Norwegian University
of Life Sciences

Master's Thesis 2019   30 ECTS
Faculty of Veterinary Medicine

Characterization and genetic investigations of the mitochondrial DNA from poultry red mite *Dermanyssus gallinae*

Anojan Rajagopal

MSc Chemistry and Biotechnology

(Cover image of *Dermanyssus gallinae* photo Ø Øines)

Acknowledgements

Oslo, 15 May 2019

Anojan Rajagopal

Abstract

The poultry red mite, *Dermanyssus gallinae,* is a global problem in the egg-laying industry. This ectoparasite has a rapid rate of proliferation with a negative impact on the birds' health, welfare and productivity resulting in severe economic consequences for poultry farmers.

Our knowledge of the pathways of this parasite is largely dependent on the fact that we have effective genetic tools for infection detection. Up to now, mitochondrial cytochrome oxidase subunit 1 has been the most widely used fragment that has been investigated for poultry red mite, and we have shown that in Norway there are two haploid groups of poultry red mite (A16 & B9), which are the most common variants.

In this thesis, we performed Nanopore MinION sequencing, for *de novo* assembly of the mitochondrial DNA of poultry red mite. In addition to the newly generated Nanopore reads, Illumina HiSeq-data from a previous project were used to complement it. Before doing Nanopore sequencing, DNA from the various haplotypes was amplified through the GenomiPhi kit.

The MinION sequencing run generated data with some depth, but with high error rate as expected. A partial mitochondrial gene arrangement in *Dermanyssus gallinae,* with 8 out of possible 13 protein-encoding genes was determined. In addition, the order of the 6 genes found in the mitochondria, was determined.

Sammendrag

Rød hønsemidd, *Dermanyssus gallinae*, er en av de aller viktigste parasittene knyttet til eggproduksjon i verden. Hvert år forårsaker midden et økonomisk tap hos eggprodusentene som bare i Europa er antatt å ligge på over 230 millioner euro. Når et hønsehus blir smittet av denne blodsugende midden, er det veldig vanskelig å behandle og bli kvitt den. Dette kan føre til vedvarende problemer i mange år.

Vår kjennskap til denne parasittens smitteveier er i stor grad avhengig av at vi har effektive genetiske verktøy for smittesporing. Frem til nå er mtCO1 stort sett vært det mest brukte fragmentet som er blitt undersøkt for hønsemidd, og vi har vist i Norge at det er to haplogrupper av hønsemidd (A16 & B9), som er de mest vanlige variantene.

I denne masteroppgaven ble det utført opplæring av standard metodikk for identifikasjon ved hjelp av molekylære verktøy, uttesting av ulike ekstraksjonsmetoder for å få fremskaffet mest mulig DNA fra enkelt midd, samt Nanopore MinION sekvensering sammen med undersøkelser av tilgjengelige sekvensdata fra tidligere Illumina HiSeq-data, i et forsøk på å sammenstille det mitokondrielle genomet fra rød hønsemidd. Før vi kjørte Nanopore sekvensering, ble DNA fra de ulike haplotypene amplifisert opp gjennom GenomiPhi.

MinION sekvenseringen genererte et datasett med noe dybde, men med høy feilrate som forventet. Et delvis mitokondrielt genarrangement i Dermanyssus gallinae ble bestemt. Av de 13 proteinkodede genene, ble 8 gener funnet, og kun 6 gener ble bestemt ut i fra rekkefølgen i mitokondriet.

Abbreviations

- CO1 – The mitochondrial cytochrome oxidase subunit 1

- COST – European Cooperation in Science and Technology

- DNA – Deoxyribonucleic acid

- dNTPs – Deoxynucleic triphosphates

- dsDNA – Double-stranded deoxyribonucleic acid

- gDNA – Genomic DNA

- HTS – High – throughput sequencing

- ITS – Internal transcribed spacer

- mtDNA – Mitochondrial DNA

- mtSSB - Mitochondrial single-stranded DNA-binding protein

- NGS – Next Generation Sequencing

- NVI – Norwegian Veterinary Institute

- ONT – Oxford Nanopore Technologies

- PCR – Polymerase chain reaction

- POLγ – DNA polymerase γ

- PRM – Poultry red mite

- WGS – Whole genome sequencing

List of figures

List of tables

Table of content

1.0 Introduction

1.1 Poultry red mite – a problem in Europe

Poultry red mite (PRM *Dermanyssus gallinae*, is a blood-sucking parasite, considered to be one of the most important parasites related to egg production in the world(Chauve 1998). It is a major worldwide concern in the egg-laying industry. Every year the mite causes a financial loss in egg production. According to Van Emous(Emous 2017) only in Europe, the damage caused by PRM infestation is estimated at around 231 million euro. This value is almost twice the estimation he made in 2005 of 130 million euro**.** An epidemiological review from COST (European Cooperation in Science and Technology), shows that 83 % or more of the European farms are infested by *D. gallinae*. This prevalence reaches 94 % in the Netherlands, Germany and Belgium. Scandinavian countries have relatively less, especially Norway with only around 11 % of PRM in farms(Mul 2013), although in recent years this has risen to about 25 % in 2018 (Animalia PRM surveillance programme). Regarding PRM prevalence outside Europe, there have also been reported high PRM infestations in Africa, Asia and Latin America (Gharbi, Sakly et al. 2013, Chu, Murano et al. 2015).

Van Emous (2017) explains this higher damage by the change of traditional cages to alternative housing systems, longer production life-cycles of the animals, and the ban of beak trimming. The consequences of the PRM infestations in the egg-laying industry include primarily a negative impact on feed conversation rate, a reduction in egg production, an increase in downgraded eggs, a higher susceptibility to poultry diseases, and more dead animals(Mul 2013).

The PRM infestation has been increasing in Europe for the past decade and is expected to further increase. One of the factors contributing to this is as mentioned, the transformation of housing systems in laying hen husbandry in EU member countries. Traditional cages for poultry birds have been banned since 2012**.** These more complex environments in the new housing systems appear to favour mite proliferation of red mite infestations. This is because it provides more hiding possibilities, thus enabling them to easily escape control measures (Sigognault Flochlay, Thomas et al. 2017).

1.2 Life Cycle of *Dermanyssus gallinae*

To better understand the poultry red mite, an understanding of its life cycle is necessary. Figure 1.1 shows an illustration of the poultry red mite at various life stages. What's interesting about this mite is that it spends most of its life away from the host. Unlike the other poultry mite, northern fowl mite, *Ornithonyssus sylvarum,* which spends its entire life on the host (Mullens, Chen et al. 2010). Poultry red mites can be tricky to find and can hide very well. In farms with battery-cage systems, mites are hidden under the conveyor belts of eggs and cage supports. In farms with slatted flooring systems, mites are hidden under the rods, in nest boxes, beneath troughs and in small cracks and crevices in the poultry house walls(Chauve 1998).
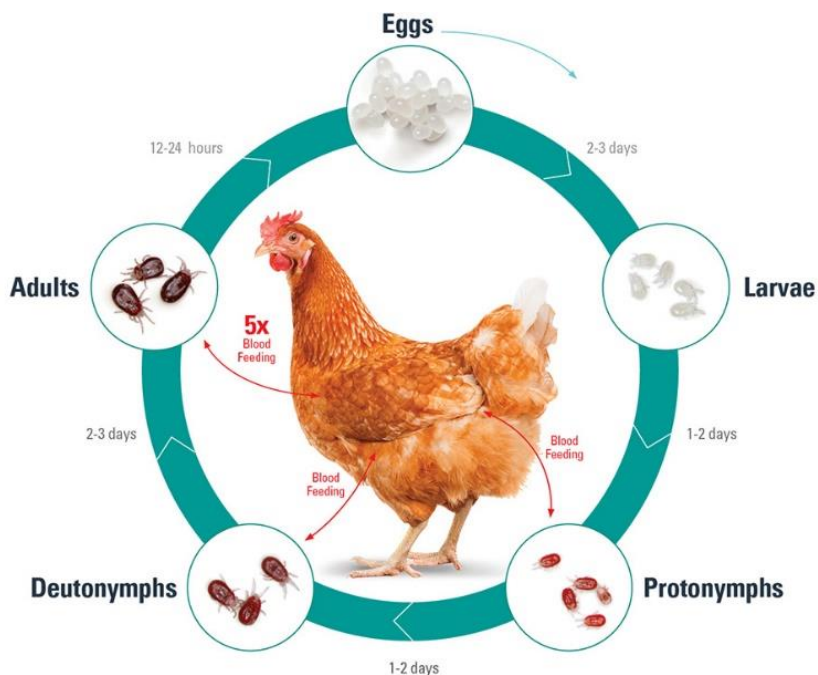
Figure 1.1. The life cycle of the poultry red mite, *Dermanyssus gallinae*. Aside from the egg, the poultry red mites have 4 lifecycle stages: larvae, protonymph, deutonymph, and adult. Image from (Sparagano, George et al. 2014).

Adult mites measure about 1 mm long, weigh about 76 µg(unfed) to 280 µg(after a blood meal), and are red in colour after feeding, but they appear grey or white without host blood in their system(Roberts 2009)**.** This mite does not only feed on the blood of poultry birds but also aviary and wild birds. It also occasionally bites mammals including man, and can thus constitute a problem, to people working in affected poultry places.

PRM typically feeds at night and stay on the birds for only 0.5-1.5 hours while feeding. It is important to note that nymphs and females suck blood while males do it only occasionally. Aside from the egg, the mite has 4 stages in its life cycle. Under warm conditions (28-30C) a larva emerges in 2-3 days. This young larva has 6 legs and does not feed. After 1-2 day it moults to the protonymphstage. The protonymph has 8 legs; it feeds and molts to a deutonymph which then feeds again before becoming an adult. In optimal conditions, the life cycle (egg-to-egg) can be as short as 7 days, which allows for the rapid growth of mite populations. PRM may live up to 8 months away from poultry host without feeding(Chauve 1998).

1.3 PRM infestation effects

In addition to the significant economic damage caused by the PRM infestation, another concern is the effects induced by PRM parasitism on the birds' health and welfare (Sigognault Flochlay, Thomas et al. 2017). Figure 1.2 shows the different effects of PRM infestation.
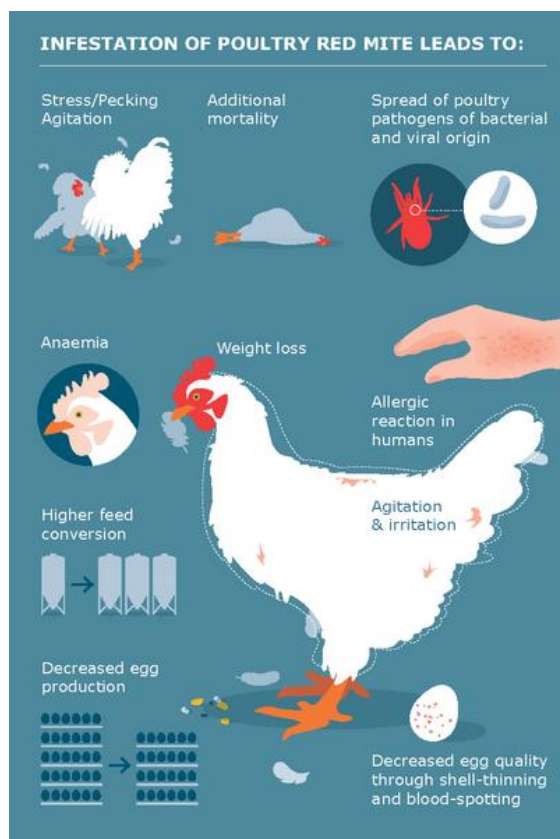
Figure 1.2: The effect of a poultry red mite infestation (Image courtesy of Wageningen Livestock Research).

Infected chickens can develop anemia due to repeated bites. In extreme case, PRM infestation burdens may be so high, the hens become severely anemic and die from blood loss alone(Chauve 1998)**.** Other effects by mite feeding are significant stress to hens, increased feed and water intake and decreased bird condition. This will also impact the production by causing a decline in egg quality (through shell thinning and spotting) and egg laying. There have also been reported aggressive feather-pecking and even cannibalistic behaviors due to the infestation(Kilpinen, Roepstorff et al. 2005). In addition to its effect on chicken's health and welfare, there is a growing impact of red mite infestation on public health. Poultry red mite can also serve as a disease vector for numerous pathogens such as *Salmonella, E. coli* and *Staphylococcus*(Valiente Moro, De Luna et al. 2009).

1.4 Control of PRM infestations

As mentioned previously, the PRM has for decades been a threat to the egg production industry. This has led to an increased research activity dedicated to controlling PRM. Successful treatments of mite infestations remains a major goal. A very limited number of chemical treatments are currently available to treat mite infestations.

Many conventional mite products have been withdrawn from European markets or banned in the past few years, because they did not comply with European or national regulatory requirements for human consumer and user safety. The organophosphate phoxim (Bymite, Bayer) is one of the few veterinary medical products registered in Europe for the treatment of PRM infestations (since 2010). Recently Exzolt from MSD was introduced as another treatment alternative, where the chemical is added in the drinking water of the birds. However, these compounds can be expensive to use correctly and not all are licensed in all countries (European Medicines Agency, 2017). Additional acaricidal spray products are available in some European countries. These sprays are mainly used during the unoccupied cleaning period between two flocks, for the treatment of the poultry house and equipment (Sigognault Flochlay, Thomas et al. 2017).

Another way is using non-chemical methods like silica-based products. On the other hand, this can be very harmful to users and animal due to the irritation of inhaled silica particles.

Predator mites are another method used, but has yet shown satisfactory efficacy(Sparagano, George et al. 2014). Another control option is heating up the house, which has been described as effective, but a very expensive approach. The development of new vaccine-based control strategies is also a promising approach, although successful long-lived vaccine response can be challenging (Price, Kuster et al. 2019). In spite of all these different solutions for controlling the mite infestation, the development of more useful, effective and innovative treatments remain in great demand. This unmet veterinary medical need has clearly been recognized by the scientific community, key opinion leader groups, the layer industry, and policy makers such as the European Union.

1.5 Mitochondria – background

Since we are going to investigate PRM using the mitochondrial (mt) DNA, an increased understanding of the mtDNA is necessary. The mitochondrion is an independent genetic system in each eukaryotic cell outside the nuclear genome**.** The mitochondrial genome contains genes that are important for cellular energetics and survival(Yasukawa and Kang 2018). Because of the increased awareness on the importance of metabolism and bioenergetics in a wide variety of human diseases, more and more mt DNA studies are performed to study the link between mtDNA sequence variation and disease development.

Mitochondria was once a free-living bacteria billions of years ago. During the evolution, the primitive eukaryotic cell endocytosed bacteria by endosymbiotically and domesticated inside the cytoplasm in a way that nucleus took some of the mitochondrial genes into it. That led to mitochondria being no longer free living, and eventually evolved into an integral part of its host. In other words, the eukaryotic cell slaved aerobic bacteria permanently. The mitochondria almost certainly evolved from bacteria. This theory is called the endosymbiotic theory(Vogt 2017).
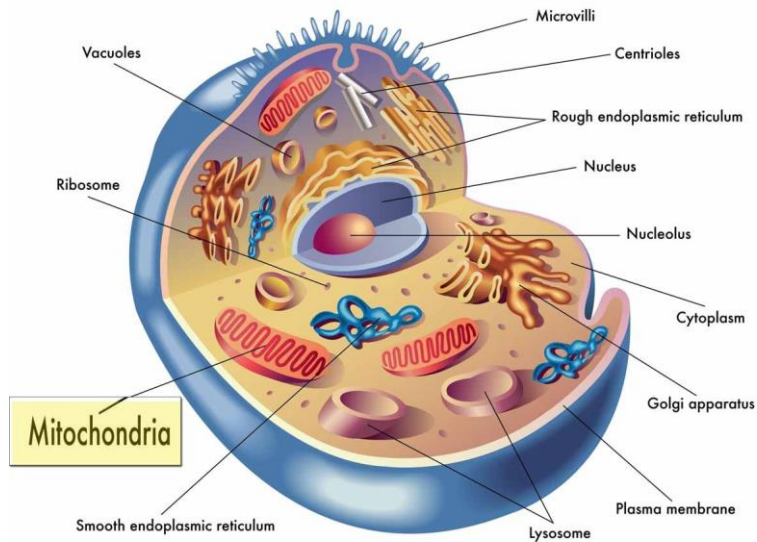
Figure 1.3: Eukaryotic cell that contains mitochondria and other cell organelles. Image from (Reyes, Valim et al. 1996).

