicates that they have likely continued to evolve independently from other global pathogen populations.

Interestingly, the level of genetic diversity observed here is similar to that found in distant microsporidian species, particularly those that are diploid (Cuomo *et al.*, 2012; Desjardins *et al.*, 2015; Ndikumana *et al.*, 2017). An SNP frequency analysis was completed for native *N. ceranae* populations resulting in a complicated allele frequency spectrum. These spectrums could indicate diploidy as there is a strong peak at a frequency of 0.5 or polyploidy as there are several other smaller peaks along the spectrum. This pattern is strikingly different from that reported by Pelin *et al.* (2015), who found that global *N. ceranae* populations appear to be tetraploid with strong allele frequency peaks at 0.25, 0.5, and 0.75. Differences in ploidy among *N. ceranae* populations could have implications regarding this pathogen and its recent hypothesized host transfer and global expansion.

Unique Genetic Diversity

Up until now, geographically separated populations of *N. ceranae* have been shown to be very genetically similar (Pelin *et al.*, 2015). In contrast, an SSU genotype was detected in Thailand populations that has not been found in other global populations of *N. ceranae*, indicating the presence of location-specific diversity. In addition to this finding, a comparison of the polymorphic sites detected between isolates from Thailand revealed much less shared variation than has been found in other global populations previously sequenced by Pelin *et al.* (2015). Fewer than 18% of SNPs analyzed are shared among all global *N. ceranae* populations, and up to 20% of SNPs are specific to a single isolate from Thailand. Native populations have an elevated amount of location-specific genetic diversity with up to 43% of detected SNPs being found exclusively in Thailand pathogen populations. The location-specific diversity detected is another unique feature of *N. ceranae* populations from Thailand and similarly indicates that these populations have evolved separately from other global populations for some time.

Patterns of Loss of Heterozygosity

Loss of heterozygosity (LOH) occurs when additional copies of an allele are lost through chromosome reduction or mitotic recombination resulting in a single allele and greatly reduced heterozygosity in that region. LOH events have been suggested to play a role in pathogen adaptability to new environments through the unmasking of potentially advantageous recessive alleles (Diogo *et al.*, 2009; Rosenberg, 2011; Lamour *et al.*, 2012). By measuring the SNP density along the genome of *N. ceranae*, three regions were identified that exhibit LOH. One of these regions has also been detected in global populations of *N. ceranae* by Pelin *et al.* (2015), but the remaining two regions are unique to populations from Thailand once again indicating these populations have undergone independent evolution. Interestingly, unlike the other analyses

already described, the LOH events detected here appear to differ based on host species, where isolates from *A. mellifera* have larger and more frequent LOH events than isolates from *A. cerana*. If this pattern of host-specificity continues as more genomes are sequenced, it could provide insight into the recent hypothesized host transfer and global expansion of *N. ceranae*.

Further Implications and Future Directions

The findings presented here reveal that a substantial amount of genetic diversity still awaits discovery in global populations of *N. ceranae*. To date, full genome sequencing has been completed for populations from across Europe and South America, with the addition of native Asian populations in this thesis. Despite the broad geographic genome sequencing already completed, some continents such as North America, Africa, and Australia have yet to be sampled from a genome perspective. Obtaining similar sequencing data from under sampled geographic locations can help further illuminate the global population dynamics of *N. ceranae* and determine whether additional genetically isolated populations exist like those located in Thailand.

Most genetic and genome studies to date have focused on *N. ceranae* infections of the Western honeybee, *A. mellifera*, due to their importance to honey production and crop pollination worldwide. The research presented here explores the genome of previously understudied pathogen populations infecting the Asian honeybee, *A. cerana*, to detect differences between hosts. In addition to these two main hosts, *N. ceranae* can infect several other organisms including the honeybee species *A. dorsata* and *A. florea* (Chaimanee *et al.*, 2013) as well as several bumblebee species (*Bombus* spp.) (Plischuk *et al.*, 2009). Future sequencing of pathogen populations from different host species will provide insight on the environmental adaptability of *N. ceranae* and potential mechanisms of host transfer.

The results on genetic diversity in native *N. ceranae* populations presented in this thesis are intriguing as genetically distinct pathogen populations have never been identified prior to this work. Unfortunately, low genome sequencing coverage prevents the formation of conclusions regarding the ploidy of these populations. Thus, future studies should prioritize the acquisition of greater coverage and higher quality genome data in order to better understand the biology and genetics of *N. ceranae*. Particularly, studies aiming to determine ploidy and SNP density across the genome will benefit from high coverage genome data as this data will contain less gaps and provide higher resolution of polymorphic sites.

There are two techniques that could be utilized in Microsporidia, particularly with *N. ceranae*, to help resolve how genome polymorphism evolves within these pathogen populations. One of these is single-cell genomics, which allows for Next-Generation Sequencing (NGS) of a single cell isolated from a population (Gawad *et al.*, 2016). This technique can provide context related to the genome diversity within and between individuals, which is not achieved when sequencing large populations of cells as is commonly done in genomic studies of microorganisms. It can also be combined with transcriptome analyses, providing an extra layer of information on the biology of this pathogen. The second technique is flow-cytometry which uses specific fluorochromes to count or sort cells. Specifically, flow-cytometry can be used to determine ploidy in a cell using DNA fluorochromes to estimate genome size (Doležel and Barto, 2005). This application could help resolve the complicated level of ploidy present in populations of *N. ceranae*.

Although these techniques are very useful, there are many challenges to using them with microsporidian species. The intracellular nature and small size of microsporidian spores make them difficult to isolate for single-cell methods, and these pathogens have hard spore walls

increasing the difficulty of using fluorochromes for flow-cytometry. In addition, some microsporidian species including *N. ceranae* have a diplokaryon - *i.e.* binucleate - nucleus further complicating genome organization. Overcoming the challenges of these techniques would provide invaluable opportunities to gather more insight into the basic biology and evolution of *N. ceranae* and pathogens like it.

Concluding Remarks

Studying populations of *N. ceranae* that are geographically separated and those that infect different hosts is required to understand the global population dynamics of this pathogen. The findings presented in this thesis reveal that native populations of *N. ceranae* exhibit unique genetic differences compared to other global pathogen populations but also have some similarities. In agreement with previously sequenced *N. ceranae* populations from around the world, populations from Thailand appear to undergo clonal reproduction. On the other hand, the native populations of *N. ceranae* described here have reduced levels of genetic diversity, contain distinct genotypes compared to other populations analyzed to date and also harbour genetic diversity specific to Thailand. Additionally, LOH was found to vary across global populations of *N. ceranae* and could possibly differ based on host species. Future studies are required to see if these patterns of genome diversity hold true in other populations of *N. ceranae*. Overall, this study provides valuable information on previously understudied native populations of *N. ceranae* and begins to illuminate interesting and novel patterns of genetic diversity that brings us one step closer to understanding the biology and genetics of this important honeybee pathogen.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**: 403–10.

Anderson, D.L. and Trueman, J.W.H. (2000) Varroa jacobsoni (Acari: Varroidae) is more than one species. *Exp. Appl. Acarol.* **24**: 165–189.

Balbiani, G. (1882) Sur les microsporidies ou psorospermies des articules. *Comptes rendus l'Académie des Sci.* **95**: 1168–1171.

Bao, E., Jiang, T., and Girke, T. (2014) AlignGraph: algorithm for secondary de novo genome assembly guided by closely related references. *Bioinformatics* **30**: i319–i328.

Beaurepaire, A.L., Truong, T.A., Fajardo, A.C., Dinh, T.Q., Cervancia, C., and Moritz, R.F.A. (2015) Host Specificity in the Honeybee Parasitic Mite, Varroa spp. in Apis mellifera and Apis cerana. *PLoS One* **10**: e0135103.

Becnel, J.J. and Andreadis, T.G. (2014) Microsporidia in Insects. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 521–570.

Bollan, K.A., Hothersall, J.D., Moffat, C., Durkacz, J., Saranzewa, N., Wright, G.A., et al. (2013) The microsporidian parasites Nosema ceranae and Nosema apis are widespread in honeybee (Apis mellifera) colonies across Scotland. *Parasitol. Res.* **112**: 751–759.

Botías, C., Anderson, D.L., Meana, A., Garrido-Bailón, E., Martín-Hernández, R., and Higes, M. (2012) Further evidence of an oriental origin for Nosema ceranae (Microsporidia: Nosematidae). *J. Invertebr. Pathol.* **110**: 108–113.

Botías, C., Martín-Hernández, R., Barrios, L., Meana, A., and Higes, M. (2013) Nosema spp. infection and its negative effects on honey bees (Apis mellifera iberiensis) at the colony level. *Vet. Res.* **44**: 25.