Eukaryotic cells contain a nucleus and other organelles (Figure 1.3). Organelles are "little organs" in the cell that carry out specific functions. Mitochondria is one of these organelles that generate energy for the cell in the form of ATP**.** Mitochondria produce ATP (energy) in far greater quantities than can be made in the cytosol of the cell**.** This is why the mitochondrion is known as the "powerhouse of the cell(Yasukawa and Kang 2018). However, mitochondria are unique organelles because they have their own genetic material – independent to the mitochondria found in the nucleus.

Human mtDNA is a double-stranded molecule of 16.6 kb (Falkenberg 2018). Mitochondrial genomes of animals differ in size, but they typically contain 37 genes (Table 1.1) that code for 13 proteins, two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes as well as containing non-coding regions of variable lengths. The proteins are essential for the respiratory chain, while the rRNA genes and tRNA genes are involved in the mitochondrial protein synthesis (Shokolenko, Wilson et al. 2014).

Table 1.1: Position and length (nt) of genes in the mitochondrion of *Stylochyrus rarior.* Table from *(Swafford and Bond 2009)*.

Gene	Gene product	Position	Length	Strand
trnM	Transfer RNA methionine	1–63	63	+
nad2	NADH dehydrogenase subunit 2	64–1026	963	+
trnW	Transfer RNA tryptophan	1025–1085	61	+
trnC	Transfer RNA cysteine	1078–1140	63	−
trnY	Transfer RNA tyrosine	1129–1190	62	−
cox1	Cytochrome *c* oxidase subunit 1	1202–2725	1524	+
cox2	Cytochrome *c* oxidase subunit 2	2725–3400	676	+
trnK	Transfer RNA lysine	3401–3463	63	+
trnD	Transfer RNA aspartic acid	3464–3524	61	+
atp8	ATP synthase F0 subunit 8	3525–3683	159	+
atp6	ATP synthase F0 subunit 6	3680–4345	666	+
cox3	Cytochrome *c* oxidase subunit 3	4359–5123	765	+
trnG	Transfer RNA glycine	5124–5186	63	+
nad3	NADH dehydrogenase subunit 3	5187–5525	339	+
trnA	Transfer RNA alanine	5525–5586	62	+
trnR	Transfer RNA arginine	5587–5645	59	+
trnN	Transfer RNA asparagine	5646–5706	61	+
trnS1	Transfer RNA serine 1	5707–5760	54	+
trnF	Transfer RNA phenylalanine	5761–5821	61	+
trnE	Transfer RNA glutamic acid	5820–5879	60	−
nad5	NADH dehydrogenase subunit 5	5880–7548	1669	−
trnH	Transfer RNA histidine	7549–7610	62	−
nad4	NADH dehydrogenase subunit 4	7606–8934	1329	−
nad4 L	NADH dehydrogenase subunit 4 L	8936–9208	273	−
trnT	Transfer RNA threonine	9225–9284	60	+
trnP	Transfer RNA proline	9285–9347	63	−
nad6	NADH dehydrogenase subunit 6	9335–9781	447	+
cob	Apocytochrome *b*	9788–10 898	1111	+
trnS2	Transfer RNA serine 2	10 899–10 961	63	+
16S	Large subunit rRNA	10 962–12 161	1200	−
trnV	Transfer RNA valine	12 162–12 233	72	−
12S	Small subunit rRNA	12 234–12 886	653	−
cr1	Control region 1	12 887–13 357	471	N/A
nad1	NADH dehydrogenase subunit 1	13 358–14 260	903	−
trnL1	Transfer RNA leucine 1	14 261–14 319	59	−
trnL2	Transfer RNA leucine 2	14 313–14 376	64	−
cr2	Control region 2	14 377–14 775	399	N/A
trnI	Transfer RNA isoleucine	14 776–14 836	61	+
trnQ	Transfer RNA glutamine	14 837–14 899	63	−

MtDNA is different from nuclear DNA in a lot of ways. It is circular, whereas nuclear DNA is linear. Figure 1.4 shows the mitochondrial genome maps of human and the mite *Stylochyrus rarior,* both consisting of all the 37 genes mentioned in Table 1.1. *Stylochyrus rarior* is from the same order (Mesostigmata) as *Dermanyssus gallinae*.
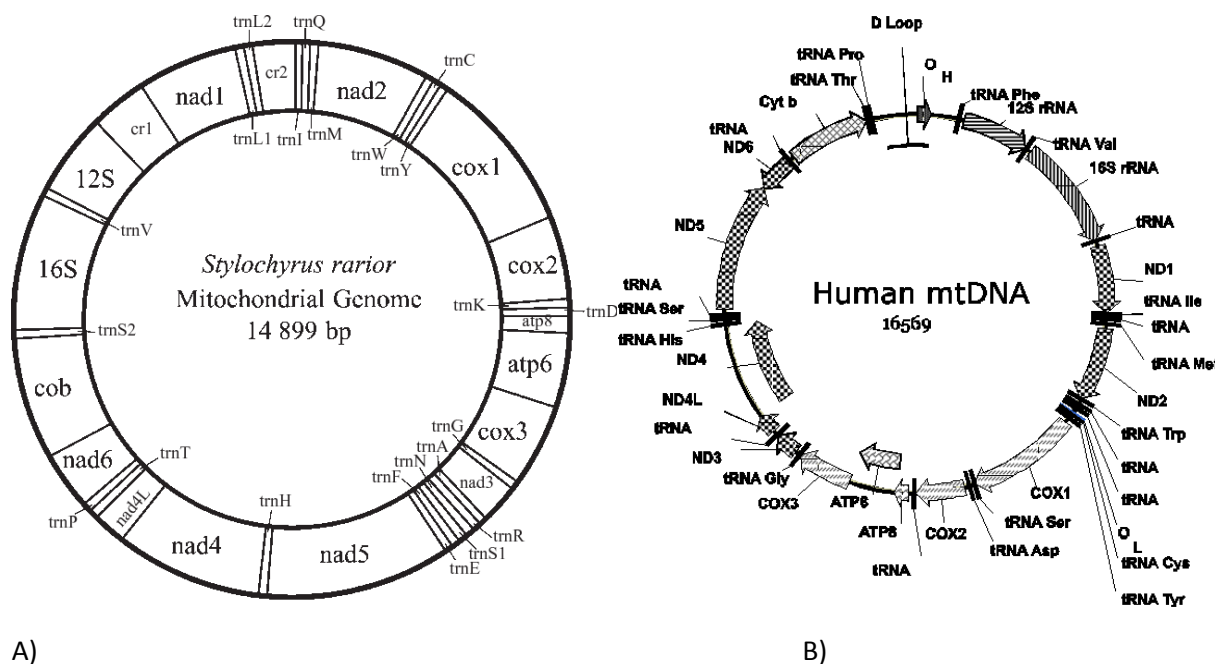
Figure 1.4: Mitochondrial genome map for A) *Stylochyrus rarior (Swafford and Bond 2009)*. B) human(Shokolenko, Wilson et al. 2014). Gene regions encoding proteins, rRNAs and tRNAs are indicated.

Nuclear DNA carries all of an organism's genetic information, which is used for growth, development, functioning and reproduction. It is important to remember that there is only one nucleus per cell, where the DNA is tightly packed into chromosomes. Human nuclear DNA consists of 46 chromosomes where you inherit 23 from your mother and 23 from your father. However, a single cell can have multiple- mitochondria and each of them 2-10 of copies of the mitochondrial genome. Unlike nuclear DNA, mtDNA consists of only one chromosome (Xia, Liu et al. 2017).

Nuclear DNA is inherited from both the mother and father, whereas mtDNA is mainly maternal. That means you never inherit mtDNA from your father (Yasukawa and Kang 2018). This is why the theory of mitochondrial Eve has developed – this is the last female that would be an ancestor of everyone on the planet, hence every human possesses direct ancestral relationship of her mitochondria (and her nuclear DNA). Since mtDNA can be regarded as solely maternal, it does not undergo recombination like nuclear DNA. This means that while the nuclear DNA present in a cell is the product of a shuffle of your parents' DNA, all changes in mtDNA have to come from mutations.

1.6 Mitochondrial DNA Replication

The mitochondrion also undergoes replication separately from nuclear DNA. The two strands of mtDNA are distinguished by nucleotide composition and are called heavy (H)- and light(L)-strands(Yasukawa and Kang 2018). Mammalian mtDNA is replicated by proteins different from those used for nuclear DNA replication. DNA polymerase γ (POLγ) is the replicative polymerase in mitochondria. This DNA polymerase is extremely accurate with very few misincorporations. POLγ cannot use double-stranded DNA as a template and a DNA helicase is therefore required at the mitochondrial replication fork. This DNA helicase is called TWINKLE. During the mtDNA replication, TWINKLE travels in front of POLγ, unwinding the double-stranded DNA template. The TWINKLE forms a hexamer and requires a fork structure (a single-stranded 5'- DNA loading site and a short 3'-tail) to load and initiate unwinding. Mitochondrial single-stranded DNA-binding protein (mtSSB) binds to the formed ssDNA, protects it against nucleases, and prevents secondary structure formation. MtSSB prevents the mitochondrial RNA polymerase (POLRMT) from initiating random RNA synthesis on the displaced strand. MtSSB enhances mtDNA synthesis by stimulating TWINKLE's helicase activity as well as increasing the processivity of POLγ(Falkenberg 2018).
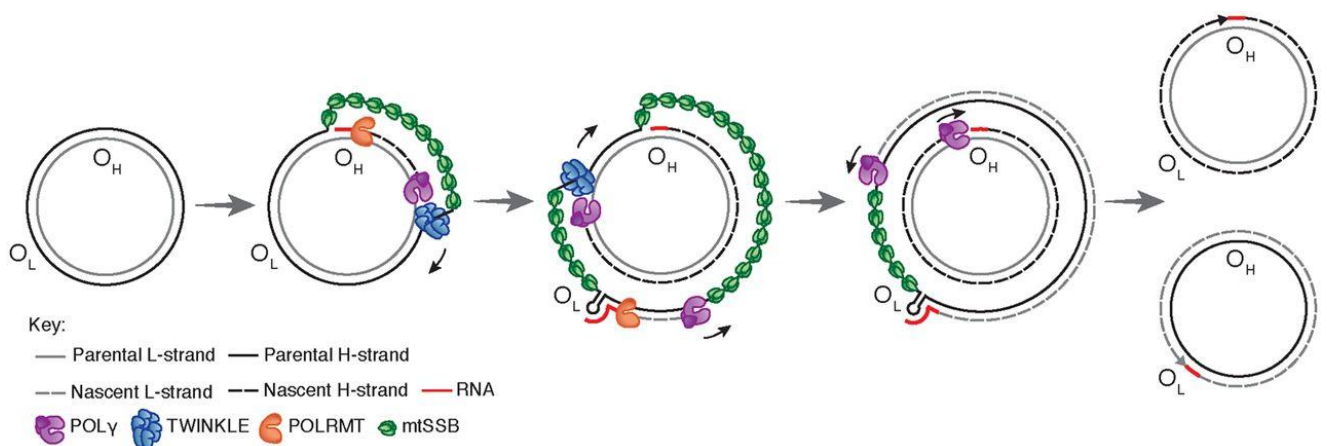
Figure 1.5: Mitochondrial DNA replication in mammalian cells that involves mitochondria-specific proteins, such as DNA polymerase γ (POLγ), TWINKLE DNA helicase, mitochondrial single-stranded DNA binding protein (mtSSB) and mitochondrial RNA polymerase (POLRMT). O_H and O_L are the major replication initiation sites of H- and L-strands under mtDNA replication model. Image from (Falkenberg 2018).

1.7 Sequencing technologies

Over the last decade, improvements in next-generation DNA sequencing technology has transformed the field of genomics, making it an important tool in modern genetic and clinical research laboratories (Dapprich, Ferriola et al. 2016). The ability to sequence whole genomes or specific genomic regions of interest gives a new insight into a variety of applications such as human health and disease, metagenomics and evolutionary biology(Goodwin, McPherson et al. 2016). One of the advantages of long-read sequencing is identifying a large structural variation in a number of regions, which is not possible through short-read sequencing like traditional, old Sanger sequencing.

1.7.1 Sanger sequencing – first generation sequencing

A biochemist called Fredrick Sanger, was the first person to sequence a genome containing single-stranded DNA (ssDNA) from a bacteriophage in 1977(Sanger, Nicklen et al. 1977). He developed a DNA sequencing method known today as Sanger sequencing. The method is based on the separation and detection of labelled dideoxynucleotides (ddNTPs) via capillary electrophoresis, shown in Figure 1.6. As labelled ddNTPs are incorporated into growing ssDNA, synthesis is terminated, creating different lengths of ssDNA. The termination process occurs because of the lack of the OH-group on the ddNTP, which is responsible for crosslinking dNTPs during synthesis. When the ssDNA fragments travel through the capillary, a light emitted by one of the four fluorochromes on terminal ddNTP is captured. Since ssDNA fragments are sorted by size, the light signal can be translated into DNA sequences (Shendure and Ji 2008).
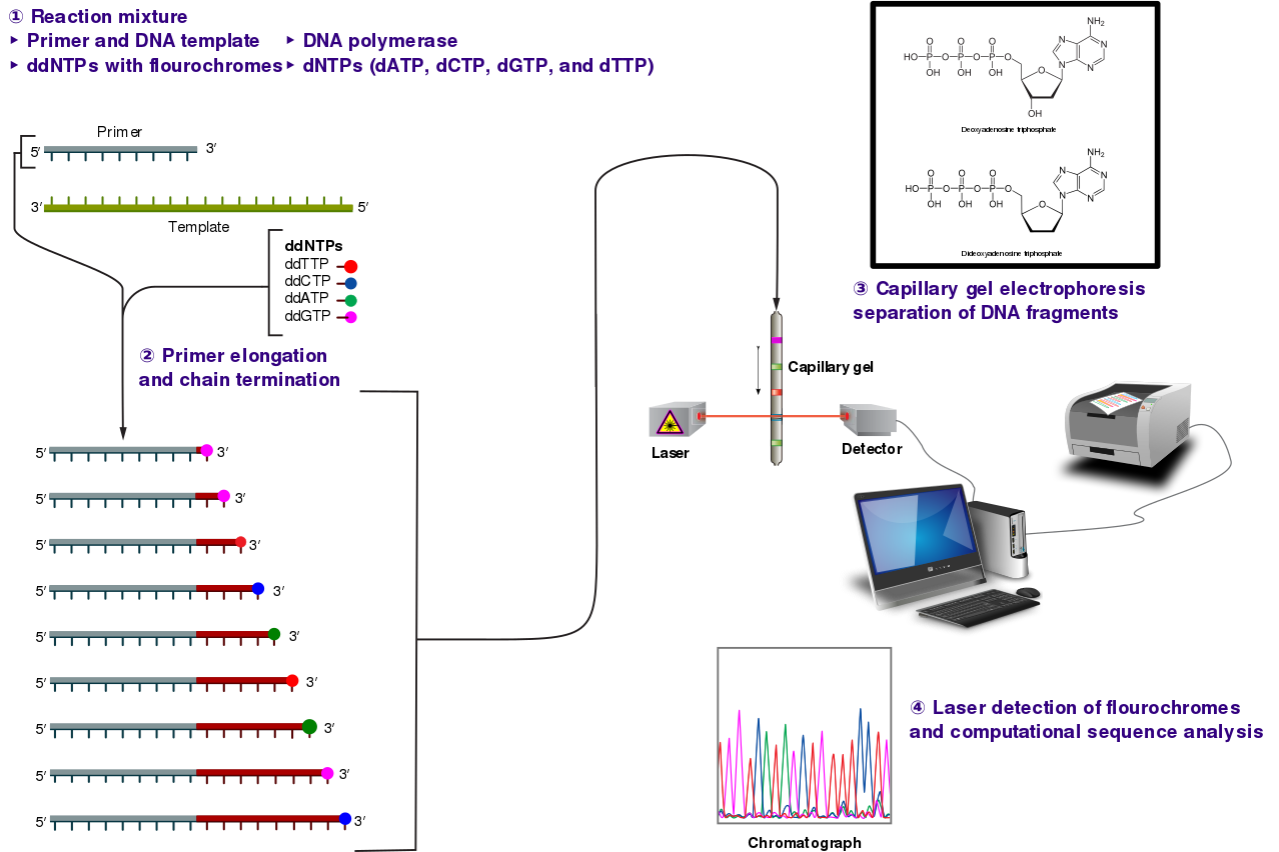
Figure 1.6. Sanger-sequencing method. **(1)** A primer is annealed to a sequence, **(2)** Reagents are added to the primer and template labelled with fluorophores. During primer elongation, the random insertion of a ddNTP instead of a dNTP terminates synthesis of the chain, because DNA polymerase cannot react with the missing hydroxyl. **(3)** The products are then separated on a single lane capillary gel, where the resulting bands are read by an imaging system. **(4)** This produces several hundred thousand nucleotides a day, data which require storage and subsequent computational analysis. The figure is taken from Wikipedia (2012).