Campbell, S.E., Williams, T.A., Yousuf, A., Soanes, D.M., Paszkiewicz, K.H., and Williams, B.A.P. (2013) The Genome of Spraguea lophii and the Basis of Host-Microsporidian Interactions. *PLoS Genet.* **9**: e1003676.

Cavalier-Smith, T. (1983) A 6-kingdom classification and a unified phylogeny. In, Shenck,H.E.A. and Schwemmler,W.S. (eds), *Endocytobiology II. Intracellular Space as Oligogenetic*. Walter de Gruyter & Co., Berlin, pp. 1027–1034.

Cavalier-Smith, T. (1998) A revised six-kingdom system of life. *Biol. Rev.* **73**: 203–266.

Chaimanee, V., Pettis, J.S., Chen, Y., Evans, J.D., Khongphinitbunjong, K., and Chantawannakul, P. (2013) Susceptibility of four different honey bee species to Nosema ceranae. *Vet. Parasitol.* **193**: 260–5.

Chen, Y., Evans, J.D., Smith, I.B., and Pettis, J.S. (2008) Nosema ceranae is a long-present and wide-spread microsporidian infection of the European honey bee (Apis mellifera) in the United States. *J. Invertebr. Pathol.* **97**: 186–188.

Chen, Y.P., Evans, J.D., Murphy, C., Gutell, R., Zuker, M., Gundensen-Rindal, D., and Pettis, J.S. (2009) Morphological, molecular, and phylogenetic characterization of Nosema ceranae, a microsporidian parasite isolated from the European honey bee, Apis mellifera. *J. Eukaryot. Microbiol.* **56**: 142–147.

Chen, Y.P. and Huang, Z.Y. (2010) Nosema ceranae, a newly identified pathogen of Apis mellifera in the USA and Asia. *Apidologie* **41**: 364–374.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Cornman, R.S., Chen, Y.P., Schatz, M.C., Street, C., Zhao, Y., Desany, B., et al. (2009) Genomic Analyses of the Microsporidian Nosema ceranae, an Emergent Pathogen of Honey Bees. *PLoS Pathog.* **5**: e1000466.

Corradi, N. (2015) Microsporidia: Eukaryotic Intracellular Parasites Shaped by Gene Loss and Horizontal Gene Transfers. *Annu. Rev. Microbiol.* **69**: 167–183.

Corradi, N., Akiyoshi, D.E., Morrison, H.G., Feng, X., Weiss, L.M., Tzipori, S., and Keeling, P.J. (2007) Patterns of Genome Evolution among the Microsporidian Parasites Encephalitozoon cuniculi, Antonospora locustae and Enterocytozoon bieneusi. *PLoS One* **2**: e1277.

Corradi, N., Gangaeva, A., and Keeling, P.J. (2008) Comparative profiling of overlapping transcription in the compacted genomes of microsporidia Antonospora locustae and Encephalitozoon cuniculi. *Genomics* **91**: 388–393.

Corradi, N. and Keeling, P.J. (2009) Microsporidia: a journey through radical taxonomical revisions. *Fungal Biol. Rev.* **23**: 1–8.

Corradi, N., Pombert, J.-F., Farinelli, L., Didier, E.S., and Keeling, P.J. (2010) The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. *Nat. Commun.* **1**: 1–7.

Corradi, N. and Selman, M. (2013) Latest Progress in Microsporidian Genome Research. *J. Eukaryot. Microbiol.* **60**: 309–312.

Corradi, N. and Slamovits, C.H. (2011) The intriguing nature of microsporidian genomes. *Brief. Funct. Genomics* **10**: 115–124.

Cuomo, C.A., Desjardins, C.A., Bakowski, M.A., Goldberg, J.M., Ma, A.T., Becnel, J.J., et al. (2012) Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res.* **22**: 2478–2488.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., et al. (2011) The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Delaplane, K.S. and Mayer, D.F. (2000) Crop pollination by bees. CABI Publishing, New York.

Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**: 2478–2483.

Desjardins, C.A., Sanscrainte, N.D., Goldberg, J.M., Heiman, D., Young, S., Zeng, Q., et al. (2015) Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat. Commun.* **6**: 7121.

Diogo, D., Bouchier, C., D'Enfert, C., and Bougnoux, M.-E. (2009) Loss of heterozygosity in commensal isolates of the asexual diploid yeast Candida albicans. *Fungal Genet. Biol.* **46**: 159–168.

Doležel, J. and Barto (2005) Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. *Ann. Bot.* **95**: 99–110.

Dong, S., Shen, Z., Xu, L., and Zhu, F. (2010) Sequence and Phylogenetic Analysis of SSU rRNA Gene of Five Microsporidia. *Curr. Microbiol.* **60**: 30–37.

Dowd, S.E., Gerba, C.P., and Pepper, I.L. (1998) Confirmation of the Human-Pathogenic Microsporidia Enterocytozoon bieneusi, Encephalitozoon intestinalis, and Vittaforma corneae in Water. *Appl. Environ. Microbiol.* **64**: 3332–3335.

Dussaubat, C., Brunet, J.-L., Higes, M., Colbourne, J.K., Lopez, J., Choi, J.-H., et al. (2012) Gut Pathology and Responses to the Microsporidium Nosema ceranae in the Honey Bee Apis mellifera. *PLoS One* **7**: e37017.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.

Edlund, T.D., Li, J., Visvesvara, G.S., Vodkin, M.H., McLaughlin, G.L., and Katiyar, S.K. (1996) Phylogenetic Analysis of β-Tubulin Sequences from Amitochondrial Protozoa. *Mol. Phylogenet. Evol.* **5**: 359–367.

Fast, N.M., Logsdon, J.M., and Doolittle, W.F. (1999) Phylogenetic Analysis of the TATA Box Binding Protein (TBP) Gene from Nosema locustae: Evidence for a Microsporidia-Fungi Relationship and Spliceosomal Intron Loss. *Mol. Biol. Evol.* **16**: 1415–1419.

Feder, A.F., Petrov, D.A., and Bergland, A.O. (2012) LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data. *PLoS One* **7**: e48588.

Franzen, C. (2004) Microsporidia: how can they invade other cells? *Trends Parasitol.* **20**: 275–279.

Fries, I., Feng, F., da Silva, A., Slemenda, S.B., and Pieniazek, N.J. (1996) Nosema ceranae n. sp. (Microspora, Nosematidae), Morphological and Molecular Characterization of a Microsporidian Parasite of the Asian Honey bee Apis cerana (Hymenoptera, Apidae). *Eur. J. Protistol.* **32**: 356–365.

Fries, I., Martín, R., Meana, A., García-Palencia, P., and Higes, M. (2006) Natural infections of Nosema ceranae in European honey bees. *J. Apic. Res.* **45**: 230–233.

Gallai, N., Salles, J.-M., Settele, J., and Vaissière, B.E. (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* **68**: 810–821.

Garrison, E. and Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. *Science* **342**: 357–360.

Gatehouse, H.S. and Malone, L.A. (1998) The Ribosomal RNA Gene Region of Nosema apis (Microspora): DNA Sequence for Small and Large Subunit rRNA Genes and Evidence of a Large Tandem Repeat Unit Size. *J. Invertebr. Pathol.* **71**: 97–105.

Gawad, C., Koh, W., and Quake, S.R. (2016) Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**: 175–88.

Gómez-Moracho, T., Bartolomé, C., Bello, X., Martín-Hernández, R., Higes, M., and Maside, X. (2015) Recent worldwide expansion of Nosema ceranae (Microsporidia) in Apis mellifera populations inferred from multilocus patterns of genetic variation. *Infect. Genet. Evol.* **31**: 87–94.

Gupta, S.K., Hossain, Z., Nanu, M.M., and Mondal, K. (2016) Impact of microsporidian infection on growth and development of silkworm Bombyx mori L. (Lepidoptera: Bombycidae). *Agric. Nat. Resour.* **50**: 388–395.

Higes, M., García-Palencia, P., Martín-Hernández, R., and Meana, A. (2007) Experimental infection of Apis mellifera honeybees with Nosema ceranae (Microsporidia). *J. Invertebr. Pathol.* **94**: 211–217.

Higes, M., Martín-Hernández, R., Botías, C., Bailón, E.G., González-Porto, A. V., Barrios, L., et al. (2008) How natural infection by Nosema ceranae causes honeybee colony collapse. *Environ. Microbiol.* **10**: 2659–2669.

Higes, M., Martín, R., and Meana, A. (2006) Nosema ceranae, a new microsporidian parasite in honeybees in Europe. *J. Invertebr. Pathol.* **92**: 93–95.