1.7.2 High throughput sequencing (HTS)

Nowadays, the sequencing yield and sequence length have changed a lot since Sanger sequencing. Next-generation sequencing technology has the capability to produce large amounts of sequences in a relatively short time compared to the traditional Sanger method(Goodwin, McPherson et al. 2016). There are several NGS platforms available today such as those from Illumina and Nanopore Technologies. The last few years have seen a drastic reduction in the cost of high-throughput sequencing, also known as next-generation sequencing. It is now possible to sequence a complete human genome within days for <$1000(Nanopore technologies). This represents a massive advancement, considering that the first human genome sequence was completed just over a decade ago after 13 years of work at a total cost of -$3 billion(National Human Genome Research Institute, 2018).

NGS techniques and bioinformatics programs have made it fast, easy and affordable to sequence and assemble complete mitochondrial genomes from almost any eukaryotic species for which total DNA has been isolated. In many cases, there is no need to purify mitochondria before sequencing. A single run of whole genomic DNA on an NGS platform, such as Illumina's Hiseq 2000 sequencing system or Nanopore, typically yields enough mtDNA-derived reads to assemble the complete mitochondrial genome (Saho R 2013).

Illumina

Illumina is one of the most popular 2nd generation technologies, because of its low cost and high yield. The basis of Illumina is the reversible-termination sequencing by synthesis with fluorescently labelled nucleotides. To put it briefly, DNA fragments are attached and distributed in a flow cell, where the sequencing reaction occurs by adding a labelled nucleotide. When the labelled nucleotide is incorporated and its fluorescent molecule is excited by a laser, the signal is registered by the machine. Afterwards, the fluorescent molecule is removed and the next nucleotide can be incorporated. DNA fragments can be sequenced from one or both sides giving single end or paired-end sequencing with a maximum read length of 300 base pairs per read (Escobar-Zepeda, Vera-Ponce de Leon et al. 2015).

Minion

While Illumina is called a second-generation sequencing platform, a newer sequencing platform called "the third generation", is capable of producing significantly larger read lengths. This method reduced the amplification bias and also the short read length problem. The time and cost reduction offered by this method is also a big advantage. However, the error rate is higher compared to Illumina (Escobar-Zepeda, Vera-Ponce de Leon et al. 2015).

Oxford Nanopore Technologies are at the forefront of genomics. The MinION from Oxford Nanopore is a pocket-sized, portable sequencing device. It is the only portable real-time device for DNA and RNA sequencing. Through the generation of long reads, the MinION can deliver high-quality whole genome sequencing with the capability to span repetitive regions and structural advantages over short-read sequencing technology like Sanger sequencing. (Oxford Nanopore Technologies).

Figure 1.7: From the left: MinION portable device(Lu, Giordano et al. 2016), and the Illumina HiSeq 2000 system to the right(Illumina 2010).

A number of kits and protocols are available for the Nanopore sequencing, which allows optimised whole genome analysis for a range of sample types and DNA input amounts. In this thesis, I will be using the rapid barcoding sequencing protocol.

Comparing Illumina and Nanopore

One of the differences between the two NGS methods is their read lengths. While Illumina produces read lengths up to 300 bp, the Nanopore technology can produce much longer read lengths. One big advantage when using Illumina is that it has more depth and smaller error rate (5%), while Nanopore has around 15-20% error rate. Illumina is also more expensive to buy and could be less practical for smaller research laboratories. Nanopore, on the other hand, has better portability. It's small enough to be handheld, but requires a laptop to operate. The goal of Nanopore technology is to make the technology sufficiently rapid, simple and inexpensive to permit the sequencing of individual genomes for clinical diagnosis. It is also important to consider that data obtained from second or third generation sequencing technologies have certain computational requirements for their analysis. The bigger the dataset generated, the higher the computational resources and more complex bioinformatics analysis are necessary (Escobar-Zepeda, Vera-Ponce de Leon et al. 2015).

1.8 Aim of this study

Until now, only fragments of mitochondrial and nuclear gene sequences have been available for PRM, and little gene information from these parasites have been available in public databases. Recently the whole mitochondrial genome of *Dermanyssus gallinae* was published (Burgess, Bartley et al. 2018), but the mtDNA is still largely unexplored. In this project, we wanted to do something about this.

In addition sequencing by Oxford Nanopore technology, we will make use of Illumina HiSeq2000-data from a previous project (Øines, unpublished). We will use these sets of PRM sequences and attempt to assemble and annotate the mitochondrial genome from poultry red mite. To first understand the currently used methodology for PRM identification using genetic tools, the approach by Øines and Brännström (2011) was carried out on the material to gain experience in DNA-preparation, standard sequence methodology and analysis. This includes DNA extraction, PCR and sequence alignment using Vector NTI (Thermo Fisher Scientific, Waltham US).

I will be using Nanopore sequencing to obtain new mtDNA sequences from poultry red mite, from the two most common mtDNA haplotypes in Norway (B9 & A16; (Oines and Brannstrom 2011)). DNA from the various haplotypes will be prepared, and we will try to amplify these through different approaches such as a modification of the mtDNA *Toxoplasma* fish method, or by kits like GenomiPhi before doing Nanopore sequencing in order to obtain sufficient starting material. To conclude, I will attempt to obtain a whole or partial draft of the mt genome from *Dermanyssus gallinae,* using MinION long-read *de novo* sequencing and supplemented by previously prepared HiSeq2000 data.

2.0 Materials

All laboratory experiments were performed at the Norwegian Veterinary Institute (NVI) in Oslo. Individual mites were collected from the surveillance programme of PRM by Animalia (kindly provided by Dr Magne Hansen) from different farms around Norway. A schematic representation of the workflow of the thesis is shown in Figure 3.1.

2.1 Kits

Table 2.1: List of commercial kits used for this thesis

Name	Area of use	Manufacturer
DNeasy Blood & Tissue	DNA extraction	QIAGEN
Qubit^{TM} dsDNA HS Assay Kit	DNA quantification	Thermo Fisher Scientific
Ready-To-Go^{TM} GenomiPhi^{TM} HY DNA Amplification Kit	DNA amplification	GE Healthcare
NucleoSpin Gel and PCR Clean-up	PCR clean-up	Macherey-Nagel
Rapid Barcoding Sequencing Kit(SQK-RBK004 & SQK-LSK108)	MinION library preparation	Oxford Nanopore Technologies
Flow Cell Wash Kit	Wash and reuse MinION Flow cell	Oxford Nanopore Technologies

2.2 Chemicals

Table 2.2: List of commercial reagents and chemicals used for this thesis.

Name	Manufacturer
Ethanol absolute	VWR
GelRed	Invitrogen
Proteinase K	QIAGEN
Nuclease-free water	Integrated DNA Technologies
GeneRuler 1 kb DNA Ladder	Fermentas
Loading dye 6x	Thermo Fisher Scientific
Agarose	VWR
TBE – Buffer x 10, pH 8,3	Made in-house
Taq DNA Polymerase	Thermo Fisher Scientific
dNTPs	QIAGEN
Primers(FCOIDG)	IDT
MgCl$_2$	Thermo Fisher Scientific
Taq-buffer	Thermo Fisher Scientific

2.3 Technical equipment

Table 2.3: List of commercial equipment used for this thesis.

Equipment	Model	Manufacturer
Centrifuge	5415 D	Eppendorf
Micro Centrifuge	Mini Star	VWR
Thermomixer comfort	5355	Eppendorf
Electrophoresis unit	Sub-Cell GT	Bio-Rad
Sterile single-use stainless Surgical blades		PARAGON
Gel imaging	ChemiDoc^{TM} XRS+	Bio-Rad
Pipette tips	ART10, 100, 200, 1000	Thermo Fisher Scientific
Pipettes	Various models	Thermo Fisher Scientific
Spectrophotometer	NanoDrop 2000	Thermo Fisher Scientific
Vortex	MS 3 Basic	IKA
Microwave	Kenwood	
Qubit	Qubit 4 Fluorometer	Thermo Fisher Scientific
Volt machine	PowerPac^{TM} Basic	Bio-Rad
PCR-strips	PCR-02-FCP-C	Axygen
PCR machine	T100^{TM} Thermal Cycler	Bio-Rad
PCR tubes	1,5 mL microtubes	Eppendorf
Mini PCR Plate Spinner	MPS 1000	Labnet
Nanopore sequence device	MinION	Oxford Nanopore Technologies
Rotator	PTR-30	Grant Instruments
Folded paper filters		Whatman

3.0 Methods
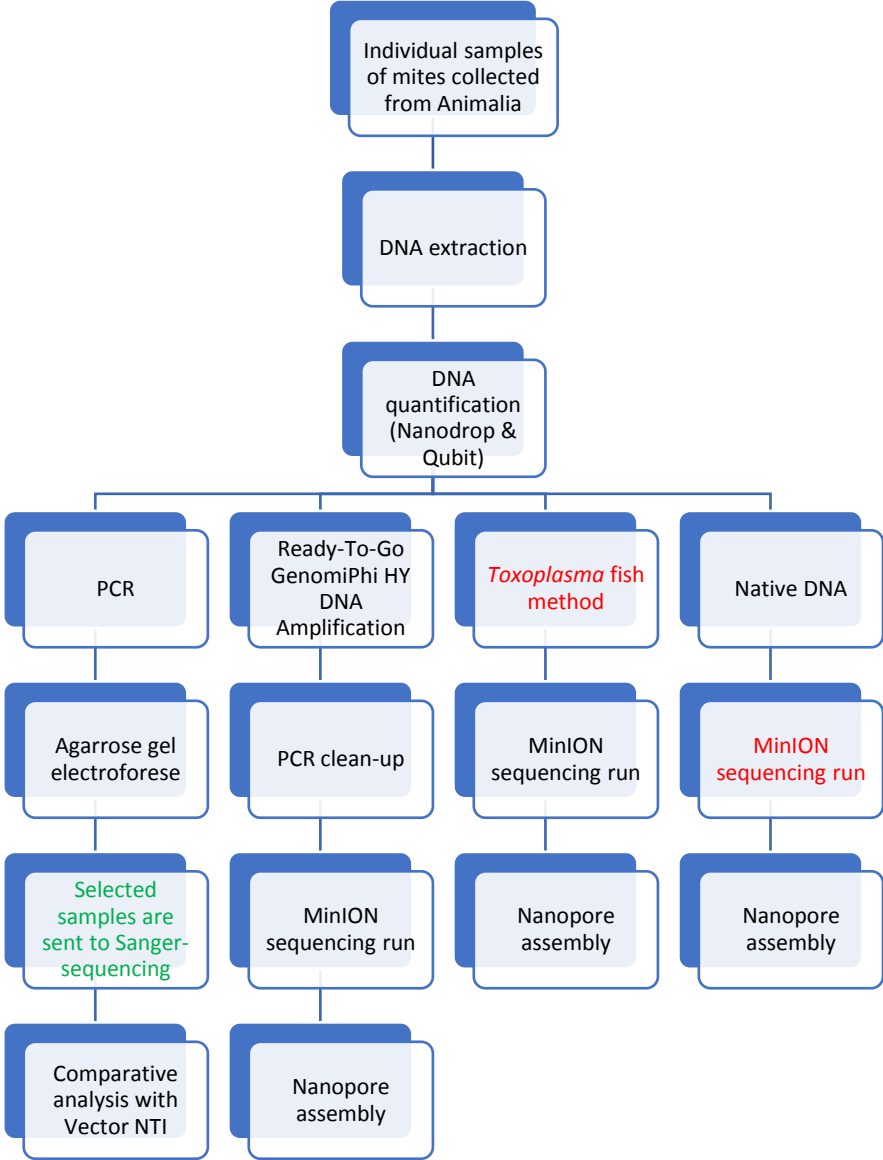
Figure 3.1: A flowchart representation of the workflow of tasks in this thesis. Green colour nodes indicate work done by other people from the VI Institute. Red colour indicates methods performed in this study, but failed to produce results.

3.1 Preparation of mites

Individual mites were placed on a soaked filter paper with water and then crushed with a pipette tip. The mite smear on the paper was cut out and transferred to a tube and marked before doing the DNA extraction. This approach is followed after Øines and Brännström protocol for mite preparation (Oines and Brannstrom 2011).

3.2 DNA extraction using QIAamp DNA Mini Kit and QIAcube instrument

DNA extraction was carried out in two ways. In the first method, DNA extraction was performed following the protocol: *Purification of total DNA from animal tissues* (Spin-Column Protocol). The second method for DNA extraction was using QIAcube, a robotic workstation for automated DNA extraction using QIAGEN spin-columns.

Whole mites are resistant to chemicals, so initial mechanical crushing of the mite was needed to make mite cells available for the digest. In both methods, the chemical lysis of the mechanically crushed material was performed using lysis buffer ATL and Proteinase K. The lysis buffer lyses the cells due to high salt concentration, while Proteinase K digests proteins in the samples, such as nucleases which destroys DNA (Thermofisher Scientific). A detailed description of this protocol is presented in Figure 3.2. While the DNA extraction methods follow almost the same steps, the QIAcube is a faster method when extracting several samples.

Figure 3.2: DNA extraction protocol using QIAamp spin-column.