Higes, M., Meana, A., Bartolomé, C., Botías, C., and Martín-Hernández, R. (2013) Nosema ceranae (Microsporidia), a controversial 21st century honey bee pathogen. *Environ. Microbiol. Rep.* **5**: 17–29.

Hirt, R.P., Logsdon, J.M., Healy, B., Dorey, M.W., Doolittle, W.F., and Embley, T.M. (1999) Microsporidia are related to Fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci.* **96**: 580–585.

Huang, W.-F., Bocquet, M., Lee, K.-C., Sung, I.-H., Jiang, J.-H., Chen, Y.-W., and Wang, C.-H. (2008) The comparison of rDNA spacer regions of Nosema ceranae isolates from different hosts and locations. *J. Invertebr. Pathol.* **97**: 9–13.

Huang, W.-F., Jiang, J.-H., Chen, Y.-W., and Wang, C.-H. (2007) A Nosema ceranae isolate from the honeybee Apis mellifera. *Apidologie* **38**: 30–37.

Ironside, J.E. (2013) Diversity and Recombination of Dispersed Ribosomal DNA and Protein Coding Genes in Microsporidia. *PLoS One* **8**: e55878.

James, T.Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J.E. (2013) Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. *Curr. Biol.* **23**: 1548–1553.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., et al. (2001) Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature* **414**: 450–453.

Kearns, C.A., Inouye, D.W., and Waser, N.M. (1998) Endangered Mutualisms: The Conservation of Plant-Pollinator Interactions. *Annu. Rev. Ecol. Syst.* **29**: 83–112.

Keeling, P.J. (2009) Five Questions about Microsporidia. *PLoS Pathog.* **5**: e1000489.

Keeling, P.J. (2014) Phylogenetic Place of Microsporidia in the Tree of Eukaryotes. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 195–202.

Keeling, P.J. and Doolittle, W.F. (1996) Alpha-tubulin from Early-Diverging Eukaryotic Lineages and the Evolution of the Tubulin Family. *Mol. Biol. Evol.* **13**: 1297–1305.

Keeling, P.J. and Fast, N.M. (2002) Microsporidia: Biology and Evolution of Highly Reduced Intracellular Parasites. *Annu. Rev. Microbiol.* **56**: 93–116.

Keeling, P.J., Fast, N.M., and Corradi, N. (2014) Microsporidian Genome Structure and Function. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 221–229.

Keeling, P.J. and McFadden, G.I. (1998) Origins of microsporidia. *Trends Microbiol.* **6**: 19–23.

Khanduja, S., Ghoshal, U., Agarwal, V., Pant, P., and Ghoshal, U.C. (2017) Identification and genotyping of Enterocytozoon bieneusi among human immunodeficiency virus infected patients. *J. Infect. Public Health* **10**: 31–40.

Klee, J., Besana, A.M., Genersch, E., Gisder, S., Nanetti, A., Tam, D.Q., et al. (2007) Widespread dispersal of the microsporidian Nosema ceranae, an emergent pathogen of the western honey bee, Apis mellifera. *J. Invertebr. Pathol.* **96**: 1–10.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., et al. (2011) PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS One* **6**: e15925.

Kofler, R., Pandey, R.V., and Schlötterer, C. (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**: 3435–3436.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**: 1870–1874.

LaFave, M.C. and Sekelsky, J. (2009) Mitotic Recombination: Why? When? How? Where? *PLoS Genet.* **5**: e1000411.

Lamour, K.H., Mudge, J., Gobena, D., Hurtado-Gonzales, O.P., Schmutz, J., Kuo, A., et al. (2012) Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen Phytophthora capsici. *Mol. Plant-Microbe Interact.* **25**: 1350–1360.

Lee, S.C., Corradi, N., Byrnes, E.J., Torres-Martinez, S., Dietrich, F.S., Keeling, P.J., and Heitman, J. (2008) Microsporidia Evolved from Ancestral Sexual Fungi. *Curr. Biol.* **18**: 1675–1679.

Lee, S.C., Corradi, N., Doan, S., Dietrich, F.S., Keeling, P.J., and Heitman, J. (2010) Evolution of the sex-Related Locus and Genomic Features Shared in Microsporidia and Fungi. *PLoS One* **5**: e10539.

Lee, S.C. and Ironside, J.E. (2014) Sex and the Microsporidia. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 231–243.

Lee, S.C., Weiss, L.M., and Heitman, J. (2009) Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. *Commun. Integr. Biol.* **2**: 414–417.

Li, H. and 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., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Martín-Hernández, R., Meana, A., Prieto, L., Salvador, A.M., Garrido-Bailón, E., and Higes, M. (2007) Outcome of Colonization of Apis mellifera by Nosema ceranae. *Appl. Environ. Microbiol.* **73**: 6331–6338.

Medici, S.K., Sarlo, E.G., Porrini, M.P., Braunstein, M., and Eguaras, M.J. (2012) Genetic variation and widespread dispersal of Nosema ceranae in Apis mellifera apiaries from Argentina. *Parasitol. Res.* **110**: 859–864.

de Miranda, J.R., Cordoni, G., and Budge, G. (2010) The Acute bee paralysis virus–Kashmir bee virus–Israeli acute paralysis virus complex. *J. Invertebr. Pathol.* **103**: S30–S47.

Nägeli, K.W. (1857) Uber die neue Krankheit der Seidenraupe und verwandte Organismen. *Oesterreichische Bot. Zeitschrift* **15**: 760–761.

Nakamura, L.K. (1996) Paenibacillus apiarius sp. nov. *Int. J. Syst. Bacteriol.* **46**: 688–693.

Nakjang, S., Williams, T.A., Heinz, E., Watson, A.K., Foster, P.G., Sendra, K.M., et al. (2013) Reduction and Expansion in Microsporidian Genome Evolution: New Insights from Comparative Genomics. *Genome Biol. Evol.* **5**: 2285–2303.

Ndikumana, S., Pelin, A., Williot, A., Sanders, J.L., Kent, M.L., and Corradi, N. (2017) Genome Analysis of Pseudoloma neurophilia : A Microsporidian Parasite of Zebrafish (Danio rerio). *J. Eukaryot. Microbiol.* **64**: 18–30.

Nebavi, F., Ayala, F.J., Renaud, F., Bertout, S., Eholie, S., Moussa, K., et al. (2006) Clonal population structure and genetic diversity of Candida albicans in AIDS patients from Abidjan (Côte d'Ivoire). *Proc. Natl. Acad. Sci.* **103**: 3663–3668.

Nei, M. (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., et al. (2013) Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. *J. Comput. Biol.* **20**: 714–737.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**: 419–420.

Paxton, R.J., Klee, J., Korpela, S., and Fries, I. (2007) Nosema ceranae has infected Apis mellifera in Europe since at least 1998 and may be more virulent than Nosema apis. *Apidologie* **38**: 558–565.

Pelin, A., Moteshareie, H., Sak, B., Selman, M., Naor, A., Eyahpaise, M.-È., et al. (2016) The genome of an Encephalitozoon cuniculi type III strain reveals insights into the genetic diversity and mode of reproduction of a ubiquitous vertebrate pathogen. *Heredity* **116**: 458–465.

Pelin, A., Selman, M., Aris-Brosou, S., Farinelli, L., and Corradi, N. (2015) Genome analyses suggest the presence of polyploidy and recent human-driven expansions in eight global populations of the honeybee pathogen Nosema ceranae. *Environ. Microbiol.* **17**: 4443–4458.

Pettis, J.S., VanEngelsdorp, D., Johnson, J., and Dively, G. (2012) Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. *Naturwissenschaften* **99**: 153–158.

Plischuk, S., Martín-Hernández, R., Prieto, L., Lucía, M., Botías, C., Meana, A., et al. (2009) South American native bumblebees (Hymenoptera: Apidae) infected by Nosema ceranae (Microsporidia), an emerging pathogen of honeybees (Apis mellifera). *Environ. Microbiol. Rep.* **1**: 131–135.

Pombert, J.-F., Selman, M., Burki, F., Bardell, F.T., Farinelli, L., Solter, L.F., et al. (2012) Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. *Proc. Natl. Acad. Sci.* **109**: 12638–12643.

Pombert, J.-F., Xu, J., Smith, D.R., Heiman, D., Young, S., Cuomo, C.A., et al. (2013) Complete Genome Sequences from Three Genetically Distinct Strains Reveal High Intraspecies Genetic Diversity in the Microsporidian Encephalitozoon cuniculi. *Eukaryot. Cell* **12**: 503–511.

Ptaszyńska, A.A., Borsu, G., Mułenko, W., and Olszewski, K. (2012) Monitoring of nosemosis in the Lublin region and preliminary morphometric studies of Nosema spp. spores. *Med. Weter.* **68**: 622–625.