3.3 Quantification – Qubit and Nanodrop

To measure the purity of dsDNA in a sample, a spectrophotometer can be used. In this thesis, Nanodrop 2000(Thermo Fisher Scientific Inc.) was used. This instrument measures the absorbance of all molecules in the sample at the given wavelength
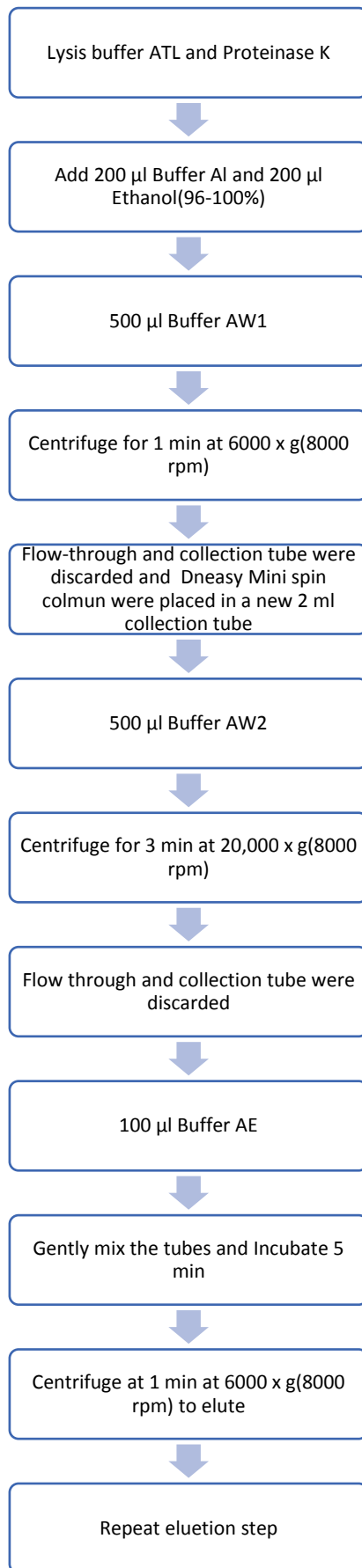
When measuring the DNA concentration, the Invitrogen Qubit 4 Fluorometer (Thermo Fisher Scientific) was used. This fluorometer quantifies DNA, RNA and proteins using very sensitive and accurate fluorescence-based Qubit ™ quantity assays (Invitrogen by Life Technologies). Genomic DNA was measured using the Invitrogen Quant-iT dsDNA High-Sensitivity kit (Invitrogen Inc.) on the Qubit fluorometer according to manufacturer's recommendations.

3.4 mtDNA fish method

This is an adaptation of a DNA concentration method used by Opsteegh (Opsteegh, Langelaar et al. 2010). This approach has also been successfully used at the NVI laboratories for the selected isolation of parasite DNA from fox faces(Oines, Isaksson et al. 2014), and parasite DNA in large meat or tissue material (Opsteegh 2010; *Toxoplasma* detection in sheep abortion samples; Øines, unpublished). This method targeted mtDNA by magnetic capture in samples using Biotin labelled DNA probes interacting with paramagnetic beads covered with streptavidin, allowing these two proteins to bind and by magnetic separation allowing mtDNA from the parasite to be captured. We adopted this approach, and by using probes that were designed towards two PRM mtDNA genes we hoped this method would allow the ratio of mtDNA vs nuclear DNA in the solution to increase.

DNA isolated from QIAcube or using the manual protocol with QIAamp columns were used as starting material. PRM DNA is specifically fished out from the solution by biotin-labelled probes that hybridize specifically with mtDNA from PRM, by raising the temperature and adding the biotin-tagged probes. The mtDNA from PRM is then removed from the solution by adding paramagnetic beads covered with streptavidin. The nucleotide fish probe with the Streptavidin-Biotin complex is finally released from the DNA after a short high-temperature step, and the plan was to use this as a template in Nanopore sequencing.
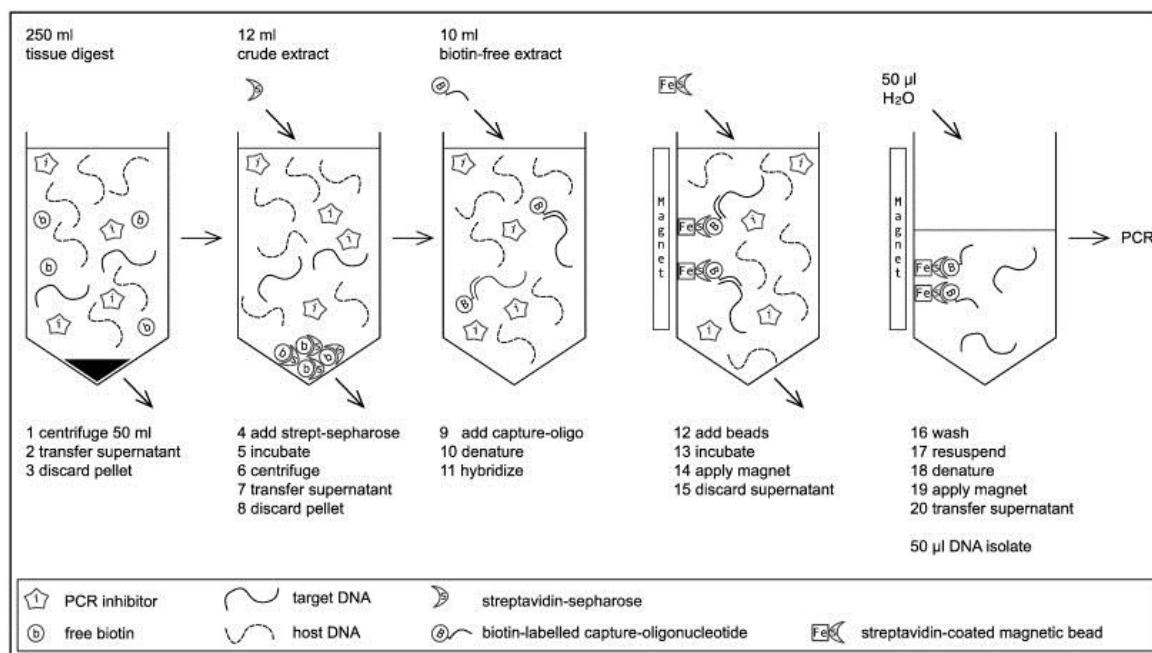
Figure 3.3: Schematic representation of *Toxoplasma* method. The Figure is taken from (Opsteegh, Langelaar et al. 2010).

The two probes with biotin were designed after the CO1 and 16S rRNA gene respectively on the mitochondria. The probes were designed by Øines and were named PRMfishCO1 and PRMfish16S. The protocol for this method has previously not been tested out for PRM, hence the protocol was based on the EM-nok method (Oines, Isaksson et al. 2014). To test if the method was successful, we performed conventional PCR (CO1) on the prepared samples.

3.5 GenomiPhi HY kit

We wanted to amplify the DNA from the different haplotypes before doing Nanopore sequencing, due to requirements of high-quality DNA in the library preparation steps. The GenomiPhi DNA Amplification Kit by GE Healthcare is a whole genome amplification kit that allow users to perform unlimited DNA tests from small or limited samples (Ge Healthcare). The already extracted DNA samples were used for this rapid whole genome amplification method. The first step in this protocol was to denature the DNA by heating in denaturation buffer and then cool. 2.5 µl DNA is mixed with 22.5 µl Water and 25 µl denaturation buffer before it was heated. The 50 µl cooled denaturized DNA template from previous step was added to each cake, and sealed with domed caps.

This was added to the freeze-dried cake which contains DNA polymerase, random hexamers, nucleotides, salts and buffer. The isothermal amplification then proceeded at 30 °C for 4 hours. After amplification, the samples were heated to 65 °C for 10 min, which was required to inactivate the exonuclease activity of the DNA polymerase which may otherwise degrade the amplification products. A control DNA (lambda) was also used and mixed with 50 µl PCR-grade water.

The GenomiPhi kit utilizes bacteriophage Phi29 DNA polymerase to exponentially amplify single- or double-stranded linear DNA templates via a strand displacement reaction, hence no thermal cycling required (GE healthcare, 2007).

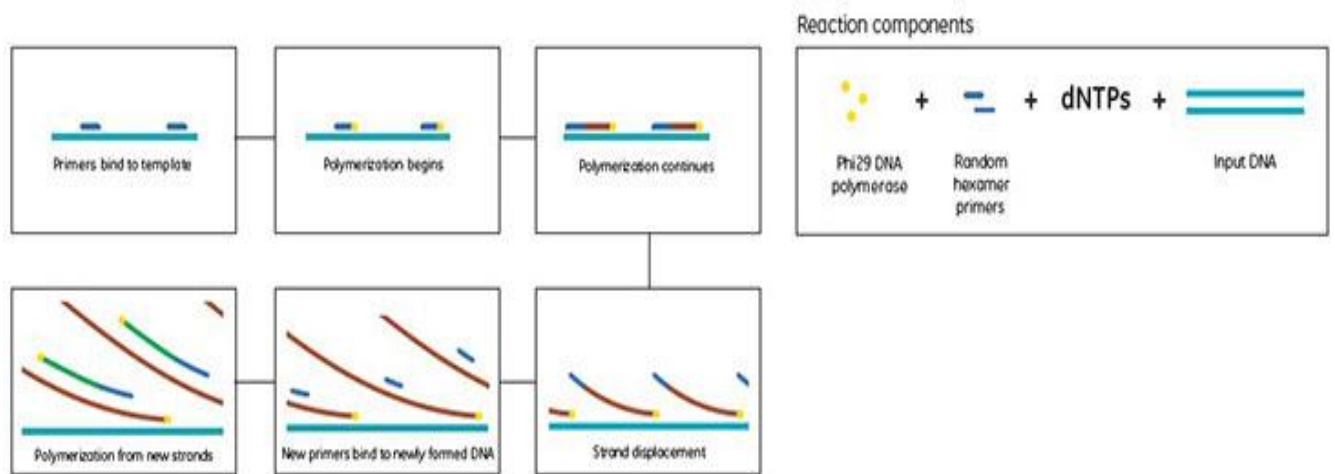**Figure 3.4:** Overview of the Ready-To-Go GenomiPhi HY DNA Amplification Kit procedure (Figure from Sigma-Aldrich)

3.6 PCR clean-up

After trying to run a conventional CO1 PCR as a control on the GenomiPhi products, the results showed weak bands. This may due to primers from the GenomiPhi kit interfere with the PCR-primers. Therefore, a purification of the GenomiPhi primers was necessary to get better results.
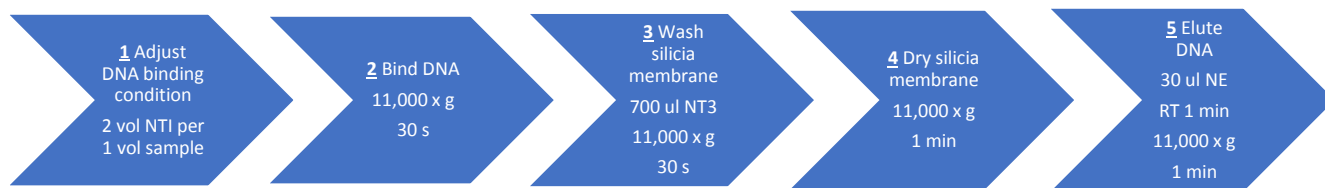
Figure 3.5: PCR clean-up flowchart protocol.

The samples were first mixed with the binding buffer NTI (Macherel Nagel). In the presence of chaotropic salt, the DNA is bound to the silica membrane of a NucleoSpin Gel and PCR Clean-up Column. Contaminations (such as nucleotides, primers, enzymes, mineral oil, dyes) are removed by simple washing steps with ethanol containing Wash Buffer NT3. And finally, the last step, the purified DNA is eluted under low salt conditions Elution Buffer NE (5 mM Tris/HCl, pH 8.5).

Since there was a small sample of GenomiPhi products, the volume was adjusted of the reaction mixture to 100 µl with PCR-grade water. 1 volume of the sample was mixed with 2 volumes of Buffer NTI (mix 100 µl PCR reaction and 200 µl Buffer NTI). A NucleoSpin Gel and PCR Clean-up Column were placed into a collection tube (2 mL). 700 µl DNA sample was loaded and centrifuged. During the washing step, 700 µL buffer NT3 were added to the Column and centrifuged according to the protocol. This step was repeated to minimize chaotropic salt carry-over. To remove the Buffer NT3 completely, it was centrifuged for 1 min. In the last elution step, the column was placed into a new 1,5 mL Eppendorf tube. 30 µL Buffer NE was added and incubated at room temperature for 1 min and centrifuged.

3.7 PCR

The purpose of Polymerase Chain Reaction (PCR) is to dramatically increase the number of copies of known fragment of DNA. PCR reaction consists of three steps repeated for many cycles. The first step is called denaturation. In this step, the double strand DNA melts open to single-stranded DNA. During step two called annealing, primers anneal to the template

DNA. The last step, elongation, the DNA polymerase attaches and copies the DNA template. All this process is carried out on an automated PCR cycle machine.

Extracted DNA was used as a template for the PCR reaction, which was carried out in 40 μL mixture containing PCR buffer, dNTP mix, each of reverse and forward primer (listed in Table 6), Taq DNA polymerase, nuclease-free water, MgCl2, and 3 μl of extracted DNA. Table 3.1 shows the PCR cycling conditions and the amount of volume of each reagent necessary for running a PCR. Initial denaturing at 95 C was performed for 2 min, followed by 37 cycles of denaturing at 95°C for 30 s, annealing at 46°C for 30 s, and elongation at 72°C for 1 min. Final elongation was performed at 72°C for 2 min. Verification of the PCR reaction can be done by analysing the PCR product on an agarose gel.

Table 3.1. Thermal cycler conditions for PCR reaction.

Hold for	2 minutes at 95°C
37 cycles	30 seconds at 95°C 30 seconds at 46°C 1 minute at 72°C
Hold for	2 minutes at 72°C
Infinity	8°C

Table 3.2: Primers used for PCR reaction.

Primer name	Sequence 5' – 3'
FCO1DG (forward)	CAT TAA TAT TAA CTG CAC CTG AGA TG
RCO1DG (reverse)	CCC GTG GAG TGT TGA AAT TCA TGA

3.8 Agarose gel electrophoresis

Agarose gel electrophoresis was used to visualize the results of PCR reactions. After running PCR, the PCR products of both QIAamp method and GenomiPhi were examined by electrophoresis on a 1% agarose gel. 8-9 µl of GelRed which binds dsDNA was added to the liquid agarose gel mixture (100 mL TBE). A gel comb was used to make wells within the gel. The gel was allowed to set for around 25-30 minutes. Following this, 10 µl of PCR product was mixed with 2 µl 6X gel loading dye, and directly applied to each well in the gel. The gel was run for 1 hour at 100 Volt using PowerPac™ Basic, Bio-Rad. After electrophoresis, the agarose gel was placed on the Molecular Imaging System (Bio-Rad Laboratories) and DNA bands were visualized. The PCR fragments' size was determined by comparing with a commercial size marker (GeneRuler 1 kb DNA Ladder).

3.9 Computational analysis – Sanger sequencing

After Sanger sequencing of partial CO1 gene, sequences were manually edited using the program Contig Express implemented in the Vector NTI software (Invitrogen. Inc.). Background noise and unresolved areas at the ends of the sequences were removed. All the sequences from the sequence reaction were then assembled to create a consensus sequence for each sample. Some fragments were incomplete and could not overlap, which meant no contigs were made. All multiple sequence alignments were made using the program AlignX in the Vector NTI software package.

3.10 Nanopore MinION sequencing

Prior to the sequencing run with the MinION, a quality control (QC) of the flow cell being utilized is required to reveal any faulty manufacturing or low quality of the flow cell in terms of the number of active nanopores. For the sequencing run, MinION software MinKNOW v. 1.4.2 was used.

The Nanopore sequencing run was performed using the Rapid Sequencing Kit SQK-RAD004 following the Rapid Barcoding Sequencing (SQK-RBK04) protocol for 1D MinION sequencing.
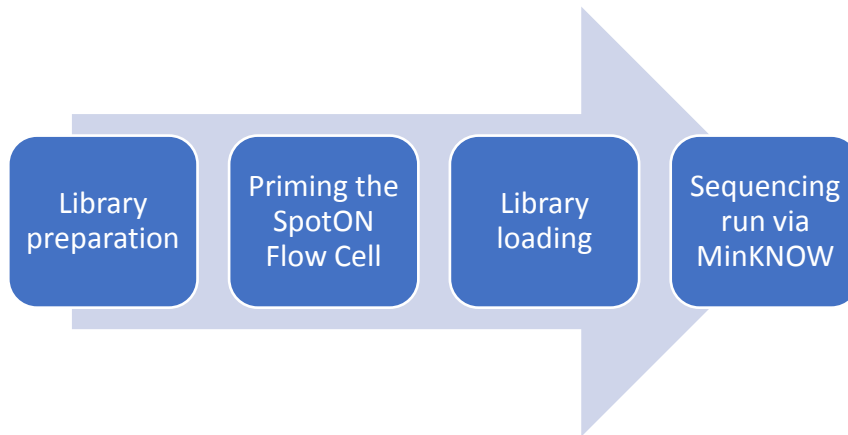
Figure 3.6: Flowchart of Nanopore MinION sequencing protocol.

3.10.1 Library preparation

Template DNA was prepared by mixing 200 ng genomic (Library A) or amplified DNA from GenomiPhi (Library B) with nuclease-free water in a 1,5 ml Eppendorf tube resulting in a total volume of 7,5 ml. The content was mixed thoroughly by inversion and spun down in a microfuge. A fragmentation mix was made consisting of the template DNA and FRM with volumes being 7,5 µl and 2,5 µl respectively, resulting in a total volume of 10 µl. The mix was incubated in a thermal cycler at 30 °C for 1 min and then at 80 °C for 1 min. An adapter ligation was performed by adding 1 µL RAD to fragmentation mix, mixed by inversion and spun down. 0,2 µl of TA Ligase Master Mix was then added to the mix, mixed by inversion, spun down and incubated for 5 min at room temperature.

3.10.2 Priming the SpotON Flow Cell

The flow cell was inspected and made sure no air bubbles were present, and that the buffer was continuous throughout the flow cell. A priming mix was made by mixing 480 µl RBF and 520 µl nuclease-free water. The flow cell priming port was then loaded with 800 µl priming mix. After 5 min, the remaining primer mix was then loaded.

3.10.3 Library loading

This step was completed during the 5 min wait time in the above step and the resulting library loading mix was loaded immediately after the final primer mix was loaded. A library loading mix was made consisting of RBF (25,5 µl), DNA library (11,0 µl), LLB(26,6 µl) and nuclease-free water(12 µl) resulting in a total volume of 75 µl. The mix was then mixed by inversion and spun down before all of it being loaded dropwise into the SpotON sample port. The sequencing run was then initiated through the MinKNOW software and ran uninterrupted for 48 hours.

3.11 Washing MinION & re-application of the new run

After the initial failed Minion run of the library from native DNA (Library A; BCI/BC2), where no data was obtained, the MinION was quickly washed in order to hopefully save some active pores for a subsequent run. The MinION was successfully reused by performing the wash protocol quickly and re-applying the second library preparation which was prepared from the GenomiPhi material (Library B; BC3/BC4).

The first thing which was performed was to stop the previous run and unplug the MinION from the computer. The figure under shows the protocol on how to wash a MinION flow cell and then add a new DNA library immediately after. The first step was to remove all the waist buffer, opened the priming port and made sure there were no bubbles. Next, 150 µl of solution A was added into the priming port. Waited 10 minutes. After the 10 minutes, 150 µl of solution B was added into the same hole in priming port. Now the priming step can be repeated. Priming buffer was added and a new DNA library from GenomiPhi (BC3/4) was

added to the sample port, and sequence restarted (washing flow cells – Nanopore protocol). This time the library preparation followed the protocol "1D PCR barcoding (96) genomic DNA (SQK-LSK108), as the reagents of the previous sequencing protocol were finished.

Add the next library
immediately

⬇

Solution A

150 µl

⬇ 10 min waiting

Solution B

150 µl

⬇

Ready for the next library
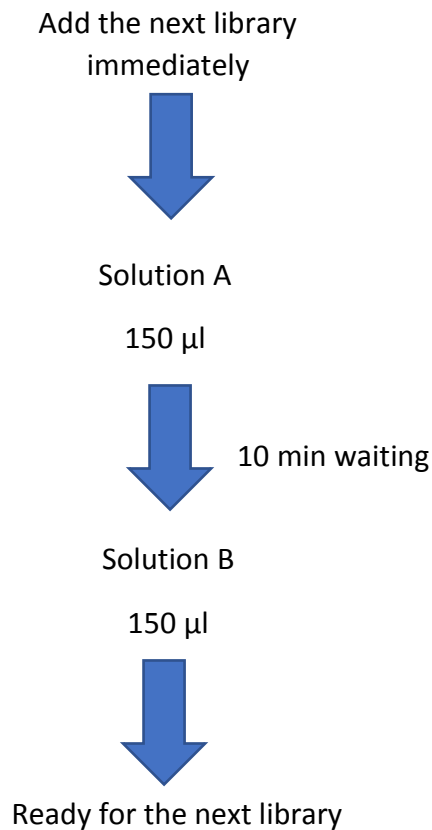
Figure 3.7: Brief procedure over washing of a flow cell.

3.12 Computational analysis – NGS

After the Nanopore MinION sequencing run, the raw data files from the sequencer was basecalled using a basecalling software (Albacore) and .fast5 files were converted to .fastq files and joined through the software package. The CLC Genomics Workbench assembly tool was used for assembly and/or mapping to reference of reads from HiSeq2000 and Nanopore MinION. The HiSeq data was obtained from a previous PRM project (Øines, unpublished), and not generated from the mites I worked with in this project. A report of the HiSeq-data can be viewed in Appendix 3.

4.0 Results

All samples from Haplotype A16 using QIAamp Spin-column were also measured using Nanodrop. However, the results of Nanodrop showed unreliable values, hence not included. As a result of this, B9 Haplotype samples were not measured with Nanodrop. Nanodrop results of A16 samples can be viewed in Appendix 1.

4.1 DNA concentration measurement using Qubit

DNA concentration from both the native DNA (QIAamp) and GenomiPhi was measured using a Qubit fluorometer. Whole genome amplification through GenomiPhi gave higher DNA concentrations. This is shown in the Figures 4.1-4.4 below.
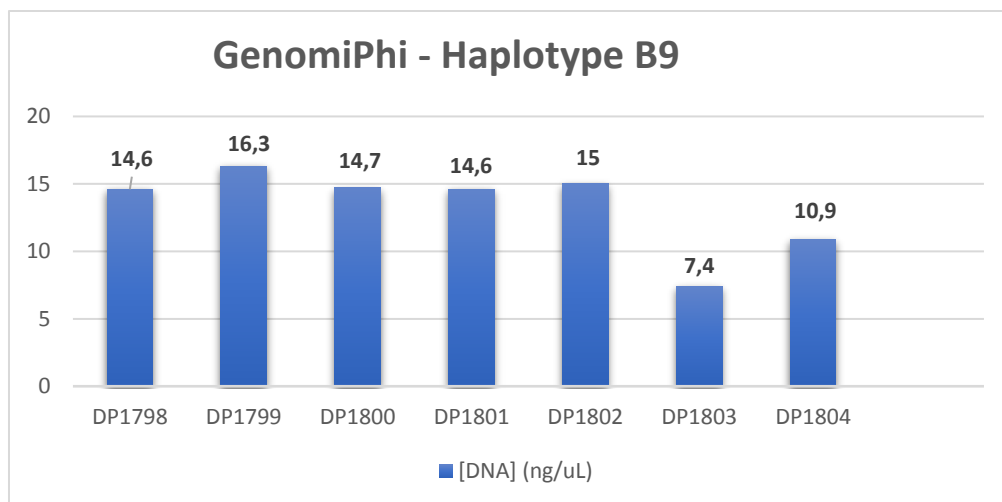
Figure 4.1: Qubit results from GenomiPhi samples of haplotype B9. The values above each peak represent the concentration of DNA (ng/µl).
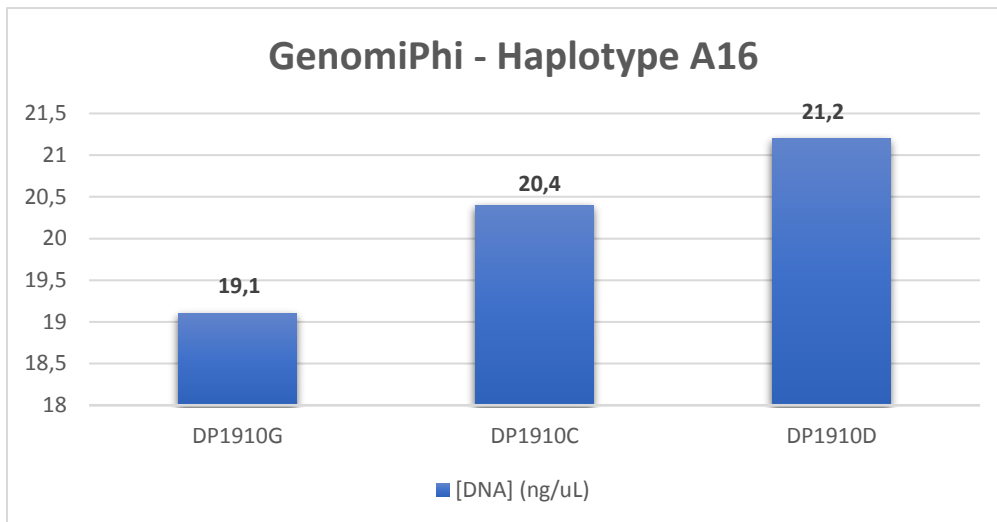
Figure 4.2: Qubit results from GenomiPhi samples of Haplotype A16. The values above each peak represent the concentration of DNA (ng/µl).
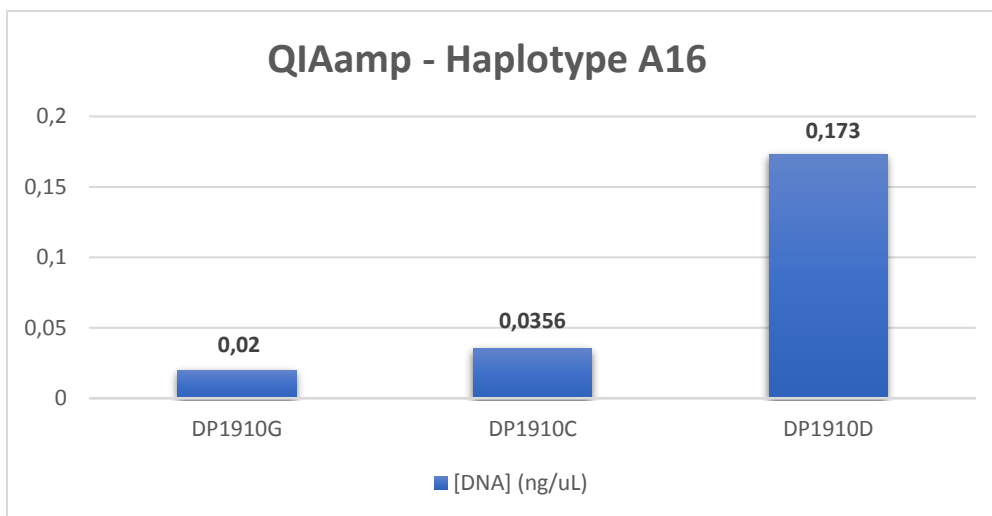
Figure 4.3: Qubit results from extracted DNA samples using the QIAamp Spin-column protocol. The values above each peak represent the concentration of DNA (ng/µl).
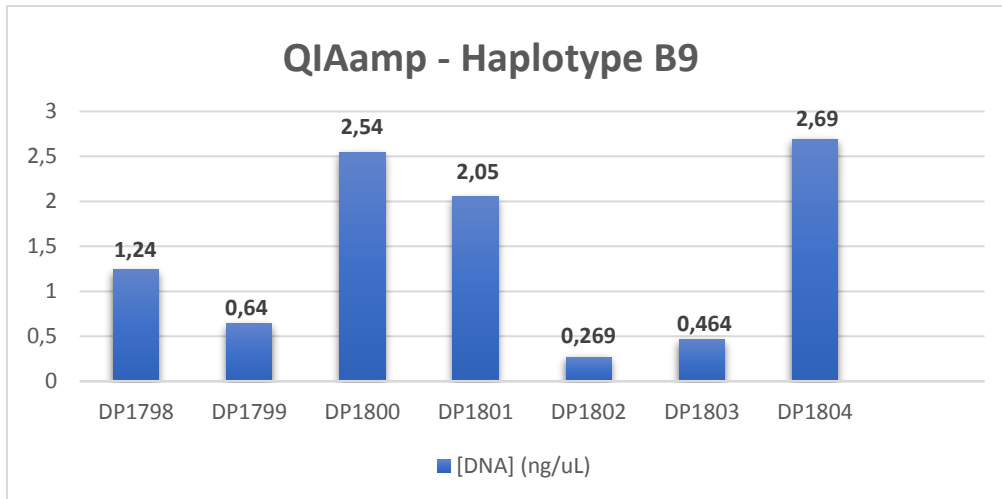
Figure 4.4: Qubit results from previous extracted DNA samples using QIAamp Spin-column protocol. The values above each peak represent the concentration of DNA (ng/μl).

4.2 Gel image of PCR products

The gel electrophoresis image of PCR products of GenomiPhi and QIAamp Spin-column haplotype A16 can be viewed in Figure 4.5 and 4.6. After performing conventional PCR (CO1) on the *Toxoplasma* fish method, results on the gel image showed no bands (figure not shown). The method was unsuccessful, hence not included prior to NGS sequencing.
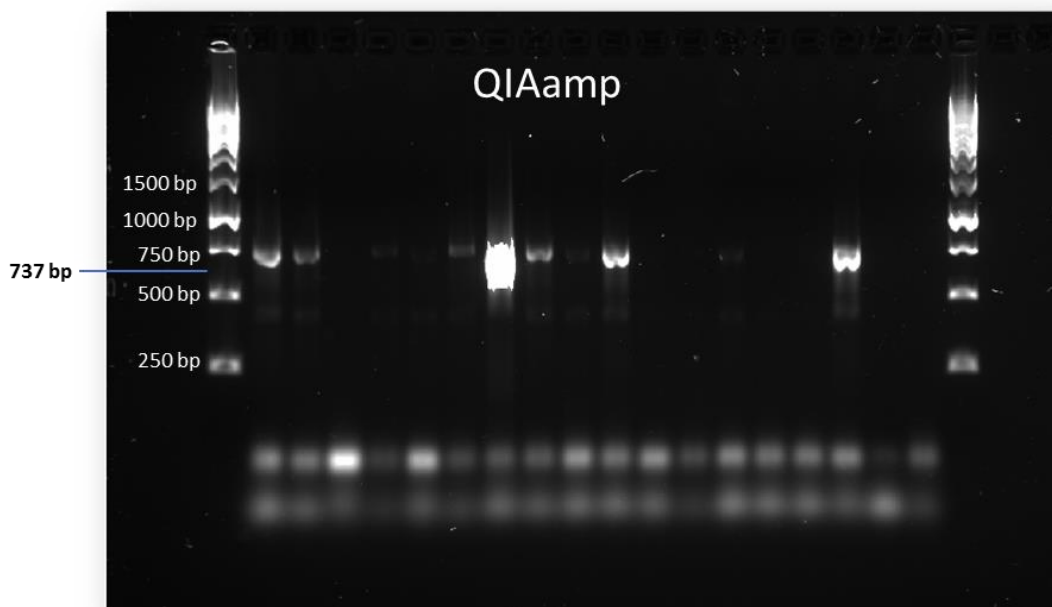
Figure 4.5: QIAamp Spin-column PCR products of haplotype A16. A 1 kb DNA Ladder was used.
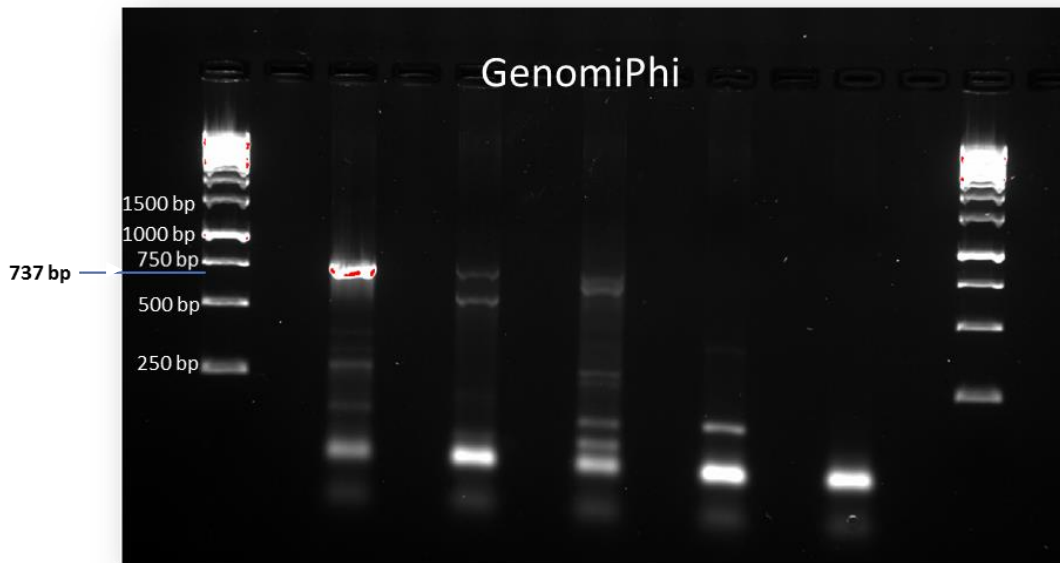
Figure 4.6: GenomiPhi PCR product of haplotype A16. A 1 kb DNA Ladder was used.