Quandt, C.A., Beaudet, D., Corsaro, D., Walochnik, J., Michel, R., Corradi, N., and James, T.Y. (2017) The genome of an intranuclear parasite, Paramicrosporidium saccamoebae, reveals alternative adaptations to obligate intracellular parasitism. *Elife* **6**: 1–19.

Rajendran, K.V., Shivam, S., Ezhil Praveena, P., Joseph Sahaya Rajan, J., Sathish Kumar, T., Avunje, S., et al. (2016) Emergence of Enterocytozoon hepatopenaei (EHP) in farmed Penaeus (Litopenaeus) vannamei in India. *Aquaculture* **454**: 272–280.

Rogers, M.B., Downing, T., Smith, B.A., Imamura, H., Sanders, M., Svobodova, M., et al. (2014) Genomic Confirmation of Hybridisation and Recent Inbreeding in a Vector-Isolated Leishmania Population. *PLoS Genet.* **10**: e1004092.

Rosenberg, S.M. (2011) Stress-Induced Loss of Heterozygosity in Candida: a Possible Missing Link in the Ability to Evolve. *MBio* **2**: e00200-11.

Rosenkranz, P., Aumeier, P., and Ziegelmann, B. (2010) Biology and control of Varroa destructor. *J. Invertebr. Pathol.* **103**: S96–S119.

Roudel, M., Aufauvre, J., Corbara, B., Delbac, F., and Blot, N. (2013) New insights on the genetic diversity of the honeybee parasite Nosema ceranae based on multilocus sequence analysis. *Parasitology* **140**: 1346–1356.

Sagastume, S., del Águila, C., Martín-Hernández, R., Higes, M., and Henriques-Gil, N. (2011) Polymorphism and recombination for rDNA in the putatively asexual microsporidian Nosema ceranae, a pathogen of honeybees. *Environ. Microbiol.* **13**: 84–95.

Sagastume, S., Martín-Hernández, R., Higes, M., and Henriques-Gil, N. (2016) Genotype diversity in the honey bee parasite Nosema ceranae: multi-strain isolates, cryptic sex or both? *BMC Evol. Biol.* **16**: 216.

Sagastume, S., Martín-Hernández, R., Higes, M., and Henriques-Gil, N. (2014) Ribosomal Gene Polymorphism in Small Genomes: Analysis of Different 16S rRNA Sequences Expressed in the Honeybee Parasite Nosema ceranae (Microsporidia). *J. Eukaryot. Microbiol.* **61**: 42–50.

Selman, M. and Corradi, N. (2011) Microsporidia: Horizonal gene transfers in vicious parasites. *Mob. Genet. Elements* **1**: 251–292.

Selman, M., Sak, B., Kváč, M., Farinelli, L., Weiss, L.M., and Corradi, N. (2013) Extremely Reduced Levels of Heterozygosity in the Vertebrate Pathogen Encephalitozoon cuniculi. *Eukaryot. Cell* **12**: 496–502.

seqtk: Toolkit for processing sequences in FASTA/Q formats. Available from: https://github.com/lh3/seqtk.

Shafer, A.B.A., Williams, G.R., Shutler, D., Rogers, R.E.L., and Stewart, D.T. (2009) Cophylogeny of Nosema (Microsporidia: Nosematidae) and bees (Hymenoptera: Apidae) suggests both cospeciation and a host-switch. *J. Parasitol.* **95**: 198–203.

Shigano, T., Hatakeyama, Y., Nishimoto, N., Watanabe, M., Yamamoto, Y., Wijonarko, A., et al. (2015) Variety and diversity of microsporidia isolated from the common cutworm Spodoptera litura in Chichijima, Ogasawara Islands. *J. Insect Biotechnol. Sericology* **84**: 69–73.

Slamovits, C.H., Fast, N.M., Law, J.S., and Keeling, P.J. (2004) Genome Compaction and Stability in Microsporidian Intracellular Parasites. *Curr. Biol.* **14**: 891–896.

Solter, L.F. (2014) Epizootilogy of Microsporidiosis in Invertebrate Hosts. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 165–194.

Stewart, K.D. and Mattox, K. (1980) Phylogeny of phytoflagellates. In, Cox,E.R. (ed), *Phytoflagellates, Development in Marine Biology*. Elsevier/North Holland, Amsterdam, pp. 433–462.

Suwannapong, G., Benbow, M.E., and Nieh, J.C. (2011) Biology of Thai honeybees: natural history and threats. Florio,R.M. (ed) Nova Science Publishers, Inc., New York.