4.3 Chromatogram of A16 Haplotype
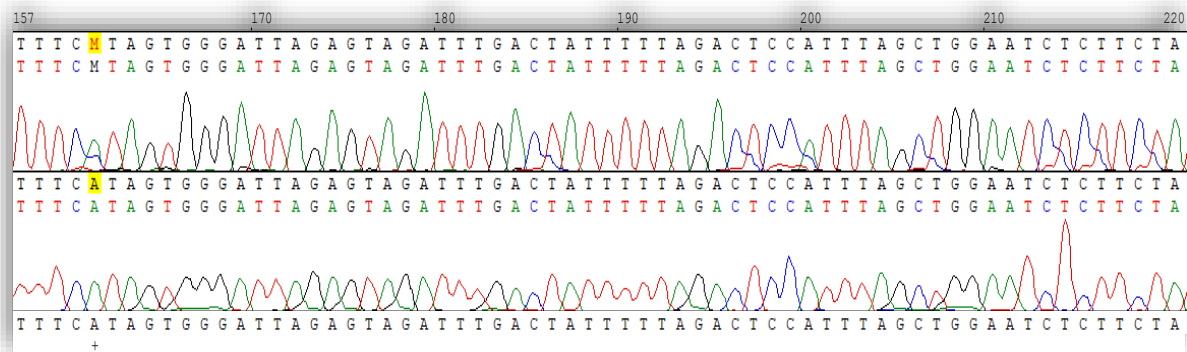
Figure 4.7: Chromatogram of a contig from haplotype A16 with forward and reverse sequence. Yellow marked ambiguous base (M) in the forward sequence indicate some background noise, but from the reverse sequence, the base is clearly an Adenine (A). Different coloured peaks correspond to different bases. The profile of positions of As is represented in green, Ts in red, Gs in black, and Cs in blue. The Figure is taken from ContigExpress in Vector NTI.

4.4 Alignment of A16 and B9 CO1 genes (partial)

```
            1                                                                                                  100
A16   (1)  CAGGTACTGGTTGAACTGTTTATCCTCCCTTATCAAATTTTAATTTTCATAGTGGGATTAGAGTAGATTTGACTATTTTTAGACTCCATTTAGCTGGAAT
B9    (1)  CAGGTACTGGTTGAACTGTTTATCCTCCTTTATCAAATTTTAATTTTCATAGTGGGATTAGGGTAGATTTGACTATTTTTAGACTCCATTTAGCTGGAAT
           101                                                                                                 200
A16 (101)  CTCTTCTATTCTTGGAGCAATTAATTTTATCACTACAATTATTAATATACGACCTAAATCGATATCATTGGAAATAATACCTCTCTTTCCATGATCTATT
B9  (101)  CTCTTCTATTCTTGGGGCGATTAATTTTATCACTACAATTATTAACATACGACCAAAATCGATATCATTGGAAATAATACCCCTCTTTCCATGATCTATT
           201                                                                                                 300
A16 (201)  TTAATTACTGCAATTTTATTACTTTTATCTCTGCCAGTTCTTGCAGGTGCAATTACTATATTATTATCTGATCGTAATTTTAATACAACTTTTTTTGATC
B9  (201)  TTAATTACTGCAATTTTATTACTTTTATCTCTGCCAGTTCTTGCAGGTGCAATTACTATATTATTATTTGATCGTAATTTTAATACAACTTTTTTTGATC
           301                                                                                                 400
A16 (301)  CTAGGGGAGGGGGAGATCCTATTTTATATCAACATTTATTTTGATTTTTTGGACATCCGGAAGTTTATATTTTAATTATTCCAGGATTTGGAATAATTTC
B9  (301)  CTAGGGGAGGGGGAGATCCTATTTTATACCAACATTTATTTTGATTTTTTGGACATCCGGAAGTTTATATTTTAATTATTCCAGGATTCGGAATAATTTC
           401                                                                                                 500
A16 (401)  CCACATTGTTTGTTATCAAACTGGAAAAAAGAAACCTTTTGGAAATATTAGAATAATCTATGCAATATTAACAATTGGTATTCTAGGATTTATTGTTTGA
B9  (401)  CCATATTGTTTGTTATCAAACTGGAAAAAAGAAACCTTTTGGGAATATTAGAATAATCTATGCAATATTAACAATTGGTATTCTAGGATTTATTGTTTGA
           501       514
A16 (501)  GCCCACCATATATT
B9  (501)  GCCCACCATATATT
```

Figure 4.8: Alignment of CO1 sequences from the two haplotypes B9 and A16 (514 bp long) using AlignX (Vector NTI). Yellow background colour marks sequence identity match between the two haplotypes, white background colour indicate differences.

4.5 Illumina HiSeq200 data

Mapping the HiSeq2000 reads using the whole *Varroa destructor* mt genome, or individual mt genes, as reference identified several mitochondrial genes, NAD1, CytB and 16S rRNA. The HiSeq data were obtained from a previous project. The analysis report can be seen in Appendix 3.

Table 4.1: CLC Genomics Workbench output table which shows an assembly summary report of the Illumina data. From the table, we see that the number of reads total are 13480,016.

	Count
References	1
Mapped reads	72
Not mapped reads	13 479 944
Total reads	13 480 016

4.6 Nanopore sequencing run

The initial QC performed on the MinION flow cell revealed the number of active pores to be >1480, which is within the recommended range (minimum 800). Prior to sequencing the second library preparation, the GenomiPhi library, a small bubble occurred on the cell, resulting in additional inactive pores in the flow cell. This was shown on the data graph during the run (not shown). Long reads were up to 17000-20000 kb long, and an overview of the read distribution from the Nanopore MinION sequencing is shown in Figure 4.9. Total reads were 76 580, shown in Table 4.2.

Table 4.2: CLC Genomics Workbench output table which shows an assembly summary report of the Nanopore MinION data.

	Count
Reference	1
Mapped reads	2553
Not mapped reads	74 027
Total reads	76 580

1.2 Distribution of read length
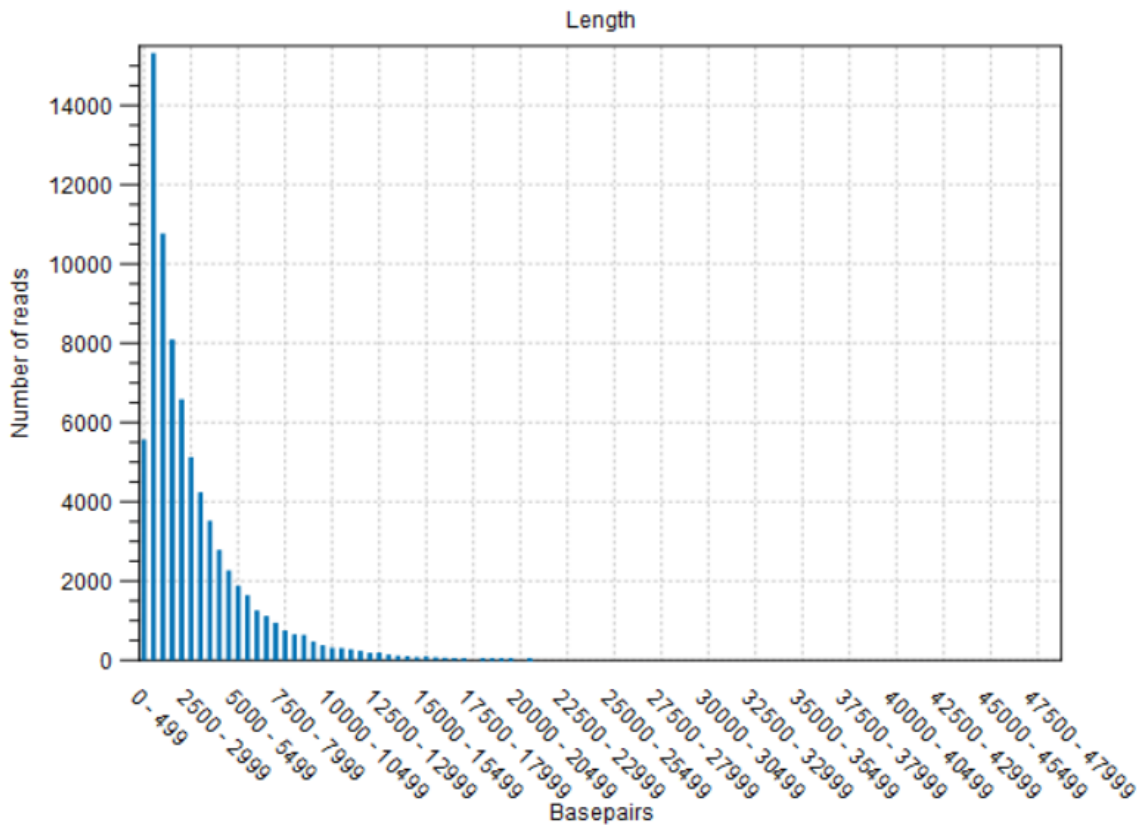
Figure 4.9: CLC Genomics Workbench output graph. Distribution of Nanopore MinION reads. The x-axis shows the number of bases associated with each read and y-axis shows the number of reads associated with length**.**

4.7 Nanopore sequence data analyses

The assembly and annotation of an unknown genome is most accurate when a genome from a closely related species is used as a reference. Following nucleotide BLAST of partial CO1 and 16S rRNA genes, it appears that *V. destructor* is the mtDNA genome (complete) from mite that is closest to *D. gallinae*, followed closely by *S. rarior* and *M. occidentalis* (Table 4.3 & 4.4). Therefore the *V. destructor* was used as a reference when trying to generate a crude assembly of the mtDNA from poultry red mite.

Table 4.3: Nucleotide BLAST of partial CO1. *Varroa destructor* gave the highest score(%) and is marked in bold.

CO1	
Stylochyrus rarior	72,22 %
Varroa destructor	**79,19 %**
Metaseiulus occidentalis	74,39 %

Table 4.4: Nucleotide BLAST of 16s rRNA.

16S rRNA	
Stylochyrus rarior	78,35 %
Varroa destructor	**79,80 %**
Metaseiulus occidentalis	77,45 %

Mapping all of the minion reads against the complete mt reference genome *V. destructor* did not produce any reliable consensus sequence output. Hence, an opposite strategy was implemented in which all protein-encoding gene sequences from the *V. destructor* reference strain (13 in total) were blasted against the entire minion data (local blast). A typical output result showing two genes (12/CYTB and 13/ND1) located within a single MinION read is shown in Figure 4.10. The sequence regions in the contigs mapped by the *V. destructor* genes were verified by a second separate nucleotide blast (NCBI).
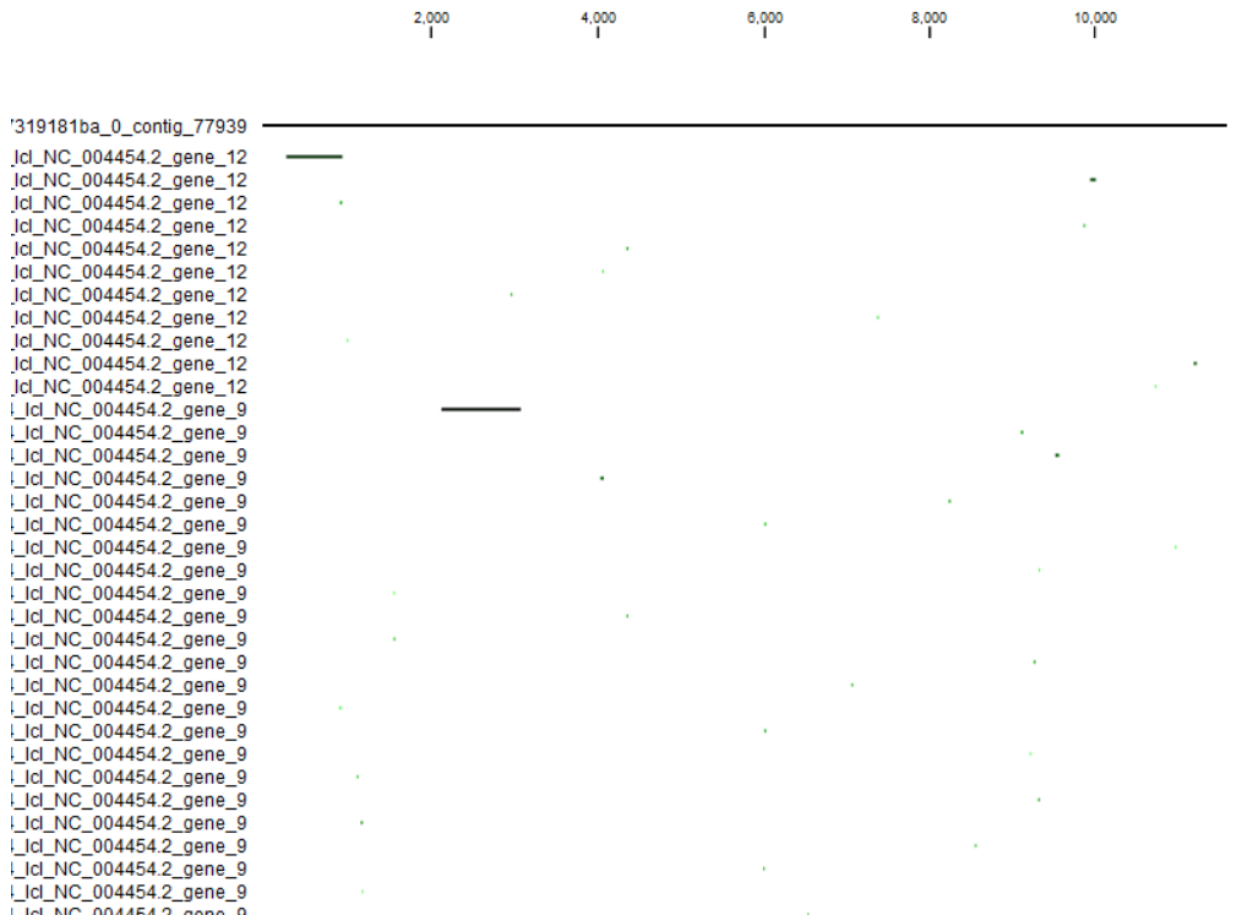
Figure 4.10: A typical local BLAST result retrieved from CLC Genomics Workbench were all *V. destructor* protein-encoding gene sequences were blasted against the MinION data.

4.8 mt genome organization

We also ran assembly using Contig express of selected MinION reads which mapped to the reference genes in order to produce larger contigs with higher sequence accuracy. These contigs were used to elucidate the order of some of the protein-encoding genes. In Figure 4.12, a partial gene arrangement is presented. Only the order of 6 out of the 8 genes identified, could be determined (Table 4.5).
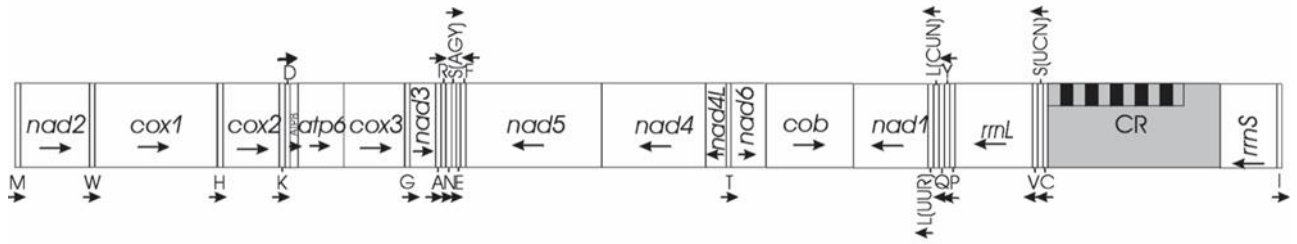
Figure 4.11: Mitochondrial gene arrangement in *Varroa destructor.* Arrows indicate the direction of gene transcription. The Figure is taken from (Navajas, Le Conte et al. 2002)*.*
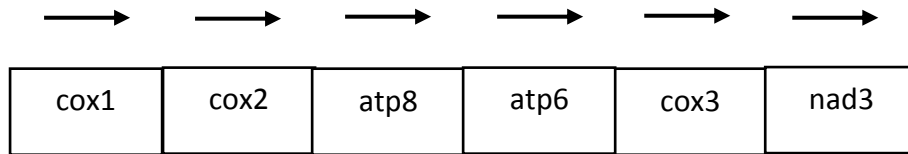
Figure 4.12: Partial gene arrangement of the mt genome from *D. gallinae* generated in this study. The gene arrangement is made out of 6 protein-coding genes.

Table 4.5: MtDNA genes found using CLC Genomics Workbench. *Varroa destructor* used as the reference genome. The whole or partial sequence reads were BLASTed to confirm that they are the correct genes (top hits with low E-values). 8 genes found of total 13 from MinION data, and 3 genes found from Illumina HiSeq-data.

MinION	HiSeq-data	
CytB	CytB	Gene 12
ND1(NADH1)	ND1	Gene 13
ND3(NADH3)		Gene 7
ND4(NADH4)		Gene 9
ND5(NADH5)		Gene 8
COX1		Gene 2
COX2		Gene 3
COX3		Gene 6
	16S rRNA	

5.0 Discussion

The poultry red mite is the most economically important parasite affecting egg production facilities worldwide. In this study, we have sequenced DNA from several mites obtained from different farms around Norway. Furthermore, we have performed deep sequencing and a partial assembly and annotation of the PRM mitochondrial genome.

5.1 Nanodrop and Qubit comparison; what do these measurements tell us?

Due to its small size (1 mm), it is difficult to obtain sufficient amounts of DNA for sequencing. This was clearly noticed when performing the gold standard of PRM extraction. The DNA concentrations were too low for both Haplotype A16 and B6, measured on the Qubit. Since the Qubit cannot say anything about the impurities, and only about concentration, Nanodrop was also tested out on the samples. However, the results of the Nanodrop (Supplementary Table 1) were not measurable, due to the low amount of DNA in the samples. Hence, the purity of the DNA was also difficult to evaluate as measured by the absorbance ratios 260/280 and 230/160. According to our results, we see that there are too many low values, which can be due to contaminants but also due to the low detection of DNA. The 260/280 ratio should measure between 1.8-2.2, and the 230/260 between 2.0-2.2. According to the results from Nanodrop, the values from both 230/260 and 260/280 were not reliable and therefore not a good way of quantifying DNA in our samples.

5.2 Gold standard of PRM identification; Chromatogram of A16 Haplotype & Alignment of A16 & B9 Haplotypes.

After discovering that the low DNA concentrations were not acceptable prior to NGS sequencing, we performed whole genome amplification to amplify DNA to a sufficient amount. The results from GenomiPhi Qubit show a much higher DNA yield. The DNA extracted from both haplotype A16 and B9 were pooled together prior to NGS sequencing. This was important to fulfil the requirement of the amount of genomic DNA prior to MinION sequence run (~400 ng genomic DNA). Even though the haplotypes are different, they are

almost identical. This was shown when performing a partial alignment of the two haplotypes (figure 22). *Toxoplasma* fish method was an alternative method used to get higher DNA yield, but the method failed and was not taken further. PCR gel image showed no bands.

5.3 Evaluation of HiSeq data

HiSeq data revealed few reads that actually mapped to the mt reference genome. It is believed that based on COX1 and 16s rDNA, it is about 80% similarity nucleotide level between *D. gallinae* and *V. destructor.* That is high enough to map with the HiSeq data, but apparently not good enough for MinION. The relatively few HiSeq reads that matched the genome was because the sequencing depth was insufficient. The total number of reads was only 13 million (Table 4.1), rather than billions. With greater depth and the use of GenomiPhi, we could probably have done much better mapping of the HiSeq data that could have complemented the MinION better. On the other hand, the HiSeq data were not planned to use in the thesis, but they were available, and they provided some information about the genes present and which genes that were matched.

5.4 Evaluation of Nanopore MinION sequencing results

Due to the failed MinION sequencing run for library A (QIAamp extraction) we only had one library left for MinION sequence running, which was GenomiPhi. The unsuccessful sequencing run was due to misinterpretation of the lab protocol. The single sequencing run performed generated reads consisting of long stretches of A's and T's, resulting in some low-quality data. More runs could have been performed, but time was limiting factor and also due to the failed first sequencing run.

When assembling or mapping our MinION data, it was not possible to run the mapping against the entire reference mt genome, which would probably be a common approach. For example, a well-annotated human genome against and sequenced genome from another human individual or closely related species. This did not work for our MinION data, because our reference mt genome seems only to be around 80 % identical at the sequence level in addition to the error of sequencing. The organization and order of genes may also be

different from the reference. The long reads can, therefore, overlap in places where there is no correct order of the genes, thus leading to no match. Instead, when we ran gene against gene, where the gene was used as "search sequence" and the contig as "reference sequence", we managed to get some mapping.

5.5 Partial Mitochondrial gene map of PRM

The main finding in this thesis was the order of several of the mitochondrial genes. However, all the 37 genes expected to be present in a standard metazoan mitogenome were not identified in this genome. Only 8 of the protein-coding genes were found (Table 4.5). One of the factors can be high A and T content. High A+T content is also common in mesostigmata species with the proportion ranging from 73 % (*S. rarior*) to 80 % (*V. destructor*) (Lingzi et al. 2019). From our partial gene organization, we can see that the order of genes in the mitochondrion of *D. gallinae* is very similar to *V. destructor*, which suggests a close evolutionary relationship between our parasite and *V. destructor.* The other missing genes were also identified, but we could not yet say the order of them. It is also not uncommon that some genes are missing, for instance, the mt genome of *Metaseiulus occidentalis* where two genes (ND3 and 6) appears to be missing(Jeyaprakash and Hoy 2007). Therefore, we cannot exclude that one or more genes are also missing in the PRM mt genome.

We knew that it was not possible to get all the rRNA, tRNA and non-coding sequences when annotating the mitochondrial genome, due to the uncertainties mentioned above. Many of the protein-coding genes gave good mapping in the CLC program, but when BLASTed the sequences were in many cases not identical enough to hit. Therefore, it was not possible to annotate and identify many of the genes with enough certainty. We managed though to obtain the order of several of the protein-coding genes. This allowed us to make a rough partial sketch of the mt genome, an important goal as well as a good starting point for future sequencing work to obtain the entire PRM mt genome.

6.0 Future work & conclusion

The development of more bioinformatics tools and expertise for metagenomics analysis is necessary. Current sequencing platforms are delivering a massive yield at a very low cost, increasing the amount of information to analyze. Using BLAST against databases such as NCBI retrieve a lot of related sequence hits with annotations that can be used to mine taxonomical information as well. A goal of this thesis was to assemble and provide a rough draft of the PRM mt genome. The results showed that there were a lot of genes found, but difficult to identify all the 37 genes present in the mitochondrion. The results show that there are many challenges remaining. To annotate the full mitochondrial genome, future work should involve Sanger sequencing and/or HiSeq sequencing of the mitochondrial genome. Sanger sequencing is probably more efficient for high accuracy of base calling and order of the genes and rRNAs sequences and regulatory regions in the genome. All in all, we have provided a partial draft genome from the mitochondrial red mite. The organization of the genes identified are very similar to *Varro destructor.*

7.0 References

Burgess, S. T. G., K. Bartley, F. Nunn, H. W. Wright, M. Hughes, M. Gemmell, S. Haldenby, S. Paterson, S. Rombauts, F. M. Tomley, D. P. Blake, J. Pritchard, S. Schicht, C. Strube, O. Oines, T. Van Leeuwen, Y. Van de Peer and A. J. Nisbet (2018). "Draft Genome Assembly of the Poultry Red Mite, Dermanyssus gallinae." Microbiol Resour Announc 7(18).

Chauve, C. (1998). "The poultry red mite Dermanyssus gallinae (De Geer, 1778): current situation and future prospects for control." Vet Parasitol 79(3): 239-245.

Chu, T. T., T. Murano, Y. Uno, T. Usui and T. Yamaguchi (2015). "Molecular epidemiological characterization of poultry red mite, Dermanyssus gallinae, in Japan." J Vet Med Sci 77(11): 1397-1403.

Dapprich, J., D. Ferriola, K. Mackiewicz, P. M. Clark, E. Rappaport, M. D'Arcy, A. Sasson, X. Gai, J. Schug, K. H. Kaestner and D. Monos (2016). "The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity." BMC Genomics 17: 486.

Emous, V. (2017). "Verwachtte schade bloedluis 21 miljoen euro." Pluimveeweb. Retrieved from https://www.pluimveeweb.nl/artikel/163578-verwachtte-schade-bloedluis-21-miljoen-euro/

Escobar-Zepeda, A., A. Vera-Ponce de Leon and A. Sanchez-Flores (2015). "The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics." Front Genet 6: 348.

European Medicines Agency(2017). "EPAR summary for the public". Retrived from https://www.ema.europa.eu/en/documents/overview/exzolt-epar-summary-public_en.pdf

Falkenberg, M. (2018). "Mitochondrial DNA replication in mammalian cells: overview of the pathway." Essays Biochem 62(3): 287-296.

Gharbi, M., N. Sakly and M. A. Darghouth (2013). "Prevalence of Dermanyssus gallinae (Mesostigmata: Dermanyssidae) in industrial poultry farms in North-East Tunisia." Parasite 20: 41.

Goodwin, S., J. D. McPherson and W. R. McCombie (2016). "Coming of age: ten years of next-generation sequencing technologies." Nat Rev Genet 17(6): 333-351.

Institute, N. H. G. R. (2018). "Human Genome Project FAQ." Retrieved from https://www.genome.gov/human-genome-project/Completion-FAQ.

Jeyaprakash, A. and M. A. Hoy (2007). "The mitochondrial genome of the predatory mite Metaseiulus occidentalis (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features." Gene 391(1-2): 264-274.

Kilpinen, O., A. Roepstorff, A. Permin, G. Norgaard-Nielsen, L. G. Lawson and H. B. Simonsen (2005). "Influence of Dermanyssus gallinae and Ascaridia galli infections on behaviour and health of laying hens (Gallus gallus domesticus)." Br Poult Sci 46(1): 26-34.

Lu, H., F. Giordano and Z. Ning (2016). "Oxford Nanopore MinION Sequencing and Genome Assembly." Genomics Proteomics Bioinformatics 14(5): 265-279.

Mul, M. (2013). Fact sheet Poultry Red Mite in Europe.

Mullens, B. A., B. L. Chen and J. P. Owen (2010). "Beak condition and cage density determine abundance and spatial distribution of northern fowl mites, Ornithonyssus sylviarum, and chicken body lice, Menacanthus stramineus, on caged laying hens." Poult Sci 89(12): 2565-2572.

Navajas, M., Y. Le Conte, M. Solignac, S. Cros-Arteil and J. M. Cornuet (2002). "The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite Varroa destructor (Acari: Mesostigmata)." Mol Biol Evol 19(12): 2313-2317.

Oines, O. and S. Brannstrom (2011). "Molecular investigations of cytochrome c oxidase subunit I (COI) and the internal transcribed spacer (ITS) in the poultry red mite, Dermanyssus gallinae, in northern Europe and implications for its transmission between laying poultry farms." Med Vet Entomol 25(4): 402-412.

Oines, O., M. Isaksson, A. Hagstrom, S. Tavornpanich and R. K. Davidson (2014). "Laboratory assessment of sensitive molecular tools for detection of low levels of Echinococcus multilocularis-eggs in fox (Vulpes vulpes) faeces." Parasit Vectors 7: 246.

Opsteegh, M., M. Langelaar, H. Sprong, L. den Hartog, S. De Craeye, G. Bokken, D. Ajzenberg, A. Kijlstra and J. van der Giessen (2010). "Direct detection and genotyping of Toxoplasma gondii in meat samples using magnetic capture and PCR." Int J Food Microbiol 139(3): 193-201.

Price, D. R. G., T. Kuster, O. Oines, E. Margaret Oliver, K. Bartley, F. Nunn, J. F. L. Barbero, J. Pritchard, E. Karp-Tatham, H. Hauge, D. P. Blake, F. M. Tomley and A. J. Nisbet (2019).

"Evaluation of vaccine delivery systems for inducing long-lived antibody responses to Dermanyssus gallinae antigen in laying hens." Avian Pathol: 1-44.

Reyes, F. G., M. F. Valim and A. E. Vercesi (1996). "Effect of organic synthetic food colours on mitochondrial respiration." Food Addit Contam 13(1): 5-11.

Roberts, V. (2009). "Diseases of farmyard poultry part 4 - External and internal parasites of chicken." National Animal Disease Information Service.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A 74(12): 5463-5467.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nat Biotechnol 26(10): 1135-1145.

Shokolenko, I. N., G. L. Wilson and M. F. Alexeyev (2014). "Aging: A mitochondrial DNA perspective, critical analysis and an update." World J Exp Med 4(4): 46-57.

Sigognault Flochlay, A., E. Thomas and O. Sparagano (2017). "Poultry red mite (Dermanyssus gallinae) infestation: a broad impact parasitological disease that still remains a significant challenge for the egg-laying industry in Europe." Parasit Vectors 10(1): 357.

Sparagano, O. A., D. R. George, D. W. Harrington and A. Giangaspero (2014). "Significance and control of the poultry red mite, Dermanyssus gallinae." Annu Rev Entomol 59: 447-466.