Suwannapong, G., Yemor, T., Boonpakdee, C., and Benbow, M.E. (2011) Nosema ceranae, a new parasite in Thai honeybees. *J. Invertebr. Pathol.* **106**: 236–241.

Tanabe, Y., Watanabe, M.M., and Sugiyama, J. (2002) Are Microsporidia really related to Fungi?: a reappraisal based on additional gene sequences from basal fungi. *Mycol. Res.* **106**: 1380–1391.

Tsaousis, A.D., Kunji, E.R.S., Goldberg, A. V., Lucocq, J.M., Hirt, R.P., and Embley, T.M. (2008) A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi. *Nature* **453**: 553–556.

VanEngelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., et al. (2009) Colony Collapse Disorder: A Descriptive Study. *PLoS One* **4**: e6481.

VanEngelsdorp, D. and Meixner, M.D. (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *J. Invertebr. Pathol.* **103**: S80–S95.

Vávra, J. and Larsson, J.I.R. (2014) Structure of Microsporidia. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 1–70.

Vávra, J. and Lukeš, J. (2013) Microsporidia and "The Art of Living Together". *Adv. Parasitol.* **82**: 253–319.

Vivarès, C.P., Gouy, M., Thomarat, F., and Méténier, G. (2002) Functional and evolutionary analysis of a eukaryotic parasitic genome. *Curr. Opin. Microbiol.* **5**: 499–505.

Vossbrinck, C.R. and Debrunner-Vossbrinck, B.A. (2005) Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol.* **52**: 131–142.

Vossbrinck, C.R., Debrunner-Vossbrinck, B.A., and Weiss, L.M. (2014) Phylogeny of the Microsporidia. In, Weiss,L.M. and Becnel,J.J. (eds), *Microsporidia: Pathogens of Opportunity*. John Wiley & Sons, Inc., Chichester, UK, pp. 203–220.

Watson, A.K., Williams, T.A., Williams, B.A.P., Moore, K.A., Hirt, R.P., and Embley, T.M. (2015) Transcriptomic profiling of host-parasite interactions in the microsporidian Trachipleistophora hominis. *BMC Genomics* **16**: 983.

Watterson, G.A. (1975) On the Number of Segregating Sites in Genetical Models without Recombination. *Theor. Popul. Biol.* **7**: 256–276.

Whipps, C.M. and Kent, M.L. (2006) Polymerase Chain Reaction Detection of Pseudoloma neurophilia, a Common Microsporidian of Zebrafish (Danio rerio) Reared in Research Laboratories. *J. Am. Assoc. Lab. Anim. Sci.* **45**: 36–39.

Wickam, H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Williams, B.A.P., Hirt, R.P., Lucocq, J.M., and Embley, T.M. (2002) A mitochondrial remnant in the microsporidian Trachipleistophora hominis. *Nature* **418**: 865–869.

Williams, B.A.P., Slamovits, C.H., Patron, N.J., Fast, N.M., and Keeling, P.J. (2005) A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc. Natl. Acad. Sci.* **102**: 10936–10941.

Williams, G.R., Shafer, A.B.A., Rogers, R.E.L., Shutler, D., and Stewart, D.T. (2008) First detection of Nosema ceranae, a microsporidian parasite of European honey bees (Apis mellifera), in Canada and central USA. *J. Invertebr. Pathol.* **97**: 189–192.

Williams, T.A., Nakjang, S., Campbell, S.E., Freeman, M.A., Eydal, M., Moore, K.A., et al. (2016) A Recent Whole-Genome Duplication Divides Populations of a Globally Distributed Microsporidian. *Mol. Biol. Evol.* **33**: 2002–2015.

Xu, Y. and Weiss, L.M. (2005) The microsporidian polar tube: A highly specialised invasion organelle. *Int. J. Parasitol.* **35**: 941–953.

Zander, E. (1909) Tierische Parasiten als Krankenheitserreger bei der Biene. *Münchener Bienenzeitung* **31**: 196–204.

Zhang, W., Ren, G., Zhao, W., Yang, Z., Shen, Y., Sun, Y., et al. (2017) Genotyping of Enterocytozoon bieneusi and Subtyping of Blastocystis in Cancer Patients: Relationship to Diarrhea and Assessment of Zoonotic Transmission. *Front. Microbiol.* **8**: 1835.