Swafford, L. and J. E. Bond (2009). "The symbiotic mites of some Appalachian Xystodesmidae (Diplopoda: Polydesmida) and the complete mitochondrial genome sequence of the mite *Stylochyrus rario*r (Berlese) (Acari: Mesostigmata : Ologamasidae)." CSIRO Publishing.

Valiente Moro, C., C. J. De Luna, A. Tod, J. H. Guy, O. A. Sparagano and L. Zenner (2009). "The poultry red mite (Dermanyssus gallinae): a potential vector of pathogenic agents." Exp Appl Acarol 48(1-2): 93-104.

Vogt, Y. (2017). "Energimotoren i cellene våre stammer fra bakterier." Apollon.

Xia, C. Y., Y. Liu, H. R. Yang, H. Y. Yang, J. X. Liu, Y. N. Ma and Y. Qi (2017). "Reference Intervals of Mitochondrial DNA Copy Number in Peripheral Blood for Chinese Minors and Adults." Chin Med J (Engl) 130(20): 2435-2440.

Yasukawa, T. and D. Kang (2018). "An overview of mammalian mitochondrial DNA replication mechanisms." J Biochem 164(3): 183-193.

Appendix

Appendix 1: Nanodrop

Supplementary Table 1: Nanodrop values from both QIAamp Spin-column (man) and QIAcube (qube).

#	Sample ID	Nucleic Acid	Unit	A260 (Abs)	A280 (Abs)	260/280	260/230	Sample type
1	Blank	-0,5	ng/ul	-0,011	-0,003	4,08	0,54	DNA
2	DP1911A_qube	2,6	ng/ul	0,052	0,006	9,24	0,11	DNA
3	DP1911A_qube replikat	3,2	ng/ul	0,065	0,023	2,88	0,12	DNA
4	DP1911B_qube	1,4	ng/ul	0,02	0,017	1,38	0,78	DNA
5	DP1911B_qube replikat	1,2	ng/ul	0,017	0,01	1,39	0,26	DNA
6	DP1911C_qube	0,7	ng/ul	0,01	0,001	1,48	0,51	DNA
7	DP1911C_qube replikat	0,6	ng/ul	0,001	0,032	10,41	0,53	DNA
8	DP1911D_qube	3,2	ng/ul	0,032	0,03	2,02	0,18	DNA
9	DP1911D_qube replikat	2,7	ng/ul	0,03	0,029	1,8	0,14	DNA
10	DP1910E_qube	1,5	ng/ul	0,029	0,033	1,04	0,18	DNA
11	DP1910E_qube replikat	2,1	ng/ul	0,033	0,018	1,27	0,15	DNA
12	DP1910F_qube	1,2	ng/ul	0,018	0,009	1,34	2,33	DNA
13	DP1910F_qube replikat	0,9	ng/ul	0,009	0,027	1,95	1,76	DNA
14	DP1910G_qube	1,3	ng/ul	0,027	0,071	0,97	-211,28	DNA
15	DP1910G_qube replikat	3,9	ng/ul	0,071	0,022	1,1	0,63	DNA
16	DP1910H_qube	1,9	ng/ul	0,022	0,031	1,73	0,15	DNA
17	DP1910H_qube replikat	2,8	ng/ul	0,03	0,03	1,84	0,17	DNA
18	Neg_qube	1,4	ng/ul	0,03	0,03	1,1	0,35	DNA
19	DP1910A_man	1,4	ng/ul	0,031	0,024	1,18	0,36	DNA
20	DP1910A_man replikat	1,9	ng/ul	0,024	0,027	1,39	0,24	DNA
21	DP1910B_man	1	ng/ul	0,027	0,024	0,82	1,71	DNA
22	DP1910B_man replikat	1	ng/ul	0,024	0,012	1,68	-22,07	DNA
23	DP1910C_man	1,1	ng/ul	0,012	0,01	2,18	0,98	DNA
24	DP1910C_man replikat	2	ng/ul	0,01	0,017	1,23	1,04	DNA
25	DP1910D_man	1,7	ng/ul	0,017	0,037	1,07	3,39	DNA
26	DP1910D_man replikat	0,8	ng/ul	0,037	0,021	1,66	3,43	DNA
27	DP1911E_man	0,5	ng/ul	0,021	0,008	1,99	-1,5	DNA
28	DP1911E_man replikat	1,1	ng/ul	0,008	0,004	2,93	-0,86	DNA
29	DP1911F_man	0,9	ng/ul	0,004	0,019	1,16	0,65	DNA
30	DP1911F_man replikat	0,9	ng/ul	0,019	0,01	1,84	0,61	DNA
31	DP1911G_man	1,2	ng/ul	0,018	0,018	0,95	0,47	DNA
32	DP1911G_man replikat	1,3	ng/ul	0,017	0,012	2,04	0,49	DNA
33	DP1911H_man	2	ng/ul	0,025	0,007	3,87	-1,67	DNA
34	DP1911H_man replikat	1,3	ng/ul	0,039	0,026	1,53	-9,04	DNA

Appendix 2: DNA pooling

Supplementary Table 2: DNA pooling prior to Nanopore library preparation

B9 - GenomiPhi

	[DNA] ng/ul	Volum(ul)	Total DNA(ng)
DP1798	14,6	145	2117
DP1799	16,3	145	2363,5
DP1800	14,7	145	2131,5
DP1801	14,6	145	2117
DP1802	15	145	2175
DP1803	7,4	145	1073
DP1804	10,9	145	1580,5
			13557,5

A16 - GenomiPhi

	[DNA] ng/ul	Volum(ul)	Total DNA(ng)
DP1910G	19,1	145	2769,5
DP1910C	20,4	145	2958
DP1910D	21,2	145	3074
			8801,5

B9 Standard

	[DNA] ng/ul	Volum(ul)	Total DNA(ng)
DP1798	1,24	30	37,2
DP1799	0,73	30	21,9
DP1800	2,58	30	77,4
DP1801	2,05	30	61,5
DP1802	0,269	30	8,07
DP1803	0,464	30	13,92
Sum			**219,99**

A16 Standard

	[DNA] ng/ul	Volum(ul)	Total DNA(ng)
DP1910G	0,02	160	3,2
DP1019C	0,0356	160	5,696
DP1919D	0,173	160	27,68
DP1914A	0,0244	194	4,7336
DP1914B	0,42	194	81,48
DP1914C	0,664	194	128,816
DP1914D	0,478	194	92,732
			344,3376

QUBIT ETTER RENSING

B9

	[DNA] ng/ul	Volum(ul)	Total DNA(ng)
DP1778	0,196	100	19,6
DP1779	0,116	100	11,6
DP1800	0,175	100	17,5
DP1801	0,219	100	21,9
DP1802	0,117	100	11,7
DP1803	0,129	100	12,9
			95,2

A16

	[DNA] ng/ul	Volum(ul)	Total DNA(ng)
DP1910G	0,066	100	6,6
DP1910C	0,186	100	18,6
DP1910D	0,216	100	21,6
DP1914A	0,071	100	7,1
DP1914B	0,579	100	57,9
DP1914C	1,06	100	106
DP1914D	0,852	100	85,2
			303

Appendix 3: HiSeq assembly report from previous project.

Analysis Report
(De novo assembly)

Hiseq 2000

Institution	:	Norwegian Veterinary Institute
Name	:	Oivind Oines
Order No.	:	1305KHS-0063
Sample Name	:	DG00

Index

1. Description of Workflow

side 60

1.1 Sequecne quality check

1.1.1 FastQC

FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis.

More information can be found here (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

1.1.2 Quality filtering

The reads were filtered before assembly such that for a pair of PE reads each read should have more than 90 % of bases with base quality greater than or equal to Q20.

1.2 Data Analysis

1.2.1 K-mer analysis

JELLYFISH is a tool for fast, memory-efficient counting of k-mers in DNA. A k-mer is a substring of length k, and counting the occurrences of all such substrings is a central step in many analyses of DNA sequence. JELLYFISH can count k-mers using an order of magnitude less memory and an order of magnitude faster than other k-mer counting packages by using an efficient encoding of a hash table and by exploiting the "compare-and-swap" CPU instruction to increase parallelism.

(http://www.cbcb.umd.edu/software/jellyfish/)

1.2.2 De novo assembly

SOAPdenovo is a novel short-read assembly method that can build a de novo draft assembly for the human-sized genomes. The program is specially designed to assemble Illumina GA short reads. It creates new opportunities for building reference sequences and carrying out accurate analyses of unexplored genomes in a cost effective way. Now the new version is available. SOAPdenovo2, which has the advantage of a new algorithm design that reduces memory consumption in graph construction, resolves more repeat regions in contig assembly, increases coverage and length in scaffold construction, improves gap closing, and optimizes for large genome.
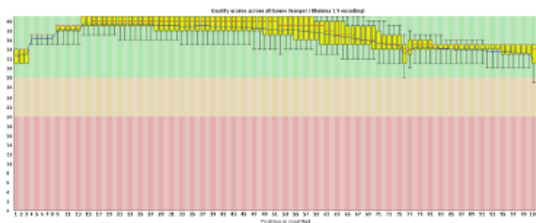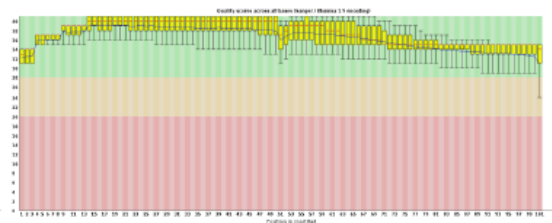
(http://soap.genomics.org.cn/soapdenovo.html)

2. Results of Data processing

2.1 Filtered data

Library	Original		Filtered	
	Total Reads	Total Bases	Total Reads	Total Bases
DG00	13,480,016	1,361,481,616	11,011,618	1,112,173,418

DG00 _1_filtered_fastqc

DG00 _2_filtered_fastqc

Sample	Total Bases	Read Count	N (%)	GC (%)	Q20 (%)	Q30 (%)
DG00	1,112,173,418	11,011,618	7.0E-4	40.14	99.1	95.16

- Total Bases : The total number of bases in reads.
- Reads Count : The number of reads identified.
- N(%) : The percentage of bases on a DNA that are N.
- GC(%) : The percentage of bases on a DNA that are either guanine or cytosine.
- Q20(%) : The percentage of bases called that have a quality score of 20 or above.
- Q30(%) : The percentage of bases called that have a quality score of 30 or above.

3. Results of Analysis

3.1 K-mer analysis
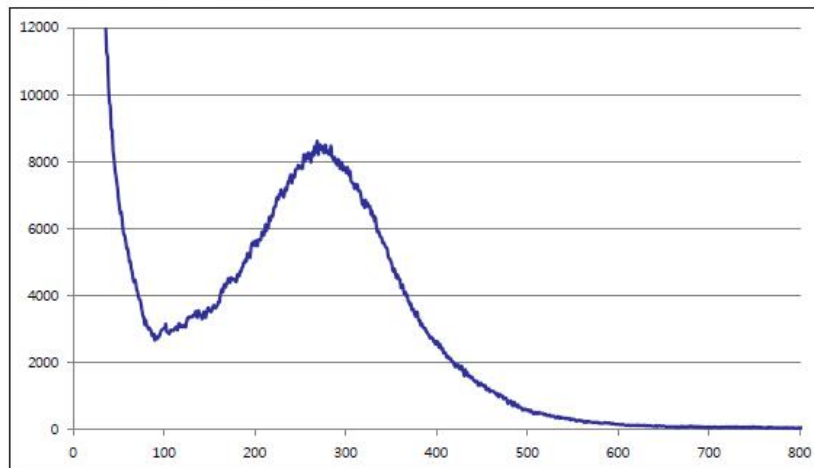
| Estimation Genome Size	1,862,056.17

side 62

3.2 Assembly Results

3.2.1 SOAPdenovo Assembly summary (length >100)

Number of scaffolds	Scaffolds sum	N50	Longest scaffold	Shortest scaffold	Average length
11,103	3,604,799	3,006	159,025	100	324

- Number of contigs : The number of contigs identified.
- Contig sum : The total number of bases in the contigs.
- N50 : An N50 means that half of all bases reside in contigs of this size or longer.
- Longest contig : The sequence size of the longest contig.
- Shortest contig : The sequence size of the shortest contig.
- Average length : The average contig size

3.2.2 Length Distribution

side 63
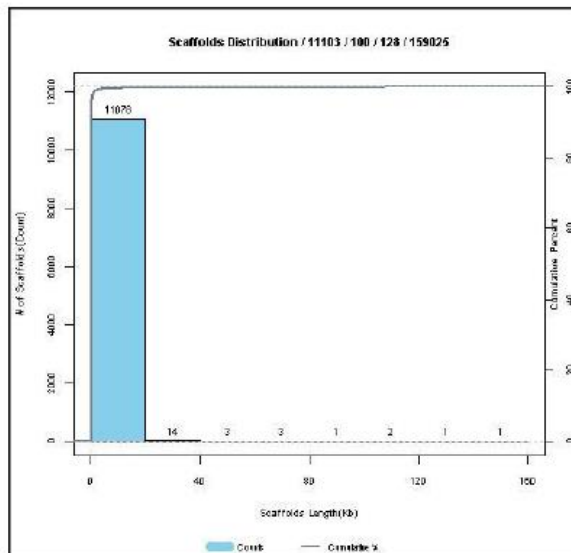
3.2.3 Contig Statistics

	Num of Scaffolds	Length	Avg.length	Sum
N10	3	116,750	138,081	414,243
N20	7	61,641	106,752	747,270
N30	15	33,649	73,078	1,096,172
N40	31	16,690	47,023	1,457,731
N50	83	3,006	21,747	1,805,021
N60	565	322	3,828	2,162,932
N70	2,390	159	1,055	2,523,423
N80	4,763	138	605	2,883,905
N90	7,692	114	421	3,244,360
N100	11,103	100	324	3,604,799

- Num of sequences : the number of sequences in the contig statistics (N10~N100)
- Length : the length of sequence in the contig statistics (N10~N100)
- Avg. length : the average length in the contig statistics (N10~N100)
- Sum : the sum of the length in the contig statistics (N10~N100)

3.2.4 GC Contents

Num of A	Num of T	Num of G	Num of C	Num of N	GC contents
1100949	1084609	690334	723382	5625	39.22%

Histogram of GC
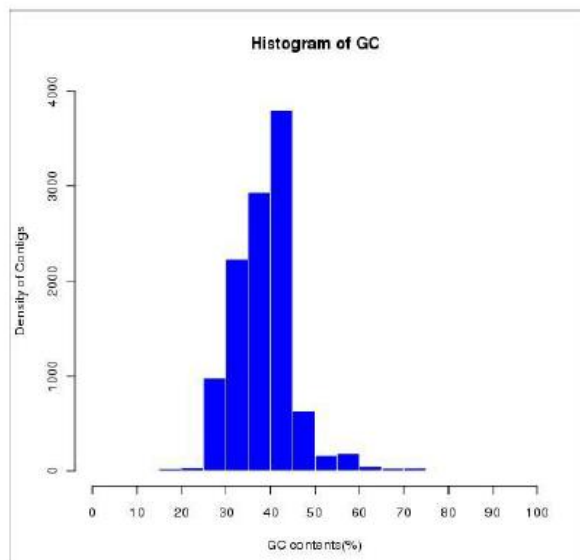
- Num of A : the total number of adenine(A)
- Num of T : the total number of thymine(T)
- Num of G : the total number of guanine(G)
- Num of C : the total number of cytosine(C)
- Num of N : the total number of ambiguous
- GC contents : the percentage of guanine-cytosine base pairs.

4. Data download

4.1 Result of Analysis

Download link : contigs sequence of assembly results

Download link : scaffolds sequence of assembly